2-17-2021

Reaction Coordinate and Thermodynamics of Base Flipping in RNA

Lev Levintov
University of New Hampshire, Durham, Lev.Levintov@unh.edu

Sanjib Paul
New York University

Harish Vashisth
University of New Hampshire, Durham, harish.vashisth@unh.edu

Comments

This is an Open Access article published by ACS Publications in Journal of Chemical Theory and Computation in 2021, available online: https://dx.doi.org/10.1021/acs.jctc.0c01199

Recommended Citation

Lev Levintov, Sanjib Paul, and Harish Vashisth. Reaction Coordinate and Thermodynamics of Base Flipping in RNA, Journal of Chemical Theory and Computation 2021 17 (3), 1914-1921, DOI: 10.1021/acs.jctc.0c01199

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

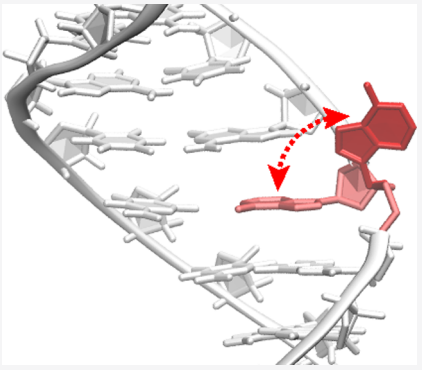
Reaction Coordinate and Thermodynamics of Base Flipping in RNA

Lev Levintov, Sanjib Paul,* and Harish Vashisth*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 1914−1921 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Base flipping is a key biophysical event involved in recognition of various ligands by ribonucleic acid (RNA) molecules. However, the mechanism of base flipping in RNA remains poorly understood, in part due to the lack of atomistic details on complex rearrangements in neighboring bases. In this work, we applied transition path sampling (TPS) methods to study base flipping in a double-stranded RNA (dsRNA) molecule that is known to interact with RNA-editing enzymes through this mechanism. We obtained an ensemble of 1000 transition trajectories to describe the base-flipping process. We used the likelihood maximization method to determine the refined reaction coordinate (RC) consisting of two collective variables (CVs), a distance and a dihedral angle between nucleotides that form stacking interactions with the flipping base. The free energy profile projected along the refined RC revealed three minima, two corresponding to the initial and final states and one for a metastable state. We suggest that the metastable state likely represents a wobbled conformation of nucleobases observed in NMR studies that is often characterized as the flipped state. The analyses of reactive trajectories further revealed that the base flipping is coupled to a global conformational change in a stem-loop of dsRNA.



1. INTRODUCTION

Interactions between nucleic acids and proteins play an essential role in various cellular processes including post-transcriptional modifications,[1−3] repair mechanisms,[4,5] and replication.[6,7] Some proteins bind to nucleic acids without introducing significant structural changes, but in other cases, binding is associated with large distortions in the structures of nucleic acids. Among other examples are enzymes that bind to nucleic acids upon opening of a specific base pair to perform a chemical reaction on the target base.[8,9] It means that the bases involved in chemical modifications have to be accessible to enzymes, preferably in a flipped out (extrahelical) state. However, it remains unclear whether base flipping occurs spontaneously or not.[10,11] Therefore, resolving atomistic details of a base-flipping event remains a fundamental problem of interest in biophysics of nucleic acids.

Studying spontaneous base flipping is a challenging process both for experimental and computational methods due to a lower likelihood of observation of flipping in a single nucleobase in otherwise stable structures of nucleic acids. On the experimental side, nuclear magnetic resonance (NMR) spectroscopy has become the leading method to study dynamics in nucleic acids due to its ability to probe fluctuations at the level of individual nucleobases.[12−14] It has been applied to study base flipping,[15−17] along with other experimental techniques including X-ray crystallography,[8] fluorescence-based assays,[18,19] melting point studies,[20] and combined approaches.[21,22] In DNA, NMR studies have shown that the lifetime of the extrahelical state of a base can be on the order of μs and that the lifetime of the intrahelical state can be in the range of *ms* depending on the stability of individual bases.[23,24] Additionally, several studies revealed that the base becomes accessible to the solvent for NMR detection when the base pair opens to a pseudodihedral angle of at least 30°, thereby indicating that the bases are still within the cutoff of a hydrogen bond formation.[25,26] Therefore, the fluctuations measured by NMR may need to be reassigned to base wobbling as opposed to flipping, and the mechanistic understanding may not be directly applicable to a base-flipping process.[22] Thus, despite key mechanistic information emerging from the application of NMR methods, there remains the need for additional analyses at the atomic level for obtaining further insights into this molecular mechanism.

On the computational side, due to limitations in conformational sampling by conventional molecular dynamics (MD) simulations, enhanced sampling methods have been applied to probe this event.[22,25−33] Among previous studies of base flipping, some have used external forces to induce base-flipping transitions,[25,29] which likely leads to a loss of critical information on key variables that may contribute to base flipping. Enhanced sampling methods also rely on the definition of an appropriate reaction coordinate (RC) which is a single variable to discriminate between a given pair of

Received: November 15, 2020
Published: February 17, 2021

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

stable states and using which key thermodynamic (e.g., free energy) properties can be computed. Although establishing an appropriate RC is challenging,[34] once it is identified, the multidimensional free energy surface can be reduced to a one-dimensional profile along the RC to obtain crucial mechanistic insights into the transition mechanism.

Many computational methods have been applied to study nucleobase stacking/unstacking in nucleic acids.[27,28,31,35−39] Several significant studies have been conducted to study the base-flipping process in DNA in association with protein binding.[40,41] Several of these studies explored simplified systems that consisted only up to three base pairs and may be limited in describing the dynamics in a larger RNA system with many base pairs.[35−37] Additionally, several previous studies were reported over a decade ago, and the force fields for nucleic acids have significantly improved in recent years.[42,43] Moreover, the candidate variables that potentially contribute to RC have not been examined systematically. Therefore, the application of simulation methods that permit systematic testing of a suitable RC is needed to improve our understanding of the mechanism of base flipping in nucleic acids.

Such techniques include the method of transition path sampling (TPS),[44−46] which has been successfully applied to study the flipping of a terminal pyrimidine base in a short DNA chain with three base pairs.[28] While most previous studies have focused on DNA due to its structural stability, we study RNA as a model system given its conformational flexibility and emerging importance in drug discovery.[47] TPS has also been applied to explore other biophysical problems including folding,[48−50] flipping of amino acids in enzymes,[51] DNA synthesis,[52] water dynamics,[53] catalysis,[54−56] nucleation,[57,58] and chemical reactions.[59] In this work, we applied TPS[44−46] simulations and the likelihood maximization methods[60] to study the base-flipping mechanism in a dsRNA molecule which has a nucleobase that can flip out (Figure 1).

2. MODEL AND METHODS

2.1. System Preparation and Simulation Details. The initial coordinates for dsRNA were obtained from the first frame of the NMR structure (PDB code: 2L2K).[16] The system was solvated in a 72 Å × 72 Å × 83 Å periodic box of TIP3P water molecules and was comprised of 39305 atoms (Figure S1). The system was neutralized with 21 Mg^{2+} ions. The temperature and pressure were maintained at 310 K and 1 atm using the Langevin thermostat and the Berendsen barostat, respectively. All MD simulations were carried out using the Amber[61] software combined with the recent RNA Amber force field developed by a Rochester group (RNA.ROC).[62] The analyses of all trajectories were carried out using the CPPTRAJ module in Amber and using the Visual Molecular Dynamics (VMD) software.[63,64]

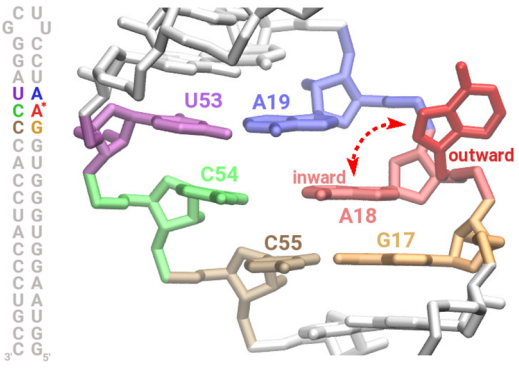


Figure 1. Sequence and structural details of bases involved in conformational transitions. (*left*) Secondary structure of dsRNA with key nucleotides highlighted. The nucleobase studied in this work is marked with an asterisk. (*right*) A side view of the three-dimensional structure of dsRNA where each key nucleotide is highlighted in a unique color and labeled. Specifically, A18 is shown both in the flipped in and out conformations. The solvated simulation domain is shown in Figure S1.

2.2. Transition Path Sampling. TPS[44,45,65] is a method to generate an ensemble of transition paths that connect a pair of initial (reactant) and final (product) states that are separated by a free energy barrier. An important step in TPS is to obtain an initial reactive trajectory (seed trajectory) that connects two well-defined stable states (e.g., the inward and outward conformations of a base in this work). The seed trajectory thus generated is further utilized to obtain the transition path ensemble by launching unbiased MD simulations (shooting trajectories) from the shooting region that is located close to the transition region. The seed trajectory does not have to be a true dynamical pathway because it will eventually reach the equilibrium transition path ensemble (TPE) by repeatedly launching successive shooting trajectories.[44,65] The shooting trajectories have a high probability of rapidly relaxing to one of the stable basins because the aimless shooting algorithm that we have used generates shooting points near the barrier region by sampling momenta afresh from the Boltzmann distribution.[60,66]

2.3. Seed Trajectory and Definitions of Stable States. We conducted four conventional MD simulations, each 150 ns long, to obtain a seed trajectory. Only one out of these four simulations exhibited a spontaneous base-flipping event (Figure S2), and therefore, it was used as our seed trajectory for building an ensemble of transition paths. For these MD simulations, we first performed 1000 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization. Then, we conducted all MD simulations in the NPT ensemble using a 2 fs time step and saved configurations every 10 ps. Prior to launching shooting simulations, we defined an order parameter (OP) that can unambiguously discriminate between the two stable states, the inward (*I*, state 1) and the outward (*O*, state 2) states, and determined the ranges of the OP for defining two stable states. These ranges were chosen to clearly separate stable basins, accommodate system fluctuations, and prevent sampling of nonreactive trajectories.[44,48,65] We selected a pseudodihedral angle as our OP that is defined by the centers of mass of four groups of atoms (Figure 2A) which has been previously identified as a potential collective variable (CV) for this system.[38] For the configuration *I*, the range of the OP was defined as $-70° < OP_1 < 70°$, and for the configuration *O*, the range of the OP was defined as $100° < OP_2 < 180°$ and $-120° < OP_2 < -180°$. From the shooting region identified in our seed trajectory, we then launched 1000 shooting trajectories, each 1 ns long. We carried out all shooting simulations in the NPT ensemble using a 2 fs time step and saved configurations every 0.25 ps. Based on the definition of our OP, 748 of them terminated in state *I*, and 252 terminated in state *O*.[60,66]

2.4. List of Collective Variables. In addition to the primary OP, a list of other potential CVs was created and monitored in all shooting trajectories. The distributions of the

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

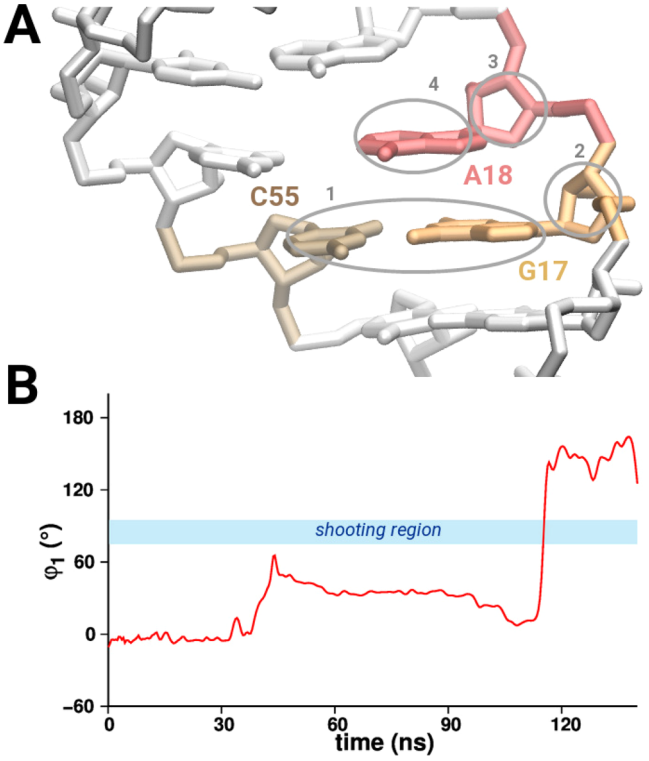


Figure 2. Details on the primary order parameter. (A) The OP is defined by the center of mass of each of the following four groups: the nitrogenous bases of C55 and G17 (labeled 1), the sugar moiety attached to G17 (labeled 2), the sugar moiety attached to A18 (labeled 3), and the nitrogenous base of A18 (labeled 4). Each key nucleotide is also uniquely colored and labeled. (B) Shown is a time trace of the primary OP in the seed trajectory (red). A cyan rectangle highlights the shooting region. See also Figure S2.

values of each of the CVs at terminal points in shooting trajectories were examined to select those CVs to be included in the construction of RC that discriminated between the I/O states. Each of the CVs was then normalized according to the following expression

$$q_k = \frac{1}{\sigma_k}(X_k - \langle X_k \rangle) \tag{1}$$

where $\langle X_k \rangle$ and σ_k, respectively, represent the mean and standard deviation of X_k (CV), for all shooting points. The normalized variable has a mean of 0 and a standard deviation of 1. Thus, $\mathbf{q}(\{q_k\})$ represents a set of CVs to be tested in construction of the refined RC.

The 12 CVs (k = 1, 2,···, 12) that were identified (Table S1) are as follows:

1. ϕ_1: the pseudodihedral angle that describes the position of A18 relative to G17. It also served as our primary OP.
2. ϕ_2: the pseudodihedral angle that describes the position of A18 relative to A19.
3. d_1: the distance between the centers of mass of G17 and A18.
4. d_2: the distance between the centers of mass of A18 and A19.
5. d_3: the hydrogen bond distance between the N_1 atom of A18 and the N_3 atom of C54.
6. α_1: the angle between A18 and C28 defined using the following three atoms: N_9 and C'_1 of A18 and C'_1 of C54.
7. α_2: the interplane angle between G17 and A18. Only heavy atoms were used to define the plane.
8. α_3: the interplane angle between A18 and A19. Only heavy atoms were used to define the plane.
9. N_W: the number of water molecules within 8 Å of A18.
10. E_1: the stacking energy between bases G17 and A18.
11. E_3: the interaction energy between bases A18 and C54.
12. E_2: the stacking energy between bases A18 and A19.

2.5. Refined Reaction Coordinate. The RC is defined as a linear combination of the identified and normalized CVs as

$$r(\{\mathbf{q}\}) = a_0 + \sum_{k=1}^{m} a_k q_k \tag{2}$$

where m is the number of OPs and is less than or equal to the total number of identified CVs, and a_k's are adjustable parameters. We applied the likelihood maximization method[60,66] to find the best set of CVs and associated a_k's that are chosen to maximize the likelihood $\ln(L)$ and have the committor function p_O defined as the probability that a transition path, initiated from a shooting point, commits to the product state, O. Per aimless shooting algorithm and likelihood maximization methods, the committor is modeled as

$$p_O(r) = \frac{1}{2}[1 + \tanh(r)] \tag{3}$$

and L is defined as

$$L = \prod_{x_k \to O} p_O(r(\mathbf{q})) \prod_{x_k \to I} [1 - p_O(r(\mathbf{q}))] \tag{4}$$

The products over $x_k \to O$ and $x_k \to I$ represent the product over all shooting points x_k committed to states O and I. By varying m in eq 2, different models of the RC were investigated (Tables S2−S4). For each model, a_k's were determined for each combination of q_k's by maximizing $\ln(L)$. The parameter a_0 was adjusted so that the transition between states I and O appears at r = 0.

The models of the RC with the same number of OPs (m = n) were then compared against each other using the maximum likelihood scores to pick the best combination of CVs. The best model with n parameters was then compared against the best model with n+1 parameters, and the significance of the addition of an extra CV was evaluated using the Bayesian information criterion (BIC),[60] which determines when additional complexity of the model shows no further improvement or increased significance because an extra parameter in the model is significant only if the likelihood increases by a value larger than the value set by BIC.[60] The BIC was applied using the following expression

$$\mathrm{BIC} = \frac{1}{2}\ln(N_{shoot}) \tag{5}$$

where N_{shoot} is the total number of shooting points that we have generated. Based on 1000 shooting trajectories that we generated and by using N_{shoot} = 1000, the BIC was computed to be 3.45 using eq 5. If $\ln(L)$ does not change more than this number on increasing model complexity, adding another parameter to RC is not considered significant.

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

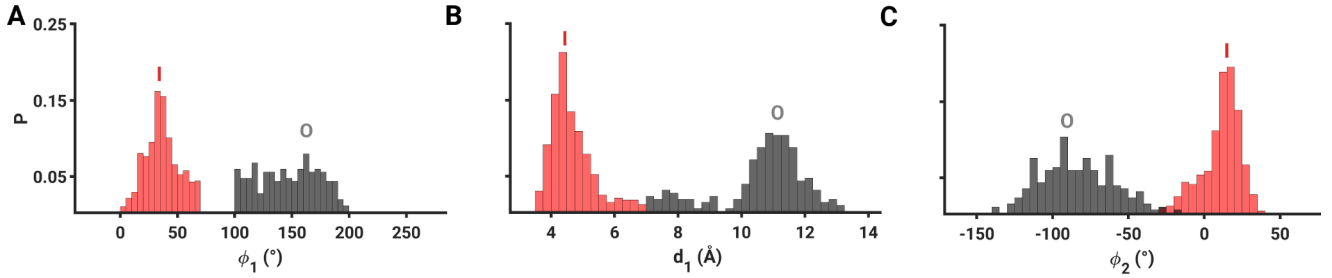


Figure 3. Population distributions of CVs at terminal points. Shown are the distributions of CVs at terminal points of transition paths for the inward (red) and outward (gray) states. (A) The pseudodihedral angle (ϕ_1) that describes the position of A18 relative to G17. (B) The distance (d_1) between the centers of mass of G17 and A18. (C) The pseudodihedral angle (ϕ_2) that describes the position of A18 relative to A19. See also Figure S3.

2.6. Free Energy Profile along RC. The potential of mean force (PMF)/free energy profile was obtained along the RC using

$$G(r) = -k_B T \ln P(r) \quad (6)$$

where k_B is Boltzmann's constant, T is the temperature, and $P(r)$ is the histogrammed population. Per Peters et al.,[60] if an RC is optimized using shooting points from TPS simulations, then the resulting RC based on the transition path ensemble is also a good RC in the equilibrium ensemble, thereby permitting eq 6 for obtaining the PMF.

3. RESULTS AND DISCUSSION

3.1. Fluctuations of the OP (ϕ_1). After defining a list of all potential CVs that can be used to describe the transition between the I and O configurations, we picked the pseudodihedral angle (ϕ_1) between G17 and A18 as our primary OP, which was used to identify the seed trajectory (Figure 2). A time trace of the OP in the seed trajectory is shown in Figure 2B, and the distribution of its values at terminal points in the transition paths is shown in Figure 3A. We defined the shooting region as the range between 75° and 95° and observed a transition in this region (at ∼117 ns) in one of four conventional MD simulations (Figure 2B). Three other conventional MD simulations that were launched from the same initial structure did not exhibit base flipping (Figure S2). The configurations from the shooting region in the seed trajectory were then used as input structures to build an ensemble of 1000 shooting trajectories.

3.2. Identification of Other Potential CVs. All of the predefined CVs (Table S1) were monitored at the terminal points of each shooting trajectory for their suitability in discriminating between two states and thus for inclusion in construction of the RC. The population distributions of all tested CVs at terminal points in shooting trajectories are shown in Figures 3 and S3. The CVs in Figure 3 (ϕ_1, d_1, and ϕ_2) were found to be the most important for constructing a model of the refined RC because these CVs exhibited distinct bimodal distributions at the terminal points where one peak was more populated at the inward state and the other was more populated at the outward state (Figure 3). These variables collectively describe the relative position of the nucleobase A18 with respect to G17 and A19 with which A18 forms stacking interactions (Figure 1). Since ϕ_1 and d_1 likely provide similar information, we anticipated that one of them may be omitted from the refined RC. Many CVs shown in Figure S3 exhibited overlapping distributions between the two states with several CVs showing larger overlaps in their distributions (e.g., Figure S3G,I) due to which of these CVs were not included in the refined RC.

3.3. Refined Reaction Coordinate (r). The refined RC was then determined using the likelihood maximization and the BIC[60,66] by testing various models of increasing complexity constructed from 12 CVs (Tables S1−S4). We found that the refined RC was a linear combination of 2 CVs, d_1 and ϕ_2. The final equation for the refined RC is

$$r = -0.84 + 0.4753 d_1 - 0.3941 \phi_2 \quad (7)$$

The addition of a third variable to the RC improves it, but according to the likelihood maximization tests, the improvement is not significant (Table S4), meaning that 2 CVs are sufficient to formulate the refined RC. This can be seen in Figure S4A which shows the histograms of the RC values across all shooting trajectories and the resulting free energy profiles in Figure S4B. These data show that, for the three-variable RC models, the free energy profiles are similar to the refined RC (Figure S4B). The evolution of the refined RC in the transition paths is shown in Figures 4A and S5. The time evolution of the RC further confirms its validity by showing that the trajectories initiated from the region near $r = 0$ terminated in one of the two stable states, and the RC is divided into two segments of the configuration space by terminating either at the inward state or at the outward state.

3.4. Free Energy Profile. The potential of mean force (PMF) profile (Figure 4B) was estimated based on the population distribution of the refined RC computed across all transition paths (Figure S6). The PMF profile exhibited three minima corresponding to the inward (I) state ($-1.65 < r < -1.56$), a metastable (M) state ($-0.87 < r < -0.75$), and the outward (O) state ($0.63 < r < 0.81$). The transition state is represented by $r = 0$. The activation free energy ($\Delta G^{\ddagger}$) and the free energy difference (ΔG) between states I and O were determined to be 1.48 and 1.0 kcal/mol, respectively. Based on the free energy profile, the outward state is less stable than the inward state which could be important for the deamination process performed by the ADAR2 enzyme.[16] Importantly, the metastable state represented the wobbling movement of A18 when the nucleotide is partially flipped out with $\phi_1 \sim 50°-65°$. We suggest that the metastable states of these types are likely observed by NMR and mischaracterized as the flipped out (outward/extrahelical) states.[16,25] Our final RC also discriminates between the metastable (wobbled) and the flipped out (outward) states.

3.5. Conformational Properties of dsRNA in the Transition Path Ensemble. We launched shooting trajectories from the shooting region that is located close to the

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

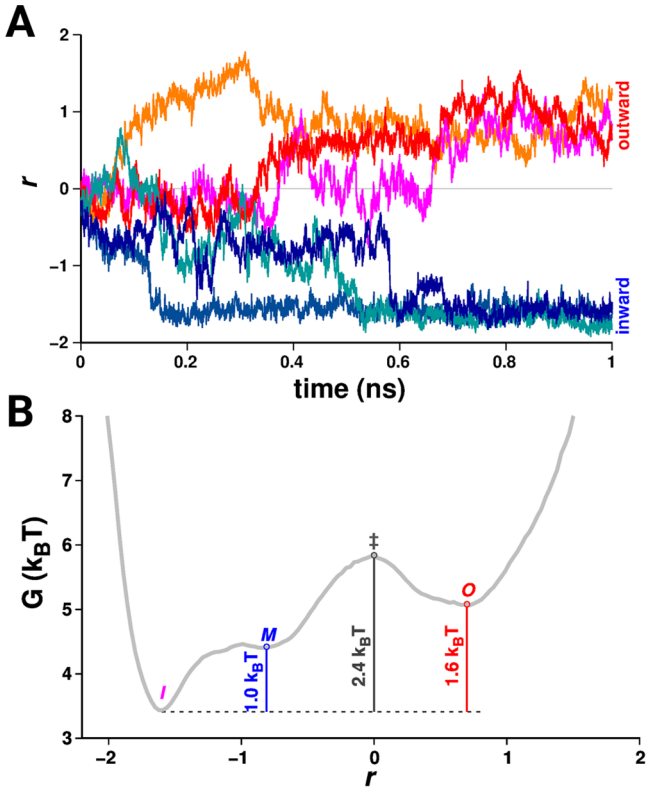


Figure 4. Evolution of the refined RC and the potential of mean force (PMF) profile. (A) The evolution of the RC along representative trajectories. See also Figure S5. (B) PMF as a function of the RC. Three vertical lines mark the free energy difference between the inward (labeled *I*) and metastable (labeled *M*) states (blue), the activation energy (dark gray; labeled ‡), and the energy difference between the inward *I* and outward (labeled *O*) states (red).

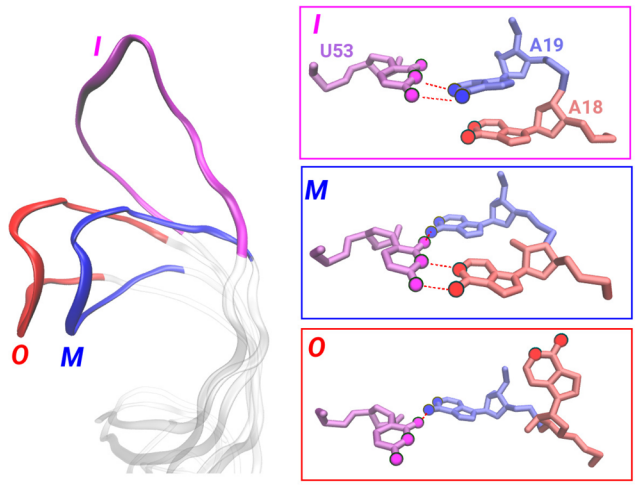


Figure 5. Global and local conformational dynamics in dsRNA. (*left*) Snapshot of global conformational changes in the RNA stem-loop derived from shooting trajectories at three different states: *I* (magenta), *M* (blue), and *O* (red). (*right*) Snapshots of the flipping site in three different states. Each key nucleotide and atoms that participate in hydrogen bonding (marked by dotted red lines) are uniquely colored.

transition region, and trajectories landed either in the reactant state (*I* and *M*) or in the product state (*O*). Based on the OP, 748 trajectories terminated in the *I* state, and 252 trajectories terminated in the *O* state. The conformations of bases in dsRNA at state *I* and the shooting region are shown in Figure S7A,B. In the shooting region, the positions of A18 and U53 are perturbed compared to the initial (*I*) conformation by ∼75−95° (ϕ_1) and ∼30−40° (using the flipping angle definition similar to ϕ_1), respectively. We observed that the flipping motion of A18 resulted not only in rearrangements in neighboring bases but also in a conformational change in a stem-loop of dsRNA (the loop highlighted in magenta/blue/red in Figure 5). Below, we discuss how the flipping of A18 affected the conformation of the stem-loop, motion of nucleotides, and hydrogen bonds between various bases in three different states (*I*, *M*, and *O*).

State *I*: During inward flipping of A18 from the transition barrier region, A18 and A19 formed the base pairs with C54 and U53, respectively, and the RNA stem-loop had an elongated conformation (*I*; Figure 5). The formation of a base pair between A19 and U53 was measured via a hydrogen bond distance of 3.5 Å between the O4 atom of U53 and the N6 atom of A19 (Figure S8A). Concomitantly, C54 partially flipped out by ∼55−60° to provide space for A18 to flip back in (Figure S8B and *I*; Figure 5). The flipping of C54 outward (Figure S8A) was not observed in the initial configuration (Figure 1).

State *M*: When A18 was in the *M* state (i.e., wobbled conformation), the RNA stem-loop was in a bent conformation relative to state *I* (*M*; Figure 5). This conformation resulted due to the interactions between a triplet of bases: A18, A19, and U53 (*M*; Figure 5). A18 formed a hydrogen bond (3.5 Å long) with U53, which was partially flipped out at the transition barrier (Figure S8C). At the same time, the initial hydrogen bond between the O4 atom of U53 and the N6 atom of A19 broke, and a new hydrogen bond formed between the O2 atom of U53 and the N6 atom of A19 (Figure S8A,D). As a result of these rearrangements, U53 formed hydrogen bonds with both A18 and A19, thus creating a triplet, which caused the RNA stem-loop to bend (*M*; Figure 5).

State *O*: In shooting trajectories that resulted in outward flipping of A18, the RNA stem-loop was also observed to undergo a bent conformation (*O*; Figure 5). Similar to the *M* state, U53 partially flipped out and disrupted the initial hydrogen bond with A19 and formed another bond between the O2 atom of U53 and the N6 atom of A19 (Figure S8D). A18 did not form any interactions with U53, but the flipping of A18 outward likely perturbed U53 and caused U53 to partially flip out. Thus, even a minor conformational change in A18 by ∼45° (*M* state) caused local rearrangements in U53 while breaking the initial hydrogen bonds with A19 which, in turn, resulted in a bent conformation of the RNA stem-loop. Overall, our mechanistic analyses of conformations of bases in the transition path ensemble revealed that the flipping of a single base (A18) in RNA is coupled not only with rearrangements in local bases but also global conformational changes in common motifs (e.g., stem-loops) found in nucleic acids.

3.6. Comparison to Previous Work. We note that Hart et al.[38] have previously studied this base-flipping event by focusing on ϕ_1 as their hypothesized RC. However, our search for the refined RC is systematic and exhaustive since we have examined a large number of CVs and their combinations using the likelihood maximization method.[60,66] The ensemble of trajectories that we have generated (totaling over 1000 ns)

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

exceeded what was used in the previous work (14.4 ns) which further helped us in identifying a refined RC. Importantly, our results showed that, for a single-variable RC model, ϕ_2 is a more important CV than ϕ_1 because ϕ_2 was ranked second, while ϕ_1 was ranked 11th (Table S2). Additionally, we estimated that a two-variable RC model is more significant for capturing the base-flipping process than a single-variable model, whether it consisted of ϕ_1 or any other variable (Table S3). In fact, even a three-variable RC model did not indicate that ϕ_1 was the most important CV out of the remaining CVs for model improvement since the model with ϕ_1 was ranked fourth (Table S4). Moreover, we also observed that the local base-flipping event in dsRNA is coupled with a global conformational transition in the stem-loop of this dsRNA. In the previous work,[38] only local rearrangements of A18 and the neighboring bases were reported, but in our work, we showed that even a partial flipping of A18 caused the stem-loop of dsRNA to bend. We also revealed that in the metastable state, U53 forms a base triplet with A18 and A19 through hydrogen bonding interactions (Figure 5) which has not been reported previously. We also note that significant work has been done to study the base-flipping process in DNA systems where several RCs have been tested using umbrella sampling simulations[32] and TPS.[28] The base flipping has also been investigated in isolated nucleotides in DNA,[35] modified DNA bases,[33] and in DNA/enzyme systems.[40,41]

4. CONCLUSIONS

Using transition path sampling combined with the likelihood maximization methods, we developed a refined reaction coordinate (RC) to describe the base-flipping mechanism in dsRNA. The refined RC is comprised of two CVs that collectively describe the relative position of the flipping base with respect to the neighboring bases, thereby showing an improved description of the base-flipping mechanism. Outside of conformational variables, we did not observe any significant improvements in our RC models on including the solvent molecules or stacking energies between the bases. However, a further examination of these coordinates may be needed for other RNA motifs (e.g., bulges) if the flipping nucleotides are not involved in base-pairing interactions unlike the system studied in this work. Our results emphasize the importance of systematic examination of CVs in constructing RC models of complex biophysical processes. We also observed that the flipping of a single base caused local rearrangements in the neighboring bases which then resulted in global structural transitions in a stem-loop of the dsRNA. We suggest that the approaches described in this work are potentially applicable to other RNA/ligand systems, for example, conformational transitions coupled to binding of a ligand molecule in an RNA element from HIV-1, as reported in our previous work.[67]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c01199.

Table S1, list of relevant CVs; Table S2, likelihood scores for single-variable reaction coordinate models; Table S3, two-variable reaction coordinate models with highest likelihood scores; Table S4, three-variable reaction coordinate models with highest likelihood scores; Figure S1, system setup; Figure S2, time traces of primary OP; Figure S3, population distributions of CVs at terminal points; Figure S4, key metrics of refined RC along with top 5 three-variable RC models; Figure S5, evolution of RC along additional representative trajectories; Figure S6, population distribution of refined RC; Figure S7, snapshots of dsRNA; and Figure S8, population distributions of several physical variables from transition path ensemble (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sanjib Paul − *Department of Chemistry, New York University, New York 10003, United States*; Email: sp6310@nyu.edu

Harish Vashisth − *Department of Chemical Engineering, University of New Hampshire, Durham 03824, New Hampshire, United States*; orcid.org/0000-0002-2087-2880; Email: harish.vashisth@unh.edu

Author

Lev Levintov − *Department of Chemical Engineering, University of New Hampshire, Durham 03824, New Hampshire, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c01199

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the financial support from the National Science Foundation (CBET-1554558) and computational support through the following resources: Premise, a central shared HPC cluster at UNH supported by the Research Computing Center; BioMade, a heterogeneous CPU/GPU cluster supported by the NSF EPSCoR award (OIA-1757371); and the NSF-supported (ACI-1548562) Extreme Science and Engineering Discovery Environment (XSEDE)[68] Comet resource at the San Diego Supercomputer Center (SDSC) under grant TG-MCB160183 (H.V.).

■ REFERENCES

(1) Nachtergaele, S.; He, C. The Emerging Biology of RNA Post-Transcriptional Modifications. *RNA Biol.* **2017**, *14*, 156−163.

(2) Bass, B. L. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu. Rev. Biochem.* **2002**, *71*, 817−846.

(3) Nishikura, K. Editor Meets Silencer: Crosstalk Between RNA Editing and RNA Interference. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 919−931.

(4) Roberts, R. J.; Cheng, X. Base Flipping. *Annu. Rev. Biochem.* **1998**, *67*, 181−198.

(5) Yang, C.-G.; Yi, C.; Duguid, E. M.; Sullivan, C. T.; Jian, X.; Rice, P. A.; He, C. Crystal Structures of DNA/RNA Repair Enzymes AlkB and ABH2 Bound to dsDNA. *Nature* **2008**, *452*, 961−965.

(6) Spies, M.; Smith, B. O. Protein-Nucleic Acids Interactions: New Ways of Connecting Structure, Dynamics and Function. *Biophys. Rev.* **2017**, *9*, 289−291.

(7) Lin, J. L. J.; Wu, C.-C.; Yang, W.-Z.; Yuan, H. S. Crystal Structure of Endonuclease G in Complex with DNA Reveals How it Nonspecifically Degrades DNA as a Homodimer. *Nucleic Acids Res.* **2016**, *44*, 10480−10490.

(8) Alian, A.; Lee, T. T.; Griner, S. L.; Stroud, R. M.; Finer-Moore, J. Structure of a TrmA-RNA Complex: a Consensus RNA Fold Contributes to Substrate Selectivity and Catalysis in m^5U

Methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6876−6881.
(9) Hong, S.; Cheng, X. DNA Base Flipping: a General Mechanism for Writing, Reading, and Erasing DNA Modifications. *Adv. Exp. Med. Biol.* **2016**, *945*, 321−341.
(10) Blainey, P. C.; van Oijen, A. M.; Banerjee, A.; Verdine, G. L.; Xie, X. S. A Base-Excision DNA-Repair Protein Finds Intrahelical Lesion Bases by Fast Sliding in Contact with DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5752−5757.
(11) Chen, Y. Z.; Mohan, V.; Griffey, R. H. Spontaneous Base Flipping in DNA and Its Possible Role in Methyltransferase Binding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *62*, 1133−1137.
(12) Salmon, L.; Yang, S.; Al-Hashimi, H. M. Advances in the Determination of Nucleic Acid Conformational Ensembles. *Annu. Rev. Phys. Chem.* **2014**, *65*, 293−316.
(13) Salmon, L.; Giambaşu, G. M.; Nikolova, E. N.; Petzold, K.; Bhattacharya, A.; Case, D. A.; Al-Hashimi, H. M. Modulating RNA Alignment Using Directional Dynamic Kinks: Application in Determining an Atomic-Resolution Ensemble for a Hairpin Using NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, *137*, 12954−12965.
(14) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The Roles of Structural Dynamics in the Cellular Functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 474−489.
(15) Cao, C.; Jiang, Y. L.; Stivers, J. T.; Song, F. Dynamic Opening of DNA During the Enzymatic Search for a Damaged Base. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1230−1236.
(16) Stefl, R.; Oberstrass, F. C.; Hood, J. L.; Jourdan, M.; Zimmermann, M.; Skrisovska, L.; Maris, C.; Peng, L.; Hofr, C.; Emeson, R. B.; Allain, F. H.-T. The Solution Structure of the ADAR2 dsRBM-RNA Complex Reveals a Sequence-Specific Readout of the Minor Groove. *Cell* **2010**, *143*, 225−237.
(17) Rangadurai, A.; Szymanski, E. S.; Kimsey, I.; Shi, H.; Al-Hashimi, H. M. Probing Conformational Transitions Towards Mutagenic Watson-Crick-Like G·T Mismatches Using Off-Resonance Sugar Carbon $R_{1\rho}$ Relaxation Dispersion. *J. Biomol. NMR* **2020**, *74*, 457−471.
(18) Altan-Bonnet, G.; Libchaber, A.; Krichevsky, O. Bubble Dynamics in Double-Stranded DNA. *Phys. Rev. Lett.* **2003**, *90*, 138101.
(19) Chen, X.; Zhou, Y.; Qu, P.; Zhao, X. S. Base-by-Base Dynamics in DNA Hybridization Probed by Fluorescence Correlation Spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 16947−16952.
(20) Spies, M. A.; Schowen, R. L. The Trapping of a Spontaneously "Flipped-out" Base from Double Helical Nucleic Acids by Host-Guest Complexation with β-cyclodextrin: the Intrinsic Base-Flipping Rate Constant for DNA and RNA. *J. Am. Chem. Soc.* **2002**, *124*, 14049−14053.
(21) Colizzi, F.; Perez-Gonzalez, C.; Fritzen, R.; Levy, Y.; White, M. F.; Penedo, J. C.; Bussi, G. Asymmetric Base-Pair Opening Drives Helicase Unwinding Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 22471−22477.
(22) Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y. Q.; Zhao, X. S. Dynamics of Spontaneous Flipping of a Mismatched Base in DNA Duplex. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8043−8048.
(23) Guéron, M.; Kochoyan, M.; Leroy, J.-L. A Single Mode of DNA Base-Pair Opening Drives Imino Proton Exchange. *Nature* **1987**, *328*, 89−92.
(24) Moe, J. G.; Russu, I. M. Kinetics and Energetics of Base-Pair Opening in 5′-d(CGCGAATTCGCG)-3′ and a Substituted Dodecamer Containing G·T Mismatches. *Biochemistry* **1992**, *31*, 8421−8428.
(25) Priyakumar, U. D.; MacKerell, A. D. Computational Approaches for Investigating Base Flipping in Oligonucleotides. *Chem. Rev.* **2006**, *106*, 489−505.
(26) Várnai, P.; Canalia, M.; Leroy, J.-L. Opening Mechanism of G·T/U Pairs in DNA and RNA Duplexes: a Combined Study of Imino Proton Exchange and Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **2004**, *126*, 14659−14667.
(27) Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. Protein-Facilitated Base Flipping in DNA by Cytosine-5-Methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 68−73.
(28) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. Atomistic Understanding of Kinetic Pathways for Single Base-Pair Binding and Unbinding in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13922−13927.
(29) Bouvier, B.; Grubmüller, H. A Molecular Dynamics Study of Slow Base Flipping in DNA Using Conformational Flooding. *Biophys. J.* **2007**, *93*, 770−786.
(30) Da, L.-T.; Yu, J. Base-Flipping Dynamics from an Intrahelical to an Extrahelical State Exerted by Thymine DNA Glycosylase During DNA Repair Process. *Nucleic Acids Res.* **2018**, *46*, 5410−5425.
(31) Kingsland, A.; Maibaum, L. DNA Base Pair Mismatches Induce Structural Changes and Alter the Free-Energy Landscape of Base Flip. *J. Phys. Chem. B* **2018**, *122*, 12251−12259.
(32) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *J. Chem. Theory Comput.* **2009**, *5*, 3105−3113.
(33) Li, H.; Endutkin, A. V.; Bergonzo, C.; Fu, L.; Grollman, A.; Zharkov, D. O.; Simmerling, C. DNA Deformation-Coupled Recognition of 8-Oxoguanine: Conformational Kinetic Gating in Human DNA Glycosylase. *J. Am. Chem. Soc.* **2017**, *139*, 2682−2692.
(34) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016**, *67*, 669−690.
(35) Norberg, J.; Nilsson, L. Conformational Free Energy Landscape of ApApA from Molecular Dynamics Simulations. *J. Phys. Chem.* **1996**, *100*, 2550−2554.
(36) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. Observation of Spontaneous Base Pair Breathing Events in the Molecular Dynamics Simulation of a Difluorotoluene-Containing DNA Oligonucleotide. *J. Am. Chem. Soc.* **1999**, *121*, 8653−8654.
(37) Brown, R. F.; Andrews, C. T.; Elcock, A. H. Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 2315−2328.
(38) Hart, K.; Nyström, B.; Öhman, M.; Nilsson, L. Molecular Dynamics Simulations and Free Energy Calculations of Base Flipping in dsRNA. *RNA* **2005**, *11*, 609−618.
(39) Häse, F.; Zacharias, M. Free Energy Analysis and Mechanism of Base Pair Stacking in Nicked DNA. *Nucleic Acids Res.* **2016**, *44*, 7100−7108.
(40) Kuznetsov, N. A.; Bergonzo, C.; Campbell, A. J.; Li, H.; Mechetin, G. V.; de los Santos, C.; Grollman, A. P.; Fedorova, O. S.; Zharkov, D. O.; Simmerling, C. Active Destabilization of Base Pairs by a DNA Glycosylase Wedge Initiates Damage Recognition. *Nucleic Acids Res.* **2015**, *43*, 272−281.
(41) Byrnes, J.; Hauser, K.; Norona, L.; Mejia, E.; Simmerling, C.; Garcia-Diaz, M. Base Flipping by MTERF1 Can Accommodate Multiple Conformations and Occurs in a Stepwise Fashion. *J. Mol. Biol.* **2016**, *428*, 2542−2556.
(42) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: a Comprehensive Overview. *Chem. Rev.* **2018**, *118*, 4177−4338.
(43) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P. Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 3288−3305.
(44) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(45) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Efficient Transition Path Sampling: Application to Lennard-Jones Cluster Rearrangements. *J. Chem. Phys.* **1998**, *108*, 9236−9245.

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921

(46) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv. Chem. Phys.* **2003**, *123*, 1−78.
(47) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-Like Small Molecules. *Nat. Rev. Drug Discovery* **2018**, *17*, 547−558.
(48) Bolhuis, P. G. Transition-Path Sampling of β-Hairpin Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12129−12134.
(49) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the Transition Coordinate for Protein Folding. *J. Chem. Phys.* **1998**, *108*, 334−350.
(50) Freitas, F. C.; Lima, A. N.; de Godoi Contessoto, V.; Whitford, P. C.; de Oliveira, R. J. Drift-Diffusion (DrDiff) Framework Determines Kinetics and Thermodynamics of Two-State Folding Trajectory and Tunes Diffusion Models. *J. Chem. Phys.* **2019**, *151*, 114106.
(51) Paul, S.; Taraphder, S. Determination of the Reaction Coordinate for a Key Conformational Fluctuation in Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2015**, *119*, 11403−11415.
(52) Radhakrishnan, R.; Schlick, T. Orchestration of Cooperative Events in DNA Synthesis and Repair Mechanism Unraveled by Transition Path Sampling of DNA Polymerase β's Closing. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 5970−5975.
(53) Xi, L.; Shah, M.; Trout, B. L. Hopping of Water in a Glassy Polymer Studied via Transition Path Sampling and Likelihood Maximization. *J. Phys. Chem. B* **2013**, *117*, 3634−3647.
(54) Quaytman, S. L.; Schwartz, S. D. Reaction Coordinate of an Enzymatic Reaction Revealed by Transition Path Sampling. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12253−12258.
(55) Paul, S.; Paul, T. K.; Taraphder, S. Orthogonal Order Parameters to Model the Reaction Coordinate of an Enzyme Catalyzed Reaction. *J. Mol. Graphics Modell.* **2019**, *90*, 18−32.
(56) Paul, S.; Paul, T. K.; Taraphder, S. Reaction Coordinate, Free energy, and Rate of Intramolecular Proton Transfer in Human Carbonic Anhydrase II. *J. Phys. Chem. B* **2018**, *122*, 2851−2866.
(57) Knott, B. C.; Molinero, V.; Doherty, M. F.; Peters, B. Homogeneous Nucleation of Methane Hydrates: Unrealistic under Realistic Conditions. *J. Am. Chem. Soc.* **2012**, *134*, 19544−19547.
(58) Lupi, L.; Hudait, A.; Peters, B.; Grünwald, M.; Mullen, R. G.; Nguyen, A. H.; Molinero, V. Role of Stacking Disorder in Ice Nucleation. *Nature* **2017**, *551*, 218−222.
(59) Leitold, C.; Mundy, C. J.; Baer, M. D.; Schenter, G. K.; Peters, B. Solvent Reaction Coordinate for an S_N2 Reaction. *J. Chem. Phys.* **2020**, *153*, 024103.
(60) Peters, B.; Trout, B. L. Obtaining Reaction Coordinates by Likelihood Maximization. *J. Chem. Phys.* **2006**, *125*, 054108.
(61) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T. E., III; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. *AMBER 2018*; University of California: San Francisco, 2018.
(62) Aytenfisu, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 900−915.
(63) Roe, D. R.; Cheatham, T. E., III PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(64) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(65) Paul, S.; Nair, N. N.; Vashisth, H. Phase Space and Collective Variable Based Simulation Methods for Studies of Rare Events. *Mol. Simul.* **2019**, *45*, 1273−1284.
(66) Peters, B.; Beckham, G. T.; Trout, B. L. Extensions to the Likelihood Maximization Approach for Finding Reaction Coordinates. *J. Chem. Phys.* **2007**, *127*, 034109.
(67) Levintov, L.; Vashisth, H. Ligand Recognition in Viral RNA Necessitates Rare Conformational Transitions. *J. Phys. Chem. Lett.* **2020**, *11*, 5426−5432.
(68) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62−74.

https://dx.doi.org/10.1021/acs.jctc.0c01199
J. Chem. Theory Comput. 2021, 17, 1914−1921