Understanding the kinetic mechanism of RNA single base pair formation

Xiaojun Xu[a,b,c], Tao Yu[a,b,c,d], and Shi-Jie Chen[a,b,c,1]

[a]Department of Physics, University of Missouri, Columbia, MO 65211; [b]Department of Biochemistry, University of Missouri, Columbia, MO 65211; [c]Informatics Institute, University of Missouri, Columbia, MO 65211; and [d]Department of Physics, Jianghan University, Wuhan, Hubei 430056, China

Edited by Ken A. Dill, Stony Brook University, Stony Brook, NY, and approved November 20, 2015 (received for review September 3, 2015)

RNA functions are intrinsically tied to folding kinetics. The most elementary step in RNA folding is the closing and opening of a base pair. Understanding this elementary rate process is the basis for RNA folding kinetics studies. Previous studies mostly focused on the unfolding of base pairs. Here, based on a hybrid approach, we investigate the folding process at level of single base pairing/stacking. The study, which integrates molecular dynamics simulation, kinetic Monte Carlo simulation, and master equation methods, uncovers two alternative dominant pathways: Starting from the unfolded state, the nucleotide backbone first folds to the native conformation, followed by subsequent adjustment of the base conformation. During the base conformational rearrangement, the backbone either retains the native conformation or switches to nonnative conformations in order to lower the kinetic barrier for base rearrangement. The method enables quantification of kinetic partitioning among the different pathways. Moreover, the simulation reveals several intriguing ion binding/dissociation signatures for the conformational changes. Our approach may be useful for developing a base pair opening/closing rate model.

RNA folding kinetics | molecular dynamics | master equation | kinetic Monte Carlo | conformational network

RNAs perform critical cellular functions at the level of gene expression and regulation (1–4). RNA functions are determined not only by RNA structure or structure motifs [e.g., tetraloop hairpins (5, 6)] but also by conformational distributions and dynamics and kinetics of conformational changes. For example, riboswitches can adopt different conformations in response to specific conditions of the cellular environment (7, 8). Understanding the kinetics, such as the rate and pathways for the conformational changes, is critical for deciphering the mechanism of RNA function (9–19). Extensive experimental and theoretical studies on RNA folding kinetics have provided significant insights into the kinetic mechanism of RNA functions (19–36). However, due to the complexity of the RNA folding energy landscape (37–46) and the limitations of experimental tools (47–55), many fundamental problems, including single base flipping and base pair formation and fraying, remain unresolved. These unsolved fundamental problems have hampered our ability to resolve other important issues, such as RNA hairpin and larger structure folding kinetics. Several key questions remain unanswered, such as whether the hairpin folding is rate-limited by the conformational search of the native base pairs, whose formation leads to fast downhill folding of the whole structure, or by the breaking of misfolded base pairs before refolding to the native structure (18, 19, 54–73).

Motivated by the need to understand the basic steps of nucleic acids folding, Hagan et al. (74) performed forty-three 200-ps unfolding trajectories at 400 K and identified both on- and off-pathway intermediates and two dominant unfolding pathways for a terminal C-G base pair in a DNA duplex. In one of the pathways, base pairing and stacking interactions are broken concomitantly, whereas in the other pathway, base stacking is broken after base pairing is disrupted. Furthermore, the unfolding requires that the Cyt diffuse away from the pairing Gua to a distance such that the C-G hydrogen bond cannot reform easily. More recently, Colizzi and Bussi (75) performed molecular dynamics (MD) pulling simulations for an RNA duplex and construct free energy landscape from the pulling simulation. The simulation showed that the base pair opening reaction starts with the unbinding of the 5′-base, followed by the unbinding of the 3′-base (i.e., the 5′-base is less stable than the 3′-base). These previous unfolding simulations offered significant insights into the pathways and transition states. However, as shown below, several important issues remain.

One intriguing problem is the rate model for base pairing. There are currently three main types of models. In the first type of model, the barrier $\Delta G_{+}^{\ddagger}$ for closing a base pair is dominated by the entropic cost ΔS for positioning the nucleotides to the base-paired configuration and the barrier $\Delta G_{-}^{\ddagger}$ for opening a base pair is the enthalpic cost ΔH for disrupting the hydrogen bonds and base stacking interactions (18, 59, 60). In the second type of model, $\Delta G_{+}^{\ddagger}$ is the net free energy change for base pairing $\Delta G = \Delta H - T\Delta S$ and $\Delta G_{-}^{\ddagger}$ is zero (76, 77). In the third type of model, $\Delta G_{\pm}^{\ddagger} = \pm \Delta G/2$ is used (78). In addition to the above three main types, other models, such as more sophisticated hybrid rate models, have been proposed (29).

In this paper, we report a hybrid method (see Fig. 1) to investigate the single base pairing process. In contrast to the previous simulations for temperature- or force-induced unfolding reactions, we directly model the folding process here (i.e., the base pair closing process). Specifically, we use MD simulations to identify the conformational clusters. Based on the network of the conformational clusters as a reduced conformational ensemble, we apply kinetic Monte Carlo (KMC) and master equation (ME) methods to elucidate the detailed roles of base pairing and stacking interactions, as well as the roles of water and ions (79–82). The study reveals previously unidentified kinetics pathways, misfolded states, and rate-limiting steps. A clear understanding of the microscopic details of the elementary kinetic move is a prerequisite for further rigorous study of large-scale RNA kinetic studies. The method described here may provide a feasible way to develop a rate model for the base pair/stack-based kinetic move set. Furthermore, the

Significance

This paper addresses a long-standing unsolved issue in RNA folding, namely, the kinetics for the very elementary steps of RNA folding. Lacking the ability to identify the exact microscopic, molecular level processes has been responsible for a variety of inaccurate predictions for RNA dynamics and RNA biochemical, such as enzymatic, reactions. Unlike previous unfolding (pulling) simulations, we here perform a direct folding study for the process. The misfolded states and the kinetic role of ions discovered in the study may provide significant insights into the physical mechanism of RNA conformational changes in biology.

Author contributions: X.X. and S.C. designed research; X.X., T.Y., and S.C. performed research; X.X., T.Y., and S.C. analyzed data; and X.X., T.Y., and S.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: chenshi@missouri.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1517511113/-/DCSupplemental.

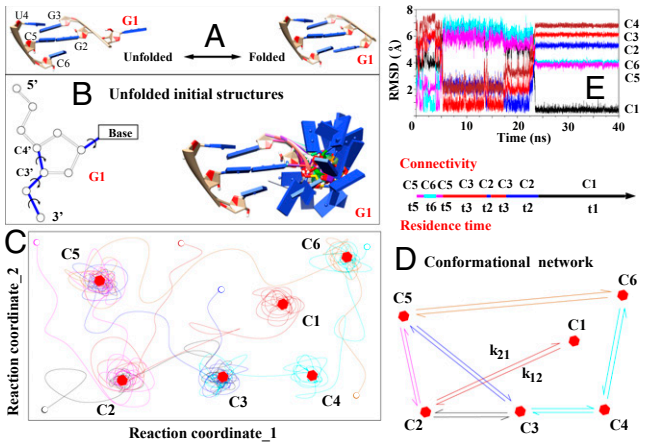

Fig. 1. (*A*) Folding of a single nucleotide (G1, red) from the unfolded (*Left*) to the native folded (*Right*) state. (*B*) Exhaustive sampling for the (discrete) conformations of the G1 nucleotide (*Right*) through enumeration of the torsion angles (formed by the blue bonds). (*C*) Schematic plot shows the trajectories on the energy landscape (depicted with two reaction coordinates for clarity) explored by the MD simulations. The lines, open circles, and hexagons denote the trajectories; the initial states; and the (centroid structures of the) clusters, respectively. (*D*) Conformational network based on six clusters. (*E*) The rmsds to the different clusters provide information about the structural changes in a MD trajectory.

mechanism of RNA single base folding may provide useful insights into many biologically significant processes, such as nucleotide flipping (83) in helicases and base pair fraying (84) (as the possible first step for nucleic duplex melting in nucleic acid enzymatic processes).

Results

Conformational Clusters. The MD trajectories give a total of 50 clusters (Fig. 2). These 50 clusters are determined from the structural closeness as measured by the rmsd. We find we can also use three order parameters to classify the clusters: the distance d_{12} between the geometric centers of all of the heavy atoms in the G1 and G2 bases and the nonbonded interaction energies E_{12} and E_{16} between the sequentially neighboring nucleotides G1 and G2 and between the pairing nucleotides G1 and C6, respectively. As shown in Fig. 2, the clusters can be separated by these three order parameters. It should be noted that the all-atom energy, as determined by the force field, is much more sensitive to the small changes in the bond lengths, bond angles, and torsional angles than the rmsd values. Even though conformations in the same structural cluster have small rmsds between each other, their E_{12} and E_{16} values may vary significantly. As a result, it might not be ideal to identify order parameters by methods like principle component analysis. In *SI Appendix*, we show the distributions as a function of other order parameters and the centroid structures of all of the 50 clusters.

As illustrated in Fig. 1 *C* and *D*, with the conformational connectivity and the intercluster transition rate constants extracted from the MD trajectories, we build the cluster-based conformational network. Detailed structural analysis showed that many of the clusters were formed due to the nonnative ribose orientation of the G1 nucleotide. Unlike the native structure, which involves stabilizing stacking interactions between the G1, G2, and G3 bases, the intermediate states (clusters) have the G1 base excursion away from the native orientation, causing weaker or no stacking interactions between the bases. Furthermore, we find that a structure of an intermediate state (cluster) is usually more flexible with larger conformational fluctuations than the native cluster M_1 (Fig. 1*E*). The base stacking interaction (E_{12}) between nucleotides G1 and G2 dominates the overall stability of G1 for the folded cluster M_1 and the partially folded clusters, such as M_3 and M_5 in Fig. 2. The different structures involve competition between the different nonbonded interactions. For example, the folded cluster M_1 has the lowest value of base pairing interaction E_{16} but a weaker base stacking interaction E_{12}. The partially folded cluster M_6 has destabilizing base pairing energy E_{16} (>0) but stabilizing stacking energy E_{12} (<0). We note that these energies do not include more distant interactions, such as those interactions between G1 and G3, U4, and C5 (a more detailed analysis is provided in *SI Appendix*).

Folding Pathway. To validate the extracted kinetic cluster network, we first bin all of the conformations (snapshots) in the trajectories into their respective clusters. From the populations of the individual conformations, we compute the time-dependent populations for each cluster (Fig. 3*A*). We then use the population of each cluster at time $t = 0$ as the initial population and the ME method to predict the subsequent populational kinetics of the clusters (Fig. 3*A*). We find that such ME-predicted populational kinetics agree with the original populations from the MD trajectories. The result suggests that the conformational cluster network may be reliable.

As shown in Fig. 2, cluster M_{23} has the largest value of d_{12}, with the G1 base totally flipped out of the stacked position, and thus has very weak base pairing interactions E_{16}. We investigate the folding process starting from this unfolded cluster. We set the initial fractional population of M_{23} to be 100%. The ME solution predicts that (Fig. 3*B*) through conversions with M_{21} and M_{44}, M_{23} quickly relaxes within 10^{-10} s. A close observation of the structures in Fig. 2 suggests that clusters M_{21}, M_{23}, and M_{44} have similar base pairing and stacking energies, with slightly different distances of d_{12}. The fast conversions (details are provided in *SI Appendix*) between them lead to a fast preequilibration between the unfolded conformations. Following such a rapid initial relaxation, M_{13} emerges between 10^{-10} s and 10^{-9} s. M_{13} has the native-like backbone but nonnative base orientations. M_{13} is a misfolded intermediate state. Subsequently, between 10^{-9} s and 10^{-8} s, part of the unfolded population is converted to M_{13}. M_{13} reaches the maximum population of $p_{M_{13}} \sim 0.1$ at $t \sim 6.0 \times 10^{-9}$ s. At around 10^{-9} s (after M_{13} reaches its peak population), the native folded cluster M_1 emerges.

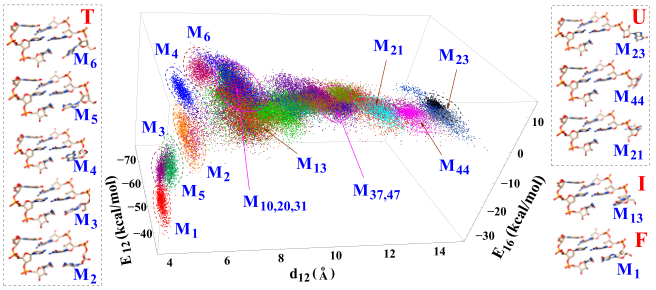

Fig. 2. Order parameters used to classify conformations: d_{12} is the distance between the geometric centers of heavy atoms between the G1 and G2 bases, and E_{12} and E_{16} are the nonbonded interactions (van der Waals and electrostatic energies given by VMD) between nucleotides G1 and G2 and between nucleotides G1 and C6, respectively. These energies characterize the base stacking and pairing interactions. The figure shows 1,000 randomly selected structures for each cluster. Ten centroid structures are shown for illustration. M_1 is the native folded structure. M_{21}, M_{23}, and M_{44} are the unfolded structures with large d_{12}. Other structures are partially folded or misfolded structures. M_2, M_3, and M_5 have strong base stacking (E_{12}) and pairing (E_{16}) interactions. M_4 and M_{13} have native-like backbone conformations, especially for the sugar ring atoms, but a partially unfolded base orientation. M_6 has nonnative conformations for both the base and the backbone of nucleotide G1. The magenta circles indicate the order parameters for the transition states. Structures of the other clusters, as well as the other order parameters, are shown in *SI Appendix*.

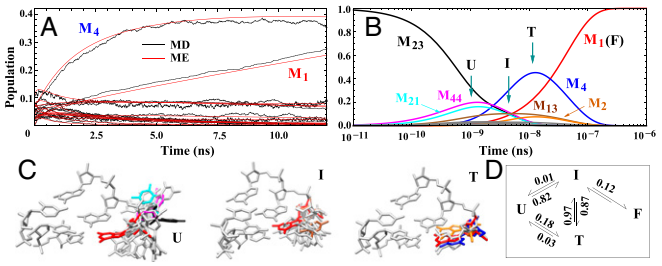

Fig. 3. (*A*) Comparison between the kinetics results from MD simulations and from the ME method (red). The ME method is based on the 50-cluster conformational network. As shown in Fig. 1*E*, we bin the MD snapshots according to the 50 clusters to obtain the time-dependent populations for each cluster from MD (shown in black). The results support the validity of the 50-cluster network model. (*B*) ME-predicted, long-time, single-nucleotide folding kinetics starting from the unfolded cluster M_{23} (black line). Based on the behavior of the populational kinetics for each cluster, we classify the 50 clusters into four states: unfolded (U), intermediate (I), trapped (T), and folded (F). (*C*) Typical structures of the three states are shown, with the folded state (F, in red) included for comparison. Structures of the clusters with significant populations in *B* are shown in *C* with the same color code. (*D*) Four-state scheme for the single-nucleotide kinetics. The numbers indicate the probabilities of the transitions between the states, (e.g., 0.87 for I → T, 0.01 for I → U, and 0.12 for I → F).

In the meantime, clusters M_4 and M_2 show a significant populational jump, with peak populations at $p_{M_4} \approx 0.5$ and $p_{M_2} \approx 0.08$ at $t \approx 1.5 \times 10^{-8}$ s. As shown in Fig. 2, M_4 has the native-like backbone configuration and (antiparallel, partially) stacked base orientation. Similar to M_4, M_2 has a native-like (stacked) base orientation and nonnative backbone position for G1. To fold to the native folded cluster, these two states first flip out their bases so they can make necessary adjustments for their backbone conformations. After the rearrangement of the backbone conformations, the bases flip back to the native configuration. Similar to M_2 and M_4, there are also slight populational changes of M_3 and M_5. These four clusters act as trapped states for the folding reaction, because the RNA must first flip out the base (detrapping) before folding to the folded state. There are transitions from the misfolded clusters of M_2, M_4, and M_{13} to the folded cluster M_1 in the time window from 10^{-8} s to 3.0×10^{-7} s. The overall folding time is about 10^{-6} s, which is consistent with the experimental results about the time of the single-strand stacking formation and helix elongation of DNA and RNA (85–88).

Based on the behavior of the time-dependent population for each cluster (details are provided in *SI Appendix*), we classify the 50 clusters into four states: unfolded (U), intermediate (I), trapped (T), and folded (F); several typical structures are illustrated in Fig. 3*C*. The U state contains unfolded clusters, such as M_{21}, M_{23}, and M_{44}. In these clusters, the base flips out and both the base and the backbone (of G1) have nonnative conformations. In the I state (e.g., M_{13}), the backbone is native-like and the base is partially flips out and is in a nonnative conformation. For such conformations, folding proceeds through rotations of the base without changing the backbone conformation and the folding is rate-limited by the process of searching for the correct (native) base conformation. For the misfolded (trapped) state, such as M_2, M_3, M_4, M_5, and M_6, the G1 base is in the stacked conformation and the backbone is in a nonnative conformation. In the folding process, to lower the overall kinetic barrier, the base stacking in a T conformation is disrupted first so that the rearrangement of the backbone conformation can occur. Indeed, we find that the G1 base flips out from the native conformation before the backbone switches to the native conformation. The overall folding kinetics can be described by the above four-state kinetic scheme (Fig. 3*D*).

We quantify the kinetic partitioning between the different pathways from the transition probabilities. As shown in Fig. 3*D*, starting from the U state, 82% of the transitions go to the I state compared with 18% to the T (trapped) state. Folding to the I and T states is rate-limited by the search for the correct (native) backbone and the base conformations, respectively. The interactions between the G1 and other atoms overcome the entropic loss involved in the G1 conformational searching process. Neither I nor T has a significant possibility to return to the U state, suggesting that the favorable interactions, such as E_{12} and E_{16}, inhibit the G1 base from flipping back to the unfolded positions. The I → T transition, with a probability of 0.87, dominates over the I → F transition, which has a smaller probability of 0.12. The kinetic barrier difference between the two transitions is around $k_BT\ ln(\frac{0.87}{0.12}) \approx 2.0\ k_BT$. Here, the T state has a small rate for the exit transitions (from T to I or to U); therefore, it is indeed a trap. From the above analysis, we conclude that the overall folding process from U to F is rate-limited by the enthalpic barrier for the T → I detrapping transition and the entropic barrier for the I → F folding transition.

Transition States. We next identify the transition states (details are provided in *SI Appendix*) for the two aforementioned rate-limiting transitions (T → I and I → F). We first focus on the I → F folding transition. We start from random unfolded clusters and use the KMC method to compute the probability P for a random initial cluster to enter the folded state F (structure M_1), instead of I (M_{13}) (Fig. 4*A*). We call the conformations with P between 0.35 and 0.65 (between the two blue lines in Fig. 4*A*) the "transition states." These transition states may be close to the actual transition states that have $P = 0.5$. The transition states (M_{10}, M_{20}, and M_{31} shown in gray in Fig. 4*B*) have native-like backbone conformations but partially folded base orientations. They all belong to the I (intermediate) state. Similar analysis gives the transition states for the detrapping transition from T (M_{13}) to I ($M_{2,4}$) (Fig. 4*C*). The two clusters M_{37} and M_{47} (Fig. 4*D*) in the transition state region (P = 0.45 and 0.55, respectively) have a slightly smaller probability to enter the trapped states (M_2 and M_4) instead of the intermediate state (M_{13}). We note that M_{37} and M_{47} also belong to the intermediate state, but with large d_{12} values as shown in Fig. 2.

Fig. 4 *A* (I → F) and *C* (I → T) shows that only a few clusters have a large forward (folding) reaction probability P (above the blue lines) and that most starting conformations would likely

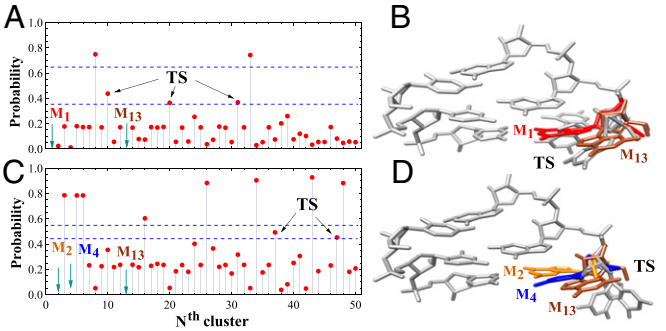

Fig. 4. (*A*) KMC-predicted probability of falling to the folded state M_1 before reaching the intermediate state M_{13} starting from the Nth cluster. (*B*) Three structures (M_{10}, M_{20}, and M_{31} in gray) of the transition states between the folded (red) and the intermediate (brown) states obtained from *A*. (*C*) KMC-predicted probability of falling to the trapped states M_2 and M_4 before reaching the intermediate state M_{13} starting from the Nth cluster. (*D*) Two structures (M_{37} and M_{47} in gray) of the transition states between the trapped (orange and blue) and intermediate (brown) states obtained from *C*. Because the kinetic cluster-based network can only give the clusters close to the transition states, we use the probability regime of (0.35, 0.65) and (0.45, 0.55) between the two dashed blue lines in *A* and *C*, respectively, to identify the transition state (TS).

enter/stay in the I state. The result implies more local minima around the I state than around the U, T, and F states. The formation of the energy minima can be understood from the interaction between the G1 nucleotide and the rest of the nucleotides. For the trapped state, the G1 base is at the native-like stacked position. The fixation of the base conformation restricts the energy minima formation. For the unfolded state, the G1 base flips out; therefore, the interaction between G1 and other nucleotides is weak. The weak interaction makes the landscape less rugged. For the intermediate state, the G1 base is not fixed to the native conformation; thus, it has greater freedom and the interaction is stronger than the unfolded state, causing a bumpy energy landscape with significantly more energy minima.

Ion Effects and Kinetic Pathways. To investigate the ion effect in the single base folding kinetics, we track the ion positions and the ion–RNA interactions in the different MD trajectories. On average, there are about six Na^+ ions within 4 Å of the van der Waals surface of the RNA. This number of "bound" ions corresponds to nearly complete charge neutralization of the RNA backbone. Na^+ ions mostly bind to the phosphate groups or the bases (*SI Appendix*, Fig. S13). The dwell time of the ions around the RNA can be as short as 5 ns. Furthermore, as shown in *SI Appendix*, Fig. S13, a small number of bound ions may dehydrate upon binding to RNA. For some tightly bound ions in the vicinity of the phosphate groups, we find that up to eight water molecules can be released, and three of them are from the first hydration shell of the ions.

The change of ion distribution along the trajectories demonstrates that ion binding events are directly coupled to the folding events and that ion binding around the base pairing region precedes the formation of the stable G1-C6 base pair. To quantify ion–RNA association, we compute the distance between the Na^+ ions and the O6, N1, N2, N4, N3, and O2 atoms of the G1 and C6 bases. These atoms are selected because they are directly involved in the G1-C6 base pair formation in the form of the O6-N4, N1-N3, and N2-O2 hydrogen bonding. These distances reflect the ion binding to the base.

In Fig. 5*B*, the O6-N4, N1-N3, and N2-O2 distances are nearly constant between $t = 20$ ns and $t = 29$ ns, suggesting the formation a stable structure during this time period. Detailed structural analysis shows that the structure is M_4 (Fig. 2), a trapped (T) state with (partially) stacked bases and a nonnative backbone conformation. Before M_4 unfolds, as shown in Fig. 5*A* and *SI Appendix*, Fig. S15, ions bind to the base pairing region to stabilize the structure. Specifically, as shown in Fig. 5*A* and *SI Appendix*, Fig. S15, from $t = 22$ to 27 ns, the small distances between Na^+ and the relevant atoms in the G1 base indicate ion binding around the base. At $t \approx 27$ ns, Na^+ ions dissociate from the base and bind on and off around the backbone. The ion dissociation from the base region causes destabilization the M_4 structure. At about $t = 29$ ns, the structure is disrupted, as shown by the large increase of the O6-N4, N1-N3, and N2-O2 distances. Structural analysis shows a T → I transition at $t = 29$ ns. At about $t = 30$ ns, the I → F transition occurs and the final stable structure is formed. After $t = 30$ ns, all of the three hydrogen bonds (O6-N4, N1-N3, and N2-O2) are stabilized. The above ion-promoted folding pathway is consistent with the KMC-predicted folding pathway (additional examples for the trajectories are provided in *SI Appendix*, Fig. S11).

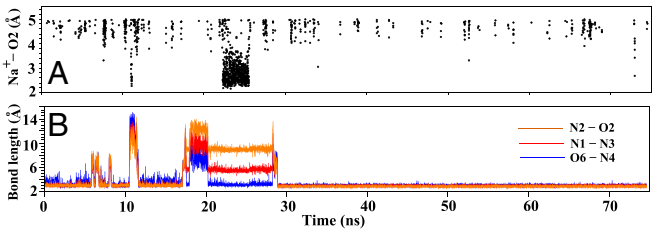

Fig. 5. (*A*) Na^+ ion distribution within a cutoff distance 5 Å from the O2 atom of C6. Ion distributions around other hydrogen bonding atoms are shown in *SI Appendix*, Fig. S15. (*B*) Donor–receptor distance distribution for the three hydrogen bonds (N2-O2, N1-N3, and O6-N4) along a typical MD simulation trajectory.

Discussion

Using an integrated approach with all-atom MD simulations, the ME method, and KMC simulations, we perform direct folding studies for single-nucleotide kinetics in a base pairing process. Based on the conformational clusters extracted from MD trajectories, we reveal a four-state kinetic mechanism. In the unfolded state U, the base of the nucleotide flips out of the folded state. In the intermediate state I, the base is in the partially unfolded conformation and the backbone is in the native orientation. In the trapped state T, the base is in the native conformation and the backbone is in the nonnative conformation. In the folded state F, both the base and the backbone are in the native state. The overall folding time ($\approx 10^{-6}$ s) is consistent with the experimental results for RNA and DNA single-strand stacking formation and helix elongation (85–88).

Physically, ion binding can lower the electrostatic repulsion (and kinetic barrier) between the nucleotides, thus inducing the close approach and interaction between RNA bases/backbone. Indeed, our results demonstrate the significant role of ions in determining the kinetic pathway. We find that ion binding can initialize folding events and stabilize misfolded as well as native states at the single-nucleotide level. In particular, ion binding to the base region can stabilize base stacking interactions, and ion dissociation from the base can destabilize the base stacking interactions and trigger a detrapping transition from a misfolded state.

The four-state kinetic scheme shows that there are two dominant barriers for the overall single base folding process: (*i*) the entropic barrier for the search of the native-like backbone orientation (U to I) and for the search of the native-like base position (U to T), and (*ii*) the enthalpic barrier for the adjustment of either the backbone (T → I) or the base (I → F) conformation. The existence of the trapped state, which was missed in previous studies (74, 75), makes the overall folding process more complex. The kinetics of this elementary folding step cannot be simply described by a two-state process (i.e., the kinetics are multistate). As a result, a rate model should consider the kinetic effects from the intermediate and trapped states. The transition state of the RNA single-nucleotide folding involves partially flipped-out base conformations stabilized by weak interactions with other nucleotides. The transition state suggests that the barrier for folding/disruption of the base pair/stack is not purely entropic/enthalpic. This finding suggests that the previous simple (ΔS, ΔH)-based rate model (18, 59, 60) should be modified. The kinetic barrier is likely a combination of entropic and enthalpic contributions. In addition, to test the Metropolis (ΔG-based) rate model (76, 77), we need to apply our method to different RNA systems with different sequences to extract the sequence dependence of the transition states. Because the overall properties of base-base stacking and hydrogen bonding are similar for different sequences, the conclusions about the folding kinetics, such as the folding pathways, kinetic intermediates/traps, and kinetic barriers derived from the present study, may be general, even though the detailed quantitative values can be different. The method developed here can be readily extended to study the more complicated processes, such as single base flipping in the middle of RNA/DNA helices and base pair fraying with the cooperative/uncooperative flipping of the two (paired) nucleotides.

Materials and Methods

To construct an effective reduced conformational ensemble, we first extract structural clusters from MD simulation. Specifically, we first exhaustively generate all of the possible (coarse-grained) nonnative structures. We then

perform MD simulations starting from each nonnative structure (as the initial state). The simulations sample hundreds of trajectories. We classify all of the snapshots in the trajectories into clusters according to the rmsd distance between the structures. We use the Visual Molecular Dynamics (VMD) (89) package to analyze the structural and energetic properties, such as the base pairing and base stacking interactions for each cluster. We implement two different nucleic acid force fields (90, 91) (AMBER ff99 and CHARMM27) with the same simulation protocol to examine the force field dependence of the predicted kinetics, and find the same general conclusions (92–94). From the kinetic connectivity between the conformations, we build a network of conformational clusters. From the KMC and ME methods for the conformational network of the clusters, we compute the detailed folding kinetics.

Initial (Unfolded) Structures. Previous pulling simulations suggested that a 3′-base is more stable than its 5′-base pairing partner. In a folding reaction, the 3′-base is likely to fold first through single-stranded base stacking, followed by folding of the 5′-base through base pairing with the 3′-base (75, 84, 95, 96). Here, we focus on the second step, namely, the folding kinetics of the 5′-base (nucleotide G1) after the 3′-base (nucleotide C6) is folded into the native state (Fig. 1*A*). Because the first step is much faster than the second step, our results may provide useful information about the overall rate process for base pair formation.

Based on the 3-bp folded structure (sequence 5′GGG3′-5′UCC3′; details are provided in *SI Appendix*), we rotate the torsional angles of the four bonds (P-O3′, O3′-C3′, C3′-C4′, and C1′-N9), shown in Fig. 1*B* in blue, to generate the complete (unfolded) conformational ensemble for the nucleotide G1. Using fivefold uniform rotational angles for each bond, we generate an ensemble of $5^4 = 625$ initial structures. Excluding structures disallowed by steric clashes, we obtain 570 viable initial structures (Fig. 1*B*).

MD Simulations. As illustrated in Fig. 1*C*, a major issue in MD simulations is that during the simulation, the sampled conformations often remain close to the initial unfolded structure. To circumvent this problem, we start from each and every unfolded conformation in the complete conformational ensemble as the initial state for the MD simulations (details are provided in *SI Appendix*). The use of the complete ensemble of the unfolded states is expected to enhance the completeness of the conformation sampling. For each initial structure, after a 500-step minimization for the whole system, we perform MD simulations for all of the atoms of the first (G1) nucleotide, as well as all of the water molecules and the ions. In total, we simulate 570 MD trajectories. For each trajectory, we run a 50-ns simulation. The simulation is terminated before 50 ns if the stable native structure is reached. The coordinates of all of the atoms are written to the NAMD (Nanoscale Molecular Dynamics program) dcd file every 5 ps.

Conformational Clusters. Clusters are identified as an ensemble of kinetically connected snapshots with small rmsds between them. To determine the clusters efficiently for a large pool of conformations, we use a two-step approach. First, from the simulated trajectories, we identify the segments in the trajectories such that the rmsds between any two snapshots in the segment are less than a cutoff value, such as 1.0 Å. For each segment, we identify the centroid structure, namely, the one with the minimum total rmsd to all of the structures in the segment. We can use a segment-centroid structure to represent a segment. As a result, each trajectory can be represented as a sequence of such segment-centroid structures. Second, after removing all of the possible redundant structures and applying the same procedure to the ensemble of the (segment-centroid) structures, we obtain a set of clusters represented by the respective centroid structures (Fig. 1 *C* and *D*). Because a cluster represents a (long-lived) macrostate consisting of consecutive conformations on the trajectory, we may view a cluster as a minimum on the free energy landscape.

Kinetic Network of Conformational Clusters. As illustrated in Fig. 1*E*, the rmsd values of the snapshots along a trajectory give the information about the transitions between the clusters and the residence time of each conformation. When a conformation falls in one of the clusters, its rmsd to the respective centroid structure would be smaller than the rmsds to the centroid structures of other clusters. For the case shown in Fig. 1*E*, the folding trajectory follows the sequential transitions of cluster C5 → cluster C6 → C5 → C3 → C2 → C3 → C2 → C1. Therefore, we can obtain the connectivities C5 ↔ C6, C5 → C3, C3 ↔ C2, and C2 → C1.

To determine the rate constants for transitions between the different clusters, for each cluster *C*, we first compute the mean residence time $\langle t^C \rangle$ from the MD trajectories. The inverse of $\langle t^C \rangle$ gives the total rate k_C^{total} for all of the transitions leaving the cluster. The rate constant $k_{CC'}$ for each individual intercluster transition $C \rightarrow C'$ can be computed from the transition frequency: $k_{CC'} = (N_{CC'}/N_C^{total})k_C^{total}$, where $N_{CC'}$ and N_C^{total} are the numbers of occurrences for the $C \rightarrow C'$ transition and for all of the exit transitions from cluster *C*, respectively (the KMC algorithm is illustrated in *SI Appendix*). For instance, the mean residence time for C1 is 5.9 ns. The total exit rate is $k_{C1}^{total} = \frac{1}{5.9\text{ns}} = 0.17\text{ns}^{-1}$. Among all of the (total of 326) exit transitions for C1, there are 293 transitions to C2, 22 to C3, and 11 to C6. Therefore, the rate for the C1 → C2 transition can be estimated as $k_{12} = \frac{293}{326} \times 0.17 = 0.152\ \text{ns}^{-1}$. Once the conformational cluster network and the intercluster rates (Fig. 1*D*) are set up, the folding kinetics of the system can be simulated by applying the KMC and ME methods to the cluster network.

Base Pairing/Stacking Interactions Analyzed by VMD. The stability of RNA structures is mainly influenced by three factors: base stacking, base pairing, and the chain flexibility. We used the VMD package (89) to calculate the nonbonded interactions between nucleotides G1 and G2 and between nucleotides G1 and C6 for the base primarily stacking and base pairing interactions, respectively. Furthermore, because the backbone is highly charged, metal ions and hydration effects can significantly influence the chain flexibility and the base–base and base–backbone interactions. Indeed, we find that ions play a critical role in initializing/promoting the folding process and in affecting the folding intermediate states and pathways.

ACKNOWLEDGMENTS. This research was supported by NIH Grant R01-GM063732.

1. Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2(12):919–929.
2. Hüttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: Hope or hype? *Trends Genet* 21(5):289–297.
3. Hirota K, et al. (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* 456(7218):130–134.
4. Morris KV (2012) *Non-Coding RNAs and Epigenetic Regulation of Gene Expression: Drivers of Natural Selection* (Caister Academic Press, Poole, UK).
5. Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal RNA: Constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci USA* 87(21):8467–8471.
6. Chakraborty D, Collepardo-Guevara R, Wales DJ (2014) Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *J Am Chem Soc* 136(52): 18052–18061.
7. Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* 15(3):342–348.
8. Haller A, Soulière MF, Micura R (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res* 44(12):1339–1348.
9. Weeks KM (1997) Protein-facilitated RNA folding. *Curr Opin Struct Biol* 7(3):336–342.
10. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282(5389):740–744.
11. Treiber DK, Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* 11(3):309–314.
12. Schroeder R, Grossberger R, Pichler A, Waldsich C (2002) RNA folding in vivo. *Curr Opin Struct Biol* 12(3):296–300.
13. Bevilacqua PC, Blose JM (2008) Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu Rev Phys Chem* 59:79–103.
14. Li PT, Vieregg J, Tinoco I, Jr (2008) How RNA unfolds and refolds. *Annu Rev Biochem* 77:77–100.
15. Zemora G, Waldsich C (2010) RNA folding in living cells. *RNA Biol* 7(6):634–641.
16. Dethoff EA, Chugh J, Mustoe AM, Al-Hashimi HM (2012) Functional complexity and regulation through RNA dynamics. *Nature* 482(7385):322–330.
17. Mustoe AM, Brooks CL, Al-Hashimi HM (2014) Hierarchy of RNA functional dynamics. *Annu Rev Biochem* 83:441–466.
18. Zhang W, Chen S-J (2002) RNA hairpin-folding kinetics. *Proc Natl Acad Sci USA* 99(4): 1931–1936.
19. Chen S-J (2008) RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* 37:197–214.
20. Woodson SA (2010) Compact intermediates in RNA folding. *Annu Rev Biophys* 39: 61–77.
21. Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res* 44(12):1312–1319.
22. Russell R, et al. (2002) Rapid compaction during RNA folding. *Proc Natl Acad Sci USA* 99(7):4266–4271.
23. Pabit SA, Sutton JL, Chen H, Pollack L (2013) Role of ion valence in the submillisecond collapse and folding of a small RNA domain. *Biochemistry* 52(9):1539–1546.
24. Schlatterer JC, et al. (2008) Hinge stiffness is a barrier to RNA folding. *J Mol Biol* 379(4):859–870.
25. Schlatterer JC, Martin JS, Laederach A, Brenowitz M (2014) Mapping the kinetic barriers of a Large RNA molecule's folding landscape. *PLoS One* 9(2):e85041.
26. Bunner AE, Beck AH, Williamson JR (2010) Kinetic cooperativity in Escherichia coli 30S ribosomal subunit reconstitution reveals additional complexity in the assembly landscape. *Proc Natl Acad Sci USA* 107(12):5417–5422.

27. Sarkar K, Meister K, Sethi A, Gruebele M (2009) Fast folding of an RNA tetraloop on a rugged energy landscape detected by a stacking-sensitive probe. *Biophys J* 97(5): 1418–1427.
28. Sarkar K, Nguyen DA, Gruebele M (2010) Loop and stem dynamics during RNA hairpin folding and unfolding. *RNA* 16(12):2427–2434.
29. Ma H, et al. (2006) Exploring the energy landscape of a small RNA hairpin. *J Am Chem Soc* 128(5):1523–1530.
30. Herschlag D, Allred BE, Gowrishankar S (2015) From static to dynamic: The need for structural ensembles and a predictive model of RNA folding and function. *Curr Opin Struct Biol* 30:125–133.
31. Solomatin SV, Greenfeld M, Chu S, Herschlag D (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 463(7281):681–684.
32. Lin JC, Thirumalai D (2013) Kinetics of allosteric transitions in S-adenosylmethionine riboswitch are accurately predicted from the folding landscape. *J Am Chem Soc* 135(44):16641–16650.
33. Koculi E, Cho SS, Desai R, Thirumalai D, Woodson SA (2012) Folding path of P5abc RNA involves direct coupling of secondary and tertiary structures. *Nucleic Acids Res* 40(16):8011–8020.
34. Hofacker IL, et al. (2010) BarMap: RNA folding on dynamic energy landscapes. *RNA* 16(7):1308–1316.
35. Kucharík M, Hofacker IL, Stadler PF, Qin J (2014) Basin Hopping Graph: A computational framework to characterize RNA folding landscapes. *Bioinformatics* 30(14): 2009–2017.
36. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 450(7173):1263–1267.
37. Chen S-J, Dill KA (2000) RNA folding energy landscapes. *Proc Natl Acad Sci USA* 97(2): 646–651.
38. Russell R, et al. (2002) Exploring the folding landscape of a structured RNA. *Proc Natl Acad Sci USA* 99(1):155–160.
39. Li PT, Bustamante C, Tinoco I, Jr (2007) Real-time control of the energy landscape by force directs the folding of RNA molecules. *Proc Natl Acad Sci USA* 104(17):7039–7044.
40. Shcherbakova I, Mitra S, Laederach A, Brenowitz M (2008) Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol* 12(6): 655–666.
41. Xu X, Chen SJ (2012) Kinetic mechanism of conformational switch between bistable RNA hairpins. *J Am Chem Soc* 134(30):12499–12507.
42. Yingling YG, Shapiro BA (2005) Dynamic behavior of the telomerase RNA hairpin structure and its relationship to dyskeratosis congenita. *J Mol Biol* 348(1):27–42.
43. Bello-Rivas JM, Elber R (2015) Simulations of thermodynamics and kinetics on rough energy landscapes with milestoning. *J Comput Chem*, 10.1002/jcc.24039.
44. Xia Z, Bell DR, Shi Y, Ren P (2013) RNA 3D structure prediction by using a coarse-grained model and experimental data. *J Phys Chem B* 117(11):3135–3144.
45. Whitford PC, Sanbonmatsu KY, Onuchic JN (2012) Biomolecular dynamics: Order-disorder transitions and energy landscapes. *Rep Prog Phys* 75(7):076601.
46. Whitford PC, Blanchard SC, Cate JH, Sanbonmatsu KY (2013) Connecting the kinetics and energy landscape of tRNA translocation on the ribosome. *PLOS Comput Biol* 9(3): e1003003.
47. Sclavi B, Sullivan M, Chance MR, Brenowitz M, Woodson SA (1998) RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279(5358): 1940–1943.
48. Nagel JHA, Gultyaev AP, Oistamo KJ, Gerdes K, Pleij CWA (2002) A pH-jump approach for investigating secondary structure refolding kinetics in RNA. *Nucleic Acids Res* 30(13):e63.
49. Ha T, et al. (1999) Ligand-induced conformational changes observed in single RNA molecules. *Proc Natl Acad Sci USA* 96(16):9077–9082.
50. Lynch DC, Schimmel PR (1974) Effects of abnormal base ionizations on Mg2 plus binding to transfer ribonucleic acid as studied by a fluorescent probe. *Biochemistry* 13(9):1852–1861.
51. Bina-Stein M, Crothers DM (1975) Localization of the structural change induced in tRNA fMET (Escherichia coli) by acidic pH. *Biochemistry* 14(19):4185–4191.
52. Wenter P, Furtig B, Hainard A, Schwalbe H, Pitsch S (2006) A caged uridine for the selective preparation of an RNA fold and determination of its refolding kinetics by real-time NMR. *ChemBioChem* 7(3):417–420.
53. Wenter P, Furtig B, Hainard A, Schwalbe H, Pitsch S (2005) Kinetics of photoinduced RNA conformational switch by real-time NMR spectroscopy. *Angew Chem Int Ed* 44(17):2600–2603.
54. Fogarty K, McPhee JT, Scott E, Van Orden A (2009) Probing the ionic atmosphere of single-stranded DNA using continuous flow capillary electrophoresis and fluorescence correlation spectroscopy. *Anal Chem* 81(1):465–472.
55. Jung J, Van Orden A (2006) A three-state mechanism for DNA hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *J Am Chem Soc* 128(4):1240–1249.
56. Ansari A, Kuznetsov SV, Shen Y (2001) Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc Natl Acad Sci USA* 98(14): 7771–7776.
57. Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc Natl Acad Sci USA* 95(15):8602–8606.
58. Wallace MI, Ying L, Balasubramanian S, Klenerman D (2001) Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc Natl Acad Sci USA* 98(10):5584–5589.
59. Zhang W, Chen S-J (2006) Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophys J* 90(3):765–777.
60. Zhang W, Chen S-J (2006) Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states. *Biophys J* 90(3):778–787.
61. Jung J, Van Orden A (2005) Folding and unfolding kinetics of DNA hairpins in flowing solution by multiparameter fluorescence correlation spectroscopy. *J Phys Chem B* 109(8):3648–3657.
62. Jung J, Ihly R, Scott E, Yu M, Van Orden A (2008) Probing the complete folding trajectory of a DNA hairpin using dual beam fluorescence fluctuation spectroscopy. *J Phys Chem B* 112(1):127–133.
63. Orden AV, Jung J (2008) Review fluorescence correlation spectroscopy for probing the kinetics and mechanisms of DNA hairpin formation. *Biopolymers* 89(1):1–16.
64. Ma H, Wan C, Wu A, Zewail AH (2007) DNA folding and melting observed in real time redefine the energy landscape. *Proc Natl Acad Sci USA* 104(3):712–716.
65. Bowman GR, et al. (2008) Structural insight into RNA hairpin folding intermediates. *J Am Chem Soc* 130(30):9676–9678.
66. Hyeon C, Thirumalai D (2008) Multiple probes are required to explore and control the rugged energy landscape of RNA hairpins. *J Am Chem Soc* 130(5):1538–1539.
67. Lee TH, et al. (2007) Measuring the folding transition time of single RNA molecules. *Biophys J* 92(9):3275–3283.
68. Woodside MT, et al. (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314(5801):1001–1004.
69. Woodside MT, et al. (2006) Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc Natl Acad Sci USA* 103(16): 6190–6195.
70. Hall KB (2012) Spectroscopic probes of RNA structure and dynamics. *Methods Mol Biol* 875:67–84.
71. Hall KB (2013) RNA does the folding dance of twist, turn, stack. *Proc Natl Acad Sci USA* 110(42):16706–16707.
72. Chen AA, García AE (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci USA* 110(42): 16820–16825.
73. Garcia AE, Paschek D (2008) Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. *J Am Chem Soc* 130(3):815–817.
74. Hagan MF, Dinner AR, Chandler D, Chakraborty AK (2003) Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *Proc Natl Acad Sci USA* 100(24):13922–13927.
75. Colizzi F, Bussi G (2012) RNA unwinding from reweighted pulling simulations. *J Am Chem Soc* 134(11):5173–5179.
76. Isambert H, Siggia ED (2000) Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. *Proc Natl Acad Sci USA* 97(12):6515–6520.
77. Cocco S, Marko JF, Monasson R (2003) Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur Phys J E Soft Matter* 10(2):153–161.
78. Flamm C, Fontana W, Hofacker IL, Schuster P (2000) RNA folding at elementary step resolution. *RNA* 6(3):325–338.
79. Kirmizialtin S, Pabit SA, Meisburger SP, Pollack L, Elber R (2012) RNA and its ionic cloud: Solution scattering experiments and atomically detailed simulations. *Biophys J* 102(4):819–828.
80. Kirmizialtin S, Silalahi AR, Elber R, Fenley MO (2012) The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys J* 102(4):829–838.
81. Kim T, Shapiro BA (2013) The role of salt concentration and magnesium binding in HIV-1 subtype-A and subtype-B kissing loop monomer structures. *J Biomol Struct Dyn* 31(5):495–510.
82. Tsui V, Case DA (2001) Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *J Phys Chem B* 105(45):11314–11325.
83. Hu J, Ma A, Dinner AR (2008) A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc Natl Acad Sci USA* 105(12):4615–4620.
84. Zgarbová M, Otyepka M, Šponer J, Lankaš F, Jurečka P (2014) Base pair fraying in molecular dynamics simulations of DNA and RNA. *J Chem Theory Comput* 10(8):3177–3189.
85. Pörschke D (1974) Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys Chem* 1(5):381–386.
86. Pörchke D (1976) The nature of stacking interations in polynucleotides. Molecular states in Oligo- and polyribocytidylic acids by relaxation analysis. *Biochemistry* 15(7):1495–1499.
87. Dewey TG, Turner DH (1980) Laser temperature jump study of solvent effects of poly (adenylic acid) stacking. *Biochemistry* 19(8):1681–1685.
88. Guéron M, Leroy JL (1995) Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol* 261:383–413.
89. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14(1):33–38, 27–28.
90. Foloppe N, MacKerell AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem* 21(2):86–104.
91. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic, molecules. *J Am Chem Soc* 117(19):5179–5197.
92. Fadrna E, et al. (2009) Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. *J Chem Theory Comput* 5(9):2514–2530.
93. Banas P, et al. (2010) Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J Chem Theory Comput* 6(12): 3836–3849.
94. Bergonzo C, Henriksen NM, Roe DR, Cheatham TE, 3rd (2015) Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* 21(9):1578–1590.
95. Fresch B, Remacle F (2014) Atomistic account of structural and dynamical changes induced by small binders in the double helix of a short DNA. *Phys Chem Chem Phys* 16(27):14070–14082.
96. Fresch B, Remacle F (2015) An atomistic view of DNA dynamics and its interaction with small binders: Insights from molecular dynamics and principal component analysis. *Advances in Atom and Single Molecule Machines* 7:17–33.