

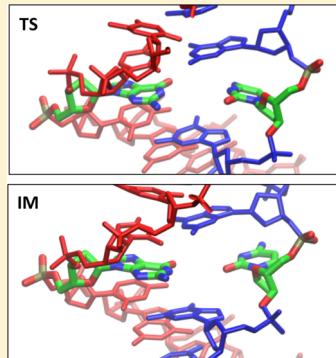
Hidden Conformation Events in DNA Base Extrusions: A Generalized-Ensemble Path Optimization and Equilibrium Simulation Study

Liaoran Cao,[†] Chao Lv,[†] and Wei Yang*,^{†,‡}

[†]Department of Chemistry and Biochemistry, and [‡]Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, United States

 Supporting Information

ABSTRACT: DNA base extrusion is a crucial component of many biomolecular processes. Elucidating how bases are selectively extruded from the interiors of double-strand DNAs is pivotal to accurately understanding and efficiently sampling this general type of conformational transitions. In this work, the on-the-path random walk (OTPRW) method, which is the first generalized-ensemble sampling scheme designed for finite-temperature-string path optimizations, was improved and applied to obtain the minimum free-energy path (MFEP) and the free-energy profile of a classical B-DNA major-groove base-extrusion pathway. Along the MFEP, an intermediate state and the corresponding transition state were located and characterized. The MFEP result suggests that a base-plane-elongation event rather than the commonly focused base-flipping event is dominant in the transition-state (TS) formation portion of the pathway; and the energetic penalty at the transition state is mainly introduced by the stretching of the Watson–Crick base pair. Moreover, to facilitate the essential base-plane-elongation dynamics, the surrounding environment of the flipped base must be intimately involved. Further taking advantage of the extended-dynamics nature of the OTPRW Hamiltonian, an equilibrium generalized ensemble simulation was performed along the optimized path; based on the collected samples, several base-flipping (opening) angle collective variables were evaluated. In correspondence with the MFEP result, the collective variable analysis result reveals that none of these commonly employed flipping (opening) angles alone can adequately represent the base-extrusion pathway, especially in the pre-TS portion. As further revealed by the collective variable analysis, the base-pairing partner of the extrusion target undergoes a series of in-plane rotations to facilitate the base-plane-elongation dynamics. A base-plane rotation angle is identified to be a possible reaction coordinate to represent these in-plane rotations. Notably, these in-plane rotation motions may play a pivotal role in determining the base-extrusion selectivity.



I. INTRODUCTION

DNA base extrusion is an important component of many biological events; this general conformational transition is a common strategy for protein machines to access to specific bases that are usually protected inside helical double-strand DNAs.^{1–12} Because of its obvious importance, base extrusion, during which base flipping is usually assumed to be the major event, has been a target of immense studies, particularly by molecular dynamics (MD) simulation methods.^{9,13–34} Despite the fact that many valuable insights have been obtained from previous investigations, key questions, such as how are base extrusions activated, which motions are responsible for the energetic costs of transition state formations, and how the base-extrusion selectivity is controlled, remain to be answered. To address these questions, it is important to describe essential events that can selectively represent the slowest dynamics of a specific base-extrusion pathway. Moreover, base extrusions may occur on time scales that are commonly intractable. Therefore, understanding base-extrusion pathways can also be instrumental to feasibly sampling this general type of conformational changes for future studies, especially on more-complex processes.

To describe DNA extrusions, a variety of collective variables have been proposed.^{16,29,32–34} A straightforward one is the base-separation distance, which can be exemplified by the distance between the pyrimidine N1 and purine N3 atoms.³² Although such base-pair separation distance can monotonically change with the progress of a base-extrusion process, it cannot selectively represent a unique path in terms of distinguishing either between the target base and its partner base-extrusion pathways or between the major and minor groove extrusion routes. To address this selectivity issue, torsion-angle-based collective variables were suggested.^{16,29,33,34} Among them, the base-opening angle by Lavery et al.³⁴ and the center-of-mass pseudo-dihedral (CPD) angle, which was originally proposed by the Mackerell group¹⁶ and further refined by Simmerling and co-workers,²⁹ are commonly applied. These reaction coordinates were designed largely based on the hypothesis that base flipping (opening) is the major event in base-extrusion processes. Notably, when a base-flipping (opening) angle collective variable was employed as the order parameter for free-energy calculations, often nontrivial simulation lengths

Received: March 12, 2013

Published: July 1, 2013

are required and calculated free-energy profiles can be very sensitive to specific base-flipping (opening) order parameter definitions and sampling lengths.^{24,29} These observations suggest that essential “hidden” events may exist in the orthogonal space perpendicular to the commonly applied collective variables; and these “hidden” events may play a crucial role in base-extrusion processes. To uncover these key “hidden” events, in this work, we seek to elucidate a B-DNA major-groove base-extrusion pathway and understand the associated essential motions.

Various approaches have been applied to map the conformational/chemical free-energy pathways of complex processes.^{20,27,35–39} To find the optimal free-energy path that represents a complex molecular process, two general strategies are often employed. One general strategy is to directly construct high-dimension free-energy surfaces along prechosen collective variables;⁴⁰ it can be achieved via the umbrella sampling (US) method,⁴¹ the blue-moon-ensemble thermodynamics integration method,⁴² the nonequilibrium steered molecular dynamics (SMD) method,⁴³ or generalized-ensemble (GE) sampling-based algorithms (such as the adaptive umbrella sampling method,⁴⁴ the metadynamics method,⁴⁵ the adaptive biasing force (ABF) method,^{46,47} the adaptive biased molecular dynamics (ABMD) method,⁴⁸ or the orthogonal space sampling methods^{49–51}). A recent work³⁹ based on the combination of the SMD and ABMD methods is an excellent example that demonstrates the usefulness of this general strategy in understanding DNA conformational changes. Direct construction of high-dimension free-energy surfaces allows multiple free-energy pathways to be simultaneously mapped; however, with the increases of collective variable dimensions, large diffusion sampling overheads, which are used to cover event-irrelevant phase regions, can boost to the level that renders it impractical. If one unique path is targeted, as an alternative strategy, chain-of-states path optimization approaches,^{52–57} among which the finite-temperature-string (FTS) method^{56,58–63} is of particular interest, allows sampling to be concentrated on phase regions near the target pathway. By nature, FTS path optimizations can be computationally demanding. To improve FTS calculation sampling efficiency, in our previous work,⁶¹ the first GE sampling-based FTS algorithm was developed; this method can naturally ensure orthogonal space structural continuity and permit more flexible usage of computing resources (for instance, with a multiwalker implementation⁶⁴). It should be noted that, different from regular GE free-energy simulation methods, by which simulated systems travel along fixed collective variable directions, GE-FTS performs random sampling along continuously updated paths. To emphasize such inherent difference, we have named the GE-based FTS path optimization scheme the “on-the-path random walk” (OTPRW) method.⁶¹

In this study, a combined OTPRW path optimization and equilibrium GE simulation strategy was used to understand a classical B-DNA major-groove base-extrusion process. To sample the target base-extrusion pathway more reliably, the OTPRW method was improved, specifically by replacing the original metadynamics kernel with an adaptive biasing force (ABF)-based kernel.^{46,47,64} Via this integration-based OTPRW (iOTPRW) approach, the MFEP and the free-energy profile of the target base-extrusion pathway were obtained simultaneously. Along the MFEP, an intermediate state and the corresponding transition state were located and characterized. The MFEP result suggests that a base-plane-elongation event

occurs before the commonly emphasized base-flipping event, the latter of which largely takes place after the formation of the transition state and the energetic penalty at the transition state is mainly introduced by the stretching motion of the Watson–Crick (WC) base pair; and to facilitate this stretching dynamics, the surrounding environment of the target base must be intimately involved. Further taking advantage of the extended-dynamics nature of the OTPRW Hamiltonian, an equilibrium GE simulation was performed along the optimized path; based on the collected samples, several collective variables were evaluated. Notably, none of the previously proposed flipping (opening)-angle collective variables alone can adequately represent the transition-state (TS) formation portion of the pathway; indeed, at the pre-TS stage, these collective variables do not take any obvious effect. The collective variable analysis also allows us to identify essential events that couple with the base-plane-elongation motion: the base-pairing partner of the extrusion target undergoes a series of in-plane rotation transitions. And an in-plane rotation angle (IPRA) can be used as a general collective variable to represent these motions. Notably, these in-plane rotation motions are likely to play a pivotal role in governing the base-extrusion selectivity.

II. THEORETICAL METHODS

II.A. A Brief Introduction to the OTPRW Sampling Method.

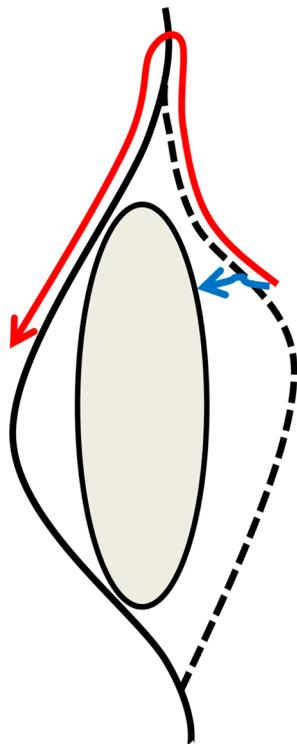
As is aforementioned, the OTPRW method is the first GE sampling strategy designed for FTS path optimizations. In FTS calculations, a set of collective variables $\theta(X) = (\theta_1(X), \dots, \theta_m(X))$ is predefined to represent the path space that can sufficiently describe a target process. Then, a string of θ -space points can be employed to depict a pathway between two end points, $Z^A = (z_1^A, \dots, z_m^A)$ and $Z^B = (z_1^B, \dots, z_m^B)$. Alternatively, a set of λ -dependent functions $Z(\lambda) = (z_1(\lambda), \dots, z_m(\lambda))$, in which $Z(0) = Z^A$ and $Z(1) = Z^B$, can be used as a continuous representation of this θ -space pathway; here, the progressing parameter λ spans from 0 to 1 to manifest the on-the-path distances from the starting point Z^A . If along a string, every single point $Z(\lambda')$ satisfies the following condition,

$$\left[M[Z(\lambda')] \frac{\partial G}{\partial \theta} [Z(\lambda')] \right]^\perp = 0 \quad (1)$$

this pathway can be called a minimum free-energy path (MFEP).^{56,58} In eq 1, M represents the diffusion tensor matrix; $(\partial G / \partial \theta) = ((\partial G / \partial \theta_1), \dots, (\partial G / \partial \theta_m))$ stands for the free-energy gradient vector; and \perp denotes the projection perpendicular to the curve. Based on eq 1, an iterative procedure of MD simulations, which are used to collect samples for $(\partial G / \partial \theta)$ calculations, and subsequent path optimization operations is generally referred as a collective-variable-space FTS algorithm.^{58–63} In FTS path optimization practices, commonly, sampling is performed on a series of independent images that are equally spaced along a to-be-updated string. Two practical issues must be noted on this original sampling strategy: (1) on-the-path structural continuity of the environmental portion (in the orthogonal space perpendicular to the path) cannot be ensured via noncommunicating samplings on independent images; (2) it is challenging for a path that represents an incorrect mechanism to be switched into the correct reaction channel. In the OTPRW sampling scheme, target molecular systems, instead of being independently constrained on noncommunicating images, are randomly propagated along to-be-updated paths to collect samples for path optimizations.⁶¹

Thereby, path sampling efficiency and robustness can be improved for the fact that orthogonal space structural continuity is naturally ensured; and certain path-space sampling that is forbidden in the original FTS sampling scheme becomes plausible (see Scheme 1).

Scheme 1. Illustration of the Path-Space Tunneling Mechanism That the OTPRW Sampling Method Enables^a



^aTwo possible pathways are depicted by the dash (the starting path) and solid (an alternative path) lines respectively. In the OTPRW sampling scheme (the red arrow), the system can evolve from the starting path to an alternative path that shares common image states; such path-space tunneling allows large energy barriers (the central hill) that are not insurmountable in the independent-image sampling scheme (the blue arrow) to be possibly bypassed.

In the OTPRW sampling design, the following extended-dynamics Hamiltonian is applied:

$$H_\lambda = H_0 + \frac{p_\lambda^2}{2m_\lambda} + \sum_{i=1}^m \frac{1}{2} K_i (\theta_i(X) - z_i(\lambda))^2 \quad (2)$$

In eq 2, H_0 represents the unmodified Hamiltonian; λ is treated as a one-dimensional (1D) dynamic particle with a mass of m_λ and its momentum is denoted as p_λ ; via the energy terms $\sum_{i=1}^m 1/2 K_i (\theta_i(X) - z_i(\lambda))^2$, the system is restrained on the immediate path described by $Z(\lambda) = (z_1(\lambda), \dots, z_m(\lambda))$. In our implementation, the λ particle is propagated based on the Langevin equation, in which the reservoir temperature is set to be the same as the system temperature. To restrain λ within the range between 0 and 1, a boundary potential (eq 3) is employed:

$$U_{\text{bound}}(\lambda) = \begin{cases} \frac{1}{2} K_{\text{bound}} \lambda^2 + f_m(0) & \lambda < 0 \\ 0 & 0 \leq \lambda \leq 1 \\ \frac{1}{2} K_{\text{bound}} (\lambda - 1)^2 + f_m(1) & \lambda > 1 \end{cases} \quad (3)$$

Considering that λ may travel out of the boundaries, when λ moves below zero, $Z(\lambda)$ is set equal to $Z(0)$; and when λ moves above one, $Z(\lambda)$ is set equal to $Z(1)$.

To realize random walks in the λ space, the on-the-path energy surface described by eq 2 must be flattened with the addition of a biasing potential $f_m(\lambda)$, the target of which is the negative of the λ -dependent free-energy profile ($-G(\lambda)$). Because $G(\lambda)$ is not known a priori and is altered upon each path optimization operation, in an OTPRW calculation, not only the path $Z(\lambda)$, but also the biasing potential $f_m(\lambda)$ should be recursively updated. In the original OTPRW implementation,⁶¹ the metadynamics method⁴⁵ was employed as the $f_m(\lambda)$ recursion kernel. As is generally known, the performance of metadynamics recursions can be sensitive to both orthogonal-space energy surface ruggedness and prechosen metadynamics parameters. Consequently, the accuracy of free-energy gradient estimations, which is essential for both path optimizations and on-the-path biasing potential updates, can be poor, particularly when instantaneous paths are not yet close to the final MFEP. In the current iOTPRW design, an ABF-like recursion strategy is employed to replace the original metadynamics-based recursion strategy, so that samples can be more robustly collected under close-to-equilibrium conditions.

II.B. The Integration-Based OTPRW Sampling Method.

The iOTPRW algorithm has two key inter-related components, respectively for adaptive updates of $f_m(\lambda)$ and $Z(\lambda)$. Correspondingly, two time intervals need to be set: ΔT_f for $f_m(\lambda)$ recursions and ΔT_Z for $Z(\lambda)$ optimizations. Usually, ΔT_Z should be much longer than ΔT_f for instance, in this study, ΔT_f was set as 0.1 ps and ΔT_Z was set as 200 ps.

The $f_m(\lambda)$ Recursion Procedure. In iOTPRW, $f_m(\lambda)$ is recursively evaluated, based on the thermodynamic integration (TI) equation:^{42,65}

$$\begin{aligned} f_m(\lambda) &= -G(\lambda) \\ &= - \int_0^\lambda \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_{\lambda'} d\lambda' \\ &= - \int_0^\lambda \sum_{i=1}^m \left\{ \frac{dz_i(\lambda)}{d\lambda} \Big|_{\lambda'} \langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_{\lambda'} \right\} d\lambda' \end{aligned} \quad (4)$$

As mentioned earlier, in OTPRW, the on-the-path Hamiltonian (eq 2) is altered upon each $Z(\lambda)$ update. Therefore, $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle$ calculations should be performed only based on samples collected within the same ΔT_Z time interval. Following the ABF approach,^{46,47,64} we can calculate $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_\lambda$, using the following equation:

$$\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_\lambda = \frac{\sum_t K_i(z_i(t) - \theta_i(X)) \delta(\lambda(t) - \lambda')}{\sum_t \delta(\lambda(t) - \lambda')} \quad (5)$$

where t denotes all the sample-collecting time steps that are within the current ΔT_Z time interval and before the current $f_m(\lambda)$ update. To avoid possible issues caused by low-precision

Table 1. The Sequence of the Target d-DNA^a

5'	G	T	C	A	G	C	G	C	A	T	G	G	3'
3'	C	A	G	T	C	G	C	G	T	A	C	C	5'

^aThe C6 base is the to-be-excluded target; and the G19 is the base-pair partner of C6.

estimation, $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_{\lambda'}$ at a state λ' is re-evaluated only when the number of samples collected for the λ' image exceeds a preset cutoff value, N_{cutoff} ; otherwise, the previous $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_{\lambda'}$ value that is utilized for last $f_m(\lambda)$ update will be reused in the current $f_m(\lambda)$ update.

Although λ is a continuous variable, for the convenience of sample collection, the λ space is linearly partitioned to a number of discrete bins; i.e., if the entire λ space is partitioned to L number of bins, the j th bin should have $(2j - 1)/(2L)$ as its bin center and sample collection boundaries of $(j - 1)/L$ and j/L . Accordingly, we can assign samples collected in the j th bin to the bin-center image $\lambda = (2j - 1)/(2L)$ for $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle_{(2j-1)/(2L)}$ calculations (eq 5). Because all of the $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle$ values are evaluated at the bin centers, using the TI formula (eq 4), we can estimate $f_m(\lambda)$ values of the bin boundary states ($0, 1/L, 2/L, \dots, 1$), based on which we can generate $f_m(\lambda)$ via the B-spline fitting method.

The $Z(\lambda)$ Optimization Procedure. In each path optimization step, a major task is to calculate the free-energy gradient vectors ($\partial G/\partial \theta$) for the bin-center states. In OTPRW, we use $\langle K_i(z_i(\lambda) - \theta_i(X)) \rangle$, which is estimated during the latest $f_m(\lambda)$ recursion step, to approximate $(\partial G/\partial \theta_i)$. It is noted that in a path optimization cycle, if the number of samples collected for an image (for instance, the j th bin) is lower than N_{cutoff} , we set $(\partial G/\partial \theta)[Z((2j - 1)/(2L))] = 0$; thereby, the images that have low numbers of samples will not be updated in the incoming optimization operation. To evolve $Z(\lambda)$, we employ the steepest descent minimization method; specifically, the point at an image center (for instance, the j th image center), $Z((2j - 1)/(2L))$, is updated to be $Z((2j - 1)/(2L)) - \{M[Z((2j - 1)/(2L))](\partial G/\partial \theta)[Z((2j - 1)/(2L))]\}^{\perp} \Delta z$, in which Δz is the path minimization step size. After the bin-center points are updated, the λ values of these newly evolved points must be recalculated to reflect the percentages of their on-the-path distances from the point Z^A , the λ value of which is always set as zero. Thereafter, we can generate the λ -dependent functions of $Z(\lambda) = (z_1(\lambda), \dots, z_m(\lambda))$ also via the B-spline fitting method.

II.C. On-the-Path Generalized Ensemble (GE) Simulation and Data Analysis. The above procedure can enable efficient convergences of $Z(\lambda)$ and $f_m(\lambda)$. When an iOTPRW reaches the convergence phase, we can fix both $Z(\lambda)$ and $f_m(\lambda)$ and carry out equilibrium GE simulations to collect samples along the optimized pathway. To obtain a higher-precision free-energy profile along λ , an on-the-path ABF simulation can be performed; in this case, only the optimized path $Z(\lambda)$ must be fixed and $f_m(\lambda)$ can be adaptively re-evaluated.

Based on the data obtained from an on-the-path equilibrium GE simulation, along any trial collective variable set \vec{s} , the free-energy landscape can be constructed based on the following equation:

$$G_0(\vec{s}) = -kT_0 \ln \sum_t \exp \left[\frac{\sum_{i=1}^m \frac{1}{2} K_i (\theta_i(X) - z_i(\lambda))^2 + f_m(\lambda(t))}{kT_0} \right] \times \delta(\vec{s}(X(t)) - \vec{s}) + \text{constant} \quad (6)$$

where k denotes the Boltzmann constant; t represents scheduled sample-collection time steps. This extended-ensemble-to-canonical-ensemble reweighting strategy has been employed in our earlier orthogonal space random walk-based simulation studies;^{50,66} it is a generalized form of the umbrella sampling formula,⁴¹ which was originally derived in the canonical-ensemble-to-canonical-ensemble reweighting framework.

III. COMPUTATIONAL DETAILS

The iOTPRW method was implemented in a customized version of the CHARMM program,^{67,68} by which all the simulations in this study were performed. The sequence of the model ds-DNA is shown in Table 1. In this well-studied B-DNA dodecamer (Figure 1), the base of Cytosine 6 (C6) is the extrusion target; the Guanine 19 (G19) is the C6's Watson–Crick (WC) partner. In the simulated system, the ds-DNA is

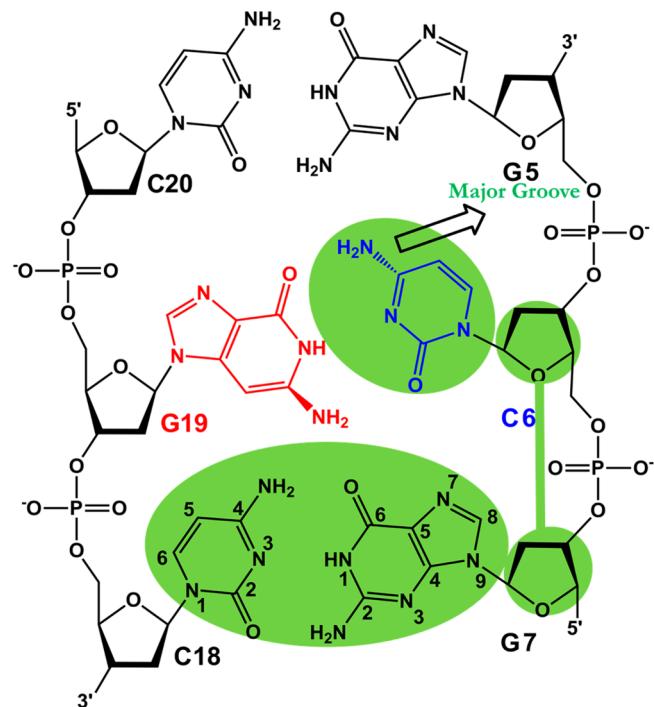


Figure 1. The chemical environment of the target base (C6) and the center-of-mass pseudo-dihedral (CPD) angle definition. The C6 base is colored in blue and the G19 base is colored in red. The major groove at the C6 position is located on the backside. Standard atom numbering of the C and G bases is shown on the G7 and C18 bases. In the CPD definition, the four atom groups are separately circled and the central pseudo-bond is drawn as a green line.

embedded in a truncated-octahedral box with 4256 water molecules, which are represented by the TIP3P model,⁶⁹ and 22 Na⁺ ions. The ds-DNA and the counterions were treated by the CHARMM27 force field.^{70,71}

III.A. The On-the-Path Random Walk (OTPRW) Path Optimization. As discussed in Section II, in a collective variable FTS calculation, a collective variable function set $\theta(X) = (\theta_1(X), \dots, \theta_m(X))$ must be prechosen to represent the path space. In this study, 10 distance collective variables were employed. As listed in Table 2, the first two distance collective

Table 2. Definition of the Collective Variable Space $\theta = (\theta_1, \dots, \theta_{10})^a$

$\theta = (\theta_1, \dots, \theta_{10})$		
the base-pair elongation	N3 ^{C6} — N1 ^{G19}	θ_1
	O2 ^{C6} — N2 ^{G19}	θ_2
the relative motions between the G19 and the neighboring base planes	N1 ^{G19} — N3 ^{C18}	θ_3
	C6 ^{G19} — C4 ^{C18}	θ_4
	N1 ^{G19} — N3 ^{C20}	θ_5
	C6 ^{G19} — C4 ^{C20}	θ_6
the neighboring base internal motions	N1 ^{G5} — N3 ^{C20}	θ_7
	N2 ^{G5} — O2 ^{C20}	θ_8
	N1 ^{G7} — N3 ^{C18}	θ_9
	N2 ^{G7} — O2 ^{C18}	θ_{10}

^aAn Ap^{Bq} naming convention is used, where symbol, “A” represents the atom type and “B” represents the base type; “p” stands for the atomic number (the numbering scheme is illustrated in Figure 1); and “q” stands for the nucleotide number. For instance, N3^{C6} denotes a nitrogen-type atom with an atomic number of 3, a C-like base type, and a nucleotide number of 6.

variables are used to describe the C6 and G19 base-separation motion; the next four distance collective variables (θ_3 – θ_6) are employed to describe relative motions between the G19 base and its neighboring base planes; and the last four distance collective variables (θ_7 – θ_{10}) are used to describe internal motions of the base planes adjacent to the C6–G19 base plane. It has been known that the 2 major order parameters, θ_1 and θ_2 , cannot distinguish specific base-extrusion pathways. In this setup, we hypothesized that the base-extrusion local environment that is represented by the remaining 8 collective variables (θ_3 – θ_{10}) allows the base-extrusion selectivity to be enforced; and this hypothesis is confirmed by the results of the path optimization calculation and the equilibrium GE simulation.

In order to generate a starting pathway, an orthogonal space tempering (OST) simulation was performed. The OST simulation was carried out with the commonly employed CPD as the order parameter; the CPD definition is shown in Figure 1. To restrict the motion along a major-groove base-extrusion pathway, the CPD value was confined between -100° and 5° . The details of the OST method can be found out in the original method paper;⁵¹ in this simulation, the orthogonal space sampling temperature was set as 1500 K. It is noted that

the purpose of this OST simulation was not to calculate a converged free-energy profile along CPD but to generate a reasonable trajectory for the construction of an initial pathway. Therefore, only a short OST trajectory that connects the two CPD end states was generated. From this OST trajectory, a total of 101 θ -space points were built to represent the initial string (Figure 2). The structures corresponding to the two end

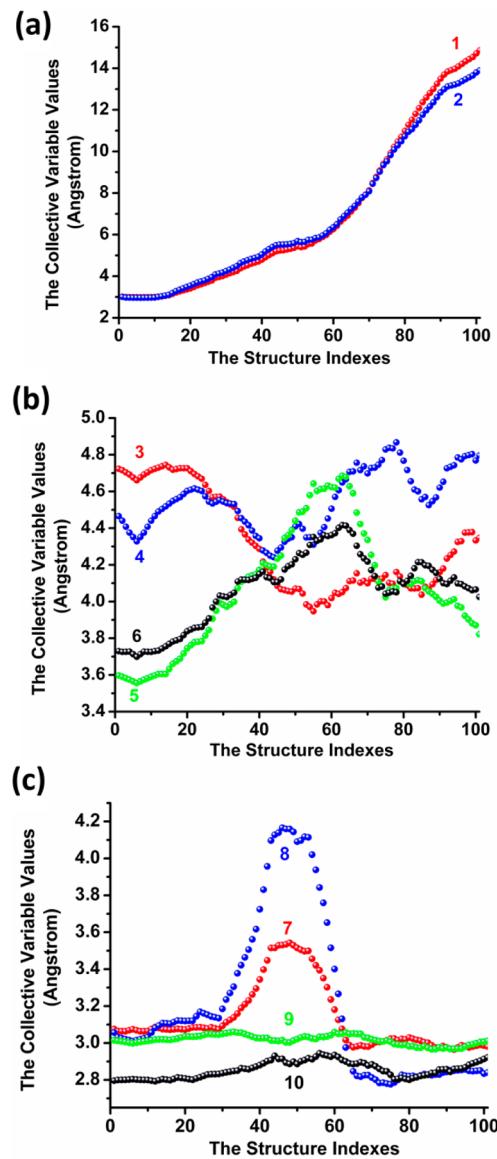


Figure 2. The θ -space points on the initial-guess path: (a) the base-pair separation collective variables; (b) the collective variables describing the interactions between the G19 base and the neighboring base planes; and (C) the collective variables describing the internal motions of the adjacent base pairs.

CPD states were used to define Z^A and Z^B . To ensure the initial path smoothness, each of the 99 intermediate points, which are shown by Figure 2, was built by averaging the θ values of 5 ps (ps) neighboring samples. Based on these θ -space points, we obtained the λ -dependent string functions $Z(\lambda) = (z_1(\lambda), \dots, z_{10}(\lambda))$ via the B-spline fitting method; here, the λ value of each point was calculated to reflect the percentage of its on-the-path distance from Z^A .

In the iOTPRW Hamiltonian (eq 2), m_λ was set as 10^3 a.m.u.; the friction coefficient in the λ Langevin dynamics was set as 2.0×10^5 a.m.u./ps; the path restraint force constant K_i was set as 100 kcal/mol/Å 2 . The λ boundary restraint constant K_{bound} (eq 3) was set as 10^5 kcal/mol. During free-energy derivative estimations and path optimizations, the λ space that spans from 0 to 1 was uniformly partitioned to 100 images. ΔT_f was set as 0.1 ps, and ΔT_z was set as 200 ps. In each path optimization step, the steepest-descent method was employed with the minimization step size Δz set as 0.005 Å 2 mol/kcal.

III.B. The On-the-Path Generalized Ensemble (GE) Simulation and Collective Variable Analysis. After ~ 12 ns (ns) of iOTPRW simulation, $Z(\lambda)$ was fixed and a 5-ns equilibrium GE simulation was performed. After the first 700 ps data was removed, the remaining samples were used to perform the collective variable analysis, particularly to assess how various collective variables are involved in the base-extrusion process. To evaluate each specific collective variable ξ , the two-dimensional (2D) free-energy surface, $G_0(\xi, \lambda)$, was constructed. For the fact that along the optimized path $Z(\lambda)$, λ represents an “ideal” reaction coordinate, the correlation relationship between ξ and λ in the free energy profile $G_0(\xi, \lambda)$ can provide a meaningful basis for the understanding of the role that ξ plays in the base-extrusion process.

III.C. The General Molecular Dynamics Simulation Setup. The molecular dynamics (MD) simulation setup was generated through the CHARMM-GUI server.⁷² The particle mesh ewald (PME) method⁷³ was applied to treat the long-range columbic interactions, while the short-range interactions were totally switched off at 12 Å. The Nôse-Hoover method⁷⁴ was employed to maintain a constant reservoir temperature at 300 K, and the Langevin piston algorithm⁷⁵ was used to maintain the constant pressure at 1 atm. The simulation time step was set to 1 fs.

IV. RESULTS AND DISCUSSION

IV.A. The Integration-Based On-the-Path Random Walk (iOTPRW) Simulation Result. The *iOTPRW Calculation Convergence Behavior*. During the 12-ns iOTPRW simulation, the progressing parameter λ repetitively traveled between the two end states (Figure 3a). The fact that there is no obvious sampling bottleneck, where λ random diffusion is possibly blocked, indicates that there is no significant “hidden” barrier in the space orthogonal to the pathway that is described by the preset collective variables; i.e., the target reaction pathway can be adequately represented by the prechosen collective variable set $\theta(X)$. Furthermore, throughout the iOTPRW simulation, all the base extrusions occur through the C6 major-groove channel; this suggests that the employed environmental collective variables ($\theta_1-\theta_{10}$) are sufficient to selectively steer base extrusions along the target pathway at least within the time scale that the iOTPRW simulation can represent.

To monitor the convergence behavior, the root-mean-square deviation (RMSD) between each pair of the estimated free-energy profiles was mapped (the inset of Figure 3b). The “maximum deviation” at each sampling length t' is defined by the maximum value at the upper right (t', t_{\max}) corner of the average RMSD map. The “maximum deviation” value represents the largest possible uncertainty within the overall simulation length; thus, it can be a fair indicator on calculation convergences. Figure 3b shows that, at ~ 9 ns, the maximum deviation dropped below 0.54 kcal/mol; and at ~ 10 ns, the

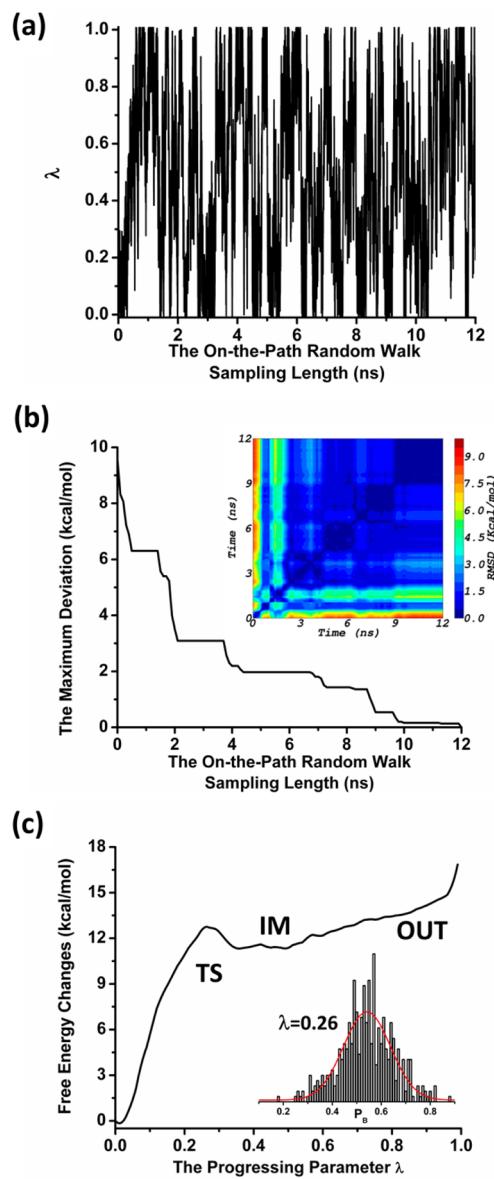


Figure 3. The iOTPRW calculation results: (a) the time-dependent progressing parameter changes; (b) the time-dependent “maximum deviation” changes (the map of the RMSD between each pair of the estimated free-energy profiles is shown in the inset); and (c) the free-energy surface along the optimized path; the committor probability distribution at the transition state is shown in the inset.

maximum deviation dropped below 0.1 kcal/mol. To monitor the path optimization convergence along a certain collective variable θ_i direction, the map of the RMSD between the corresponding functions $z_i(\lambda)$ at each pair of simulation lengths was generated. As shown by the right panels of Figure 5 (shown later in this work), ~ 7.5 ns were necessary for the collective variables that describe the base-pair separation, such as θ_1 (Figure 5a, shown later in this work), to enter the convergence phase; and it takes less than 9 ns for the collective variables that describe the base-extrusion environment changes, such as θ_4 (Figure 5b, shown later in this work) and θ_7 (Figure 5c, also shown later in this work), to enter their convergence phases. Furthermore, the convergence behaviors of the collective variable functions and the free-energy profile are in good agreement. It should be noted that, in an iOTPRW simulation,

usually, the majority of sampling time is used to optimize collective variable functions; and when these collective variable functions enter their convergence phase, sampling time required for the final free-energy convergence should be relatively short.

The Energetic and Structural Details of the MFEP Result. Based on the final free-energy profile (Figure 3c), we can identify two free-energy minimum regions: the intrahelical region at $\lambda = 0.03$ and an intermediate (IM) region, in which λ broadly spans between 0.38 and 0.50. The free energy of the IM region is ~ 11.6 kcal/mol higher than that of the intrahelical region. The transition state (TS) between these two free-energy minima is located around the state of $\lambda = 0.26$; and the corresponding free-energy barrier height is ~ 13.1 kcal/mol. To verify this transition state, a committor probability analysis⁷⁶ was performed. For this analysis, we carried out a 500-ps equilibrium simulation based on the on-the-path Hamiltonian, in which $Z(\lambda)$ is set as the optimized string and λ is fixed at 0.26; to obtain the committor value for each of the 500 collected samples, we generated 100 MD trajectories, each of which was initiated with a different random seed. As shown in the inset of Figure 3c, the calculated committor probabilities form a normal distribution that is centered at ~ 0.5 (0.54 from the Gaussian fitting, with the R^2 value of 0.84); this result further confirms the quality of the path optimization convergence and supports the fact that the state of $\lambda = 0.26$ is a valid transition state. In contrary to what is generally hypothesized, at the transition state (Figure 4a), the C6 base

point to each other, the WC hydrogen bonding interactions are mostly lost. Furthermore, we can tell from the TS and IM structures (see Figures 4a and 4b) that, at the transition state, the pathway starts to switch from the base-pair-elongation mode to the base-flipping mode; specifically, when the system moves forward from the TS region, the C6 base starts to flip out from the “dry interior”, so that it can be stabilized by water molecules in the major groove (see Figure 4c). Apparently, in this base-extrusion process, the base-pair stretching and base-flipping motions proceed in a stepwise fashion; the energetic penalty at the transition state is mainly introduced by the activation of the earlier motion, whose role may simply be preparing a ready environment for the latter motion to occur effortlessly.

Earlier umbrella sampling (US) studies reveal that base-extrusion free-energy profile calculations are sensitive not only to reaction-coordinate definitions but also to force-field treatments.^{9,24,29} Here, we focus our comparison with the result of a recent US study,²⁴ in which the identical DNA sequence and energy function were employed and the CPD collective variable was employed as the order parameter. In the region of $\lambda \in (0,0.5)$, the free-energy profiles obtained by the US and iOTPRW calculations are consistent, in terms of general shape and even free-energy height of the IM region, which is estimated to be ~ 12.0 kcal/mol by the US calculation. The major difference lies around the pre-TS and TS regions: the free-energy barrier height was calculated to be 16.0 kcal/mol by the US method, which is ~ 2.9 kcal/mol higher than the MFEP free-energy barrier; and the US free-energy profile is more rugged in the pre-TS region, whereas the iOTPRW free-energy surface is very smooth. Employing the equilibrium GE simulation samples, we recalculated the CPD-dependent potential of mean force (PMF); although these samples were collected along the optimized path, given the fact that the CPD may not be an ideal reaction coordinate, the CPD-based free-energy barrier is estimated to be 11.9 kcal/mol, which is ~ 1.2 kcal/mol lower than the MFEP free-energy barrier. As elaborated in the next section, these discrepancies suggest the fact that a base-flipping collective variable alone may not be sufficient to guide effective sampling of the TS formation, which, actually, is predominantly a base-plane elongation event. It is worth noting that the CPD-dependent PMF should be contributed by the microstates along all the possible paths, each of which corresponds to a unique flipped-out (OUT) end state, while the MFEP free-energy profile should describe the free-energy change along a single optimized path and it is usually dependent on the end-state. In the region of $\lambda \in (0,0.5)$, the to-be-extruded base is still confined in a narrow area; in this region, a converged US free-energy profile along a correct reaction coordinate should be comparable with the MFEP free-energy profile. However, in the region of $\lambda \in (0.5,1.0)$, the C6 base flips away from the tightly bound IM region and moves toward the predefined OUT state (see Figure 3c). Thus, in this region, the MFEP free-energy profile is dependent on the end-state and cannot be directly compared with the US calculation result; as the matter of fact, in the region of $\lambda \in (0.5,1.0)$, the PMF along a simple collective variable is likely to have large entropic contribution from all possible pathways. Notably, the focus of this study is mainly on the region of $\lambda \in (0,0.5)$, because this is the key region where the TS is formed and crucial “hidden” events are also involved.

The Collective Variable Changes along the MFEP. The above MFEP free-energy profile analysis further suggests that,

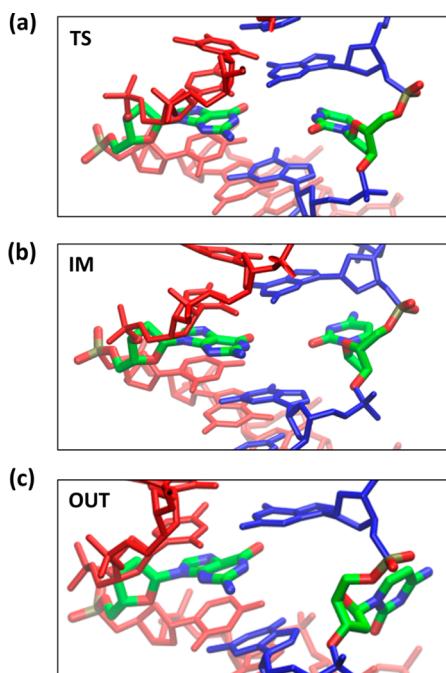


Figure 4. Schematics of the structures of (a) the transition state (TS), (b) the intermediate state (IM), and (c) the flipped-out end-state (OUT).

actually has not yet been extruded from the intrahelical interior and the C6–G19 base pair still remains in the same plane. Apparently, during the transition-state formation, the stretching of the WC base plane, rather than the base flipping, is the dominant event. As the consequence of the base-pair stretching, at the transition state, although the WC moieties still directly

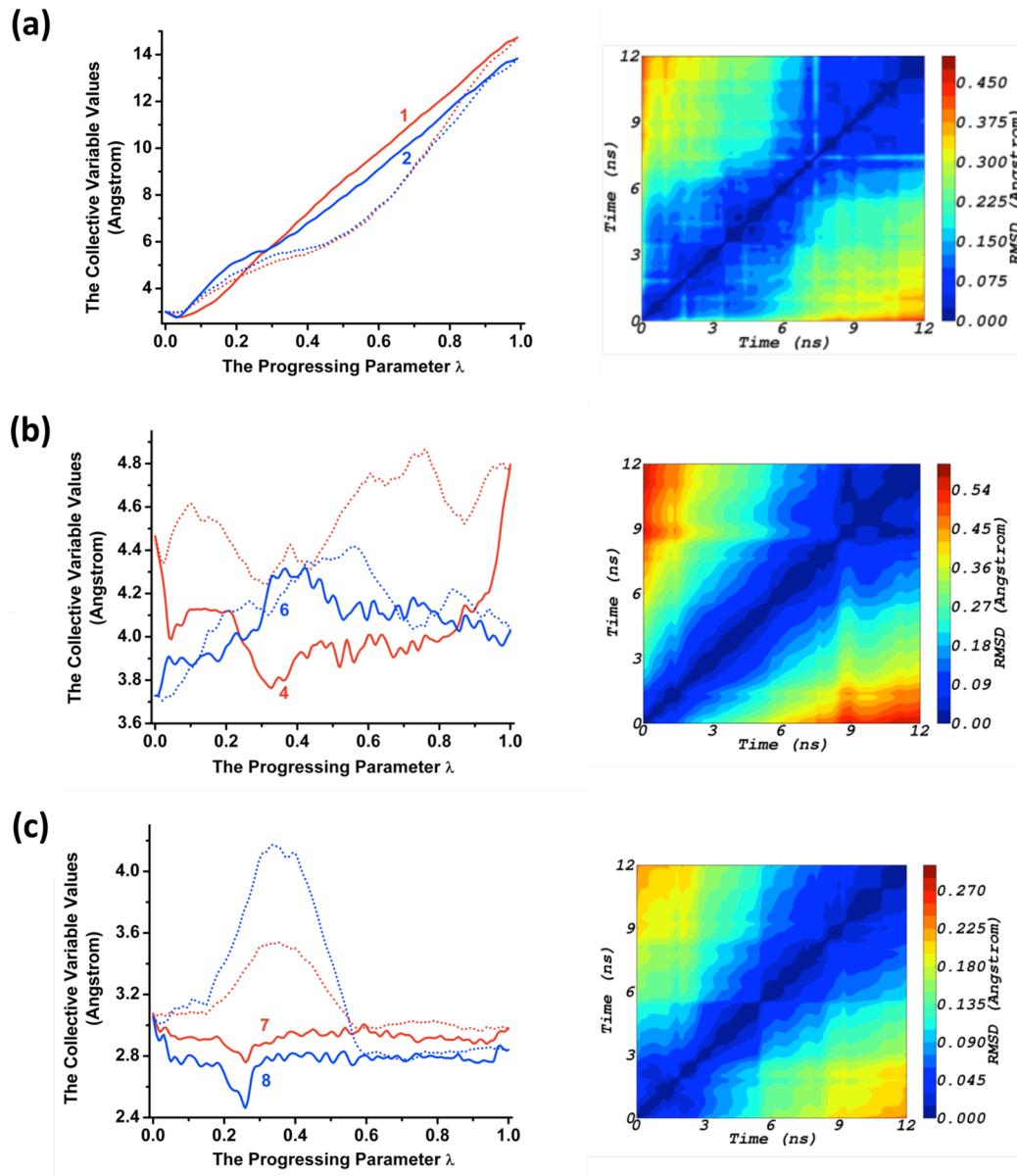


Figure 5. Evolutions of the essential collective variable candidates along the initial (the dotted lines) and optimized (the solid lines) pathways: (a) base-pair separation variable changes; (b) the changes of the variables describing the interactions between the G19 base and the neighboring base planes; and (c) the changes of the variables describing the internal motions of the neighboring base pairs. The left panel shows the evolutions of the representative essential collective variable candidates, and the right panel shows the maps of the RMSD between the representative collective variable functions obtained at each pair of simulation lengths.

at the pre-TS stage of the process, the associated events may be too complex for a single collective variable, such as the CPD, to represent. From the optimized MFEP, we can identify seven collective variables ($\theta_1-\theta_3$, θ_4 , $\theta_6-\theta_8$) as essential coordinate candidates, based on the fact that these coordinates have distinct on-the-path changes, particularly in the $\lambda \in (0, 0.5)$ region.

The left panels of Figure 5 show how these collective variables evolve along the optimized path. It is noted that the tiny beginning portion of an MFEP pathway (for instance, along this MFEP path, the region between the starting point Z^A and the free-energy minimum point $Z(\lambda = 0.03)$) is usually not informative but a necessary consequence of the path-optimization setup, because it is always preferable for the starting point to be chosen as a point before the reactant free-

energy minimum. The left panel of Figure 5a displays the on-the-path evolutions of the base-pair separation collective variables, θ_1 and θ_2 , which show that the MFEP transition state ($\lambda = 0.26$) has a tighter structure (with both θ_1 and θ_2 being slightly smaller than 6 Å) than the transition state obtained by a previous US study,¹⁶ the θ_1 value of which is ~7 Å. This structural difference provides a reasonable explanation why the free-energy barrier from the US calculation may have been overestimated. Notably, at the early stage of the path, the two base-pair separation distances change in a nonsynchronous manner. As observed from the trajectory, along this part of the pathway, indeed, an interbase rocking motion occurs around the vector along θ_1 ; i.e., the C6 plane then rocks away from the original WC base plane. Starting from $\lambda = 0.2$, the rocking motion is reversed and when the system approaches the

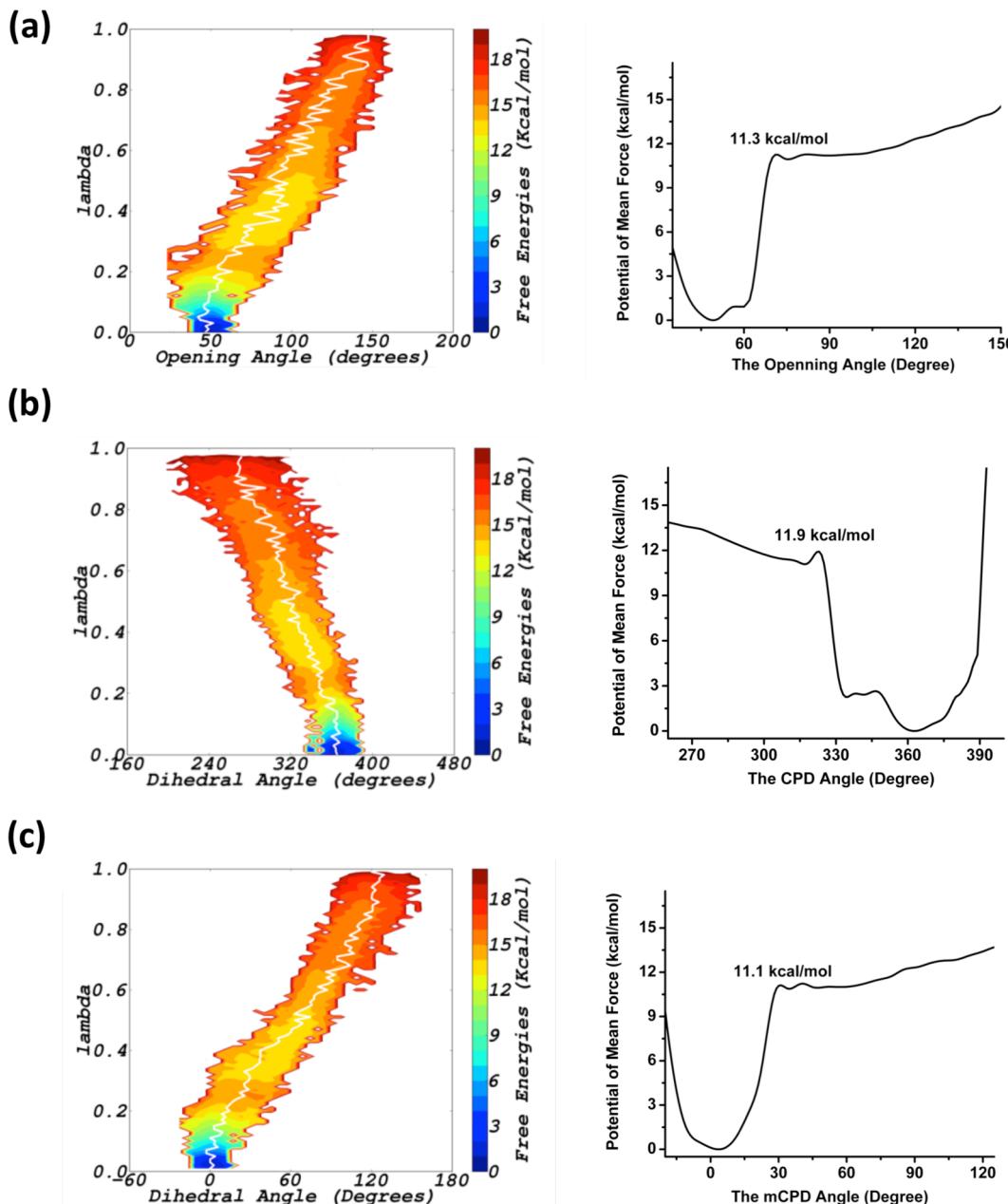


Figure 6. Two-dimensional (2D) free-energy profiles along each of the three examined collective variables (in the horizontal direction) and the progressing parameter (in the vertical direction): (a) the base-opening angle is the target collective variable, (b) the original CPD is the target collective variable, and (c) the mCPD is the target collective variable. The white lines in these free-energy profiles represent the average collective variable values along the path.

transition state, the broken WC base plane is reformed. Such two-step rocking motions may play a role in helping the strain introduced by the base-plane elongation motion to be released. The left panel of Figure 4b displays the progresses of the collective variables that describe the relative motions between the G19 plane and its adjacent base planes, θ_4 and θ_6 (θ_3 is not shown because it has similar on-the-path evolution behavior to that of θ_6). Obviously, during the formation of the transition state, the interbase-plane motions must be actively involved. The same is true for the C6-plane rocking motions; these motions may also be able to facilitate the stretching of the WC base pair. More importantly, as discussed in the next section, these motions can be crucial in selectively steering the base-extrusion process along the target channel. The left panel of

Figure 4c display the progresses of the collective variables that describe the internal motion of the adjacent G5–C20 plane, θ_7 and θ_8 . These on-the-path evolutions show that, simultaneous with the TS formation, the G5–C20 base plane is buckled; after the transition state, accompanying the flipping of the C6 base, the G5–C20 plane relaxes back to its canonical WC shape. These neighboring base-plane responses are clearly induced by the C6–G19 elongation motion. Interestingly, along the optimized path, the structure of the other neighboring (G7–G18) plane (represented by θ_9 and θ_{10}) does not have any obvious change. Such asymmetric responses further indicate that the base-extrusion environment plays a crucial role in selecting base-extrusion channels.

In summary, a well-converged MFEP and the corresponding free-energy profile were generated via the iOTPRW simulation. To our knowledge, this is the first time that the base-plane-elongation motion rather than the commonly focused base flipping motion is identified as the key component, specifically the major event responsible for the TS formation, in a base-extrusion process. Moreover, we determine that the base-pair stretching and base-flipping motions are largely dissociative. Furthermore, it is revealed that some intricate motions involving neighboring base planes actively participate in this part of the base-extrusion pathway. It is worth noting that the base-pair elongation motion and associated changes of the surrounding base-pair structures also were reported in an earlier study.¹⁶ Here, we further verify these observations with more-accurate details and, more importantly, we identify the crucial roles that they play in the base-extrusion process.

IV.B. The On-the-Path GE Simulation and the Collective Variable Analysis. The iOTPRW result allows us to identify several essential coordinate candidates from the employed collective variable set. Because of the fact that these collective variables were defined in the context of a specific process, the obtained path information may not be directly transferable for other DNA base-extrusion studies to use. Therefore, it is necessary to perform further analysis to derive relatively simple and transferable representations of the motions that may be generally meaningful. In addition, it is of importance to revisit previously proposed collective variables and re-evaluate them based on the obtained MFEP.

Analysis on the Flipping (Opening)-Angle Collective Variables. As explained in Section III.B, in our collective variable analysis, the free-energy profile ($G_0(\xi, \lambda)$) must be constructed to map the relationship between a collective variable ξ , and the “ideal” reaction coordinate λ . Here, we focus our examination on three flipping (opening) angle collective variables: the base-opening angle, the original CPD, and a modified CPD (mCPD). As shown by the left panel of Figure 6, all of these two-dimensional free-energy surfaces agree well with the free-energy profile from the path optimization calculation (see Figure 3c); for instance, the free-energy barriers in these free-energy profiles are all close to 13.1 kcal/mol. Such agreement indicates that the collective variable set $\theta(X) = (\theta_1(X), \dots, \theta_m(X))$ has adequately included representation of each of the three base-flipping order-parameter functions. More importantly, Figure 6 shows that these flipping-angle-based collective variables have no obvious change at the pre-TS stage of the base-extrusion process; this further confirms that the TS formation portion of the pathway is predominantly the base-pair-elongation event, rather than the base-flipping event. In terms of path selectivity, although the three commonly applied flipping-angle-based collective variables were designed to enforce the final base flipping direction, as shown in Figure 6, these collective variables cannot guide effective sampling until λ is close to 0.2. Furthermore, based on these high-quality equilibrium GE ensemble samples, we calculated the PMFs along these three collective variables (the right panels of Figure 6). In correspondence with the above understanding, these PMFs have only moderate flipping-angle-dependent changes until these flipping angles enter the PMF barrier regions, where sharp increases occur. It should be noted that the PMF value at a collective variable state is contributed by all the relevant microstate regions; because of the exponential relationship, the energetics of the lowest free-energy microstate region are likely to dominate the

corresponding PMF value. Therefore, if a collective variable is not an ideal reaction coordinate, the correct PMF barrier height is likely to be lower than the MFEP barrier height. In the present study, all the flipping angle collective variables show broad distributions at every λ state (see the left panel of Figure 6); consequently, in these PMFs, the contribution of the MFEP transition state is totally overshadowed by the energetics of the neighboring lower free-energy regions (along λ), in particular the IM regions. As shown by the right panel of Figure 6, the PMF barriers along the three flipping (opening) angles are, respectively, 11.3, 11.9, and 11.1 kcal/mol, which are \sim 1.2–2.0 kcal/mol lower than the MFEP barrier height; as a matter of fact, these values mostly reflect the free-energy height of the MFEP IM region (11.6 kcal/mol; see Figure 3). In addition, these three collective variables show different behaviors, in terms of how they correlate with λ ; these differences explain why US-based free-energy profile calculations are sensitive to flipping-angle definitions.

The Selectivity-Governing Motions. According to the above analysis, in base-extrusion simulations, it is crucial to have the base-pair-elongation motion properly represented in reaction coordinate definitions. In the current study, the base-pair separation distance collective variables (θ_1 and θ_2) are included for this purpose. As mentioned in the Introduction, these collective variables alone cannot selectively describe a unique path, in terms of distinguishing either between the target base and partner base extrusion pathways or between the major and minor groove extrusion routes. Thus, a question arises: What motions govern the path selectivity at the pre-TS stage of the extrusion process? Based on the details of on-the-path collective variable changes discussed in Section IV.A, among the identified essential coordinate candidates, θ_1 and θ_2 mainly take care of the base-pair-elongation motion; θ_7 and θ_8 only describe a neighboring plane internal response motion. Moreover, these collective variables do not have any obvious change until λ approaches 0.2 (see the left panel of Figure 6). Therefore, θ_3 , θ_4 , and θ_6 are the only candidates that are possibly responsible for the target path extrusion selectivity. Interestingly, all three of these collective variables were introduced to describe how the C6–G19 base plane, specifically the G19 base, moves relative to the neighboring base planes. By surveying our simulation trajectories, indeed, the on-the-path changes of these three relative variables follow a series of G19-base in-plane rotation motions. To represent these rotation motions in a general manner, as illustrated in Figure 7a, we can utilize the vector generated via the projection of $N9^{G7} \rightarrow C6^{C18}$ on the G7–C18 reference plane and the vector generated via the projection of $C8^{G19} \rightarrow C2^{G19}$ on the reference C6–G19 reference plane to define an in-plane rotation angle (IPRA) function; here, the G7–C18 and C6–G19 reference planes are referred to the corresponding base planes in an “ideal” B-DNA structure, upon which the obtained structures from the trajectory that contain the $N9^{G7} \rightarrow C6^{C18}$ and $C8^{G19} \rightarrow C2^{G19}$ vectors are superimposed. As shown in Figure 7b, different from the base-flipping angle collective variables, the IPRA undergoes distinct transitions throughout the entire process. At the very beginning of the process, with the decreasing of the IPRA value, the C6–G19 base plane slightly rotates into the minor groove region. At $\lambda \approx 0.08$, the base plane starts to rotate reversely; therefore, through the concurrent base-pair-elongation motion, the C6 base can be selectively extruded into the target major groove region. At the state of $\lambda \approx 0.16$, although the WC hydrogen bonding

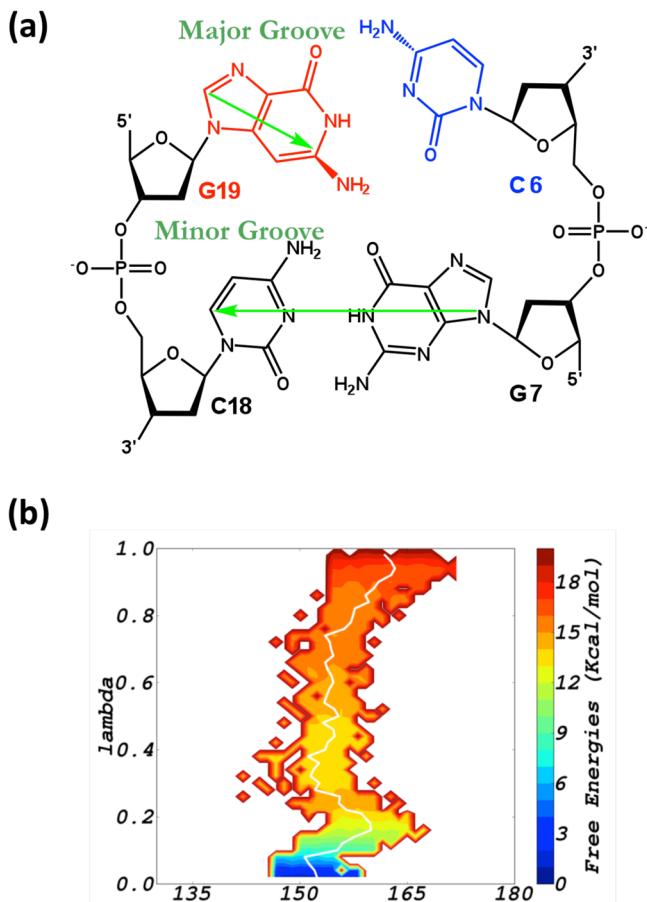


Figure 7. The in-plane rotation motions in the DNA base extrusion: (a) the in-plane rotation angle (IPRA) function is defined based on the vectors of N9G7 → C6^{C18} and C8^{G19} → C2^{G19} (the G19 major-groove rotations progress with increasing IPRA value), and (b) the two-dimensional free-energy profiles along the IPRA (in the horizontal direction) and the progressing parameter (in the vertical direction).

functional groups of C6 are not yet exposed (Figure 4a), the C6 base has been sufficiently delivered into the major groove region. In the following portion of the process, the G19 base starts to relax back toward its canonical ds-DNA position (Figure 7b) while, because of the continuing separation between the C6 and G19 bases, the C6 base still remains in the major groove region. Apparently, these in-plane rotation motions are not only essential for the base-extrusion path selectivity but also for the enabling of the base-plane-elongation dynamics in the confined ds-DNA environment. In the region of $\lambda \in (0, 0.16)$, without any obvious base-flipping motion, a free-energy increase of ~ 9.5 kcal/mol has already been built up (Figure 3c). After the transition state (Figure 4a), when the base-pair separation generates enough room for C6 to be mobilized inside the elongated WC structure, the C6 base starts to flip away from the “dry interior” of the B-DNA (Figure 4b) and at the same time, the G19 base almost completes its in-plane rotation and returns to the canonical ds-DNA location (Figure 7b). As also shown in Figure 7b, starting at $\lambda \approx 0.75$, the G19 base rotates away from its canonical ds-DNA position. Such unusual base rotation is likely to be caused by the MFEP end-state definition, which, as discussed earlier, can greatly influence the near-end-state MFEP result; and it also explains the sharp increase in the corresponding part of the free-energy profile (Figure 3c).

The above analysis further confirms the fact that, at the early stage of the C6-base major-groove extrusion process, the base-plane-elongation motion, which can create room for the mobilization of the C6 base, is the predominant event. Therefore, not surprisingly, all the commonly employed flipping-angle collective variables fail to represent the essential initial portion of the target base-extrusion pathway. In addition, we identify G19-base in-plane rotations as essential motions that may govern the base-extrusion selectivity.

V. CONCLUDING REMARKS

In this work, the on-the-path random walk (OTPRW) method, which is the first generalized-ensemble (GE) sampling-based finite-temperature-string algorithm, was improved to be an integration-based OTPRW (iOTPRW) algorithm for the understanding of a B-DNA major-groove base-extrusion pathway. From our iOTPRW minimum free-energy path (MFEP) optimization calculation, we identified a base-plane-elongation event rather than a base-flipping event as the dominant motion in the transition state (TS) formation portion of the process; moreover, we also show that, to facilitate the essential base-plane-elongation dynamics, the surrounding environment of the flipped base must be actively involved. Further taking the advantage of the extended-dynamics nature of the OTPRW Hamiltonian, an on-the-path equilibrium simulation was performed along the optimized path. The analysis of the collected samples reveals that none of these commonly employed flipping-angle based collective variables alone can adequately represent the base-extrusion pathway. Furthermore, the collective variable analysis also allows us to uncover the essential hidden events that couple with the base-plane-elongation motion and govern the base-extrusion selectivity: the base-pairing partner of the extrusion target undergoes a series of in-plane rotation motions to accommodate the base-plane-elongation dynamics. A base-plane rotation angle is identified as a general reaction coordinate to represent these in-plane rotations.

DNA base extrusion is a crucial component of many biomolecular processes. We believe that our understanding will shed light on related future studies. It is worth noting that, in this study, we have employed a relatively simple model system and only focused our analysis on the motions surrounding the base-extrusion site. As one could expect, in more-complex systems, more-extended and larger-scale dynamics may be involved. Then, if the iOTPRW method is applied, carefully choosing collective variable sets is definitely critical. Here, the term “base extrusion” instead of “base flipping” or “base opening” is employed to define the entire target process in order to differentiate it from the composition events including the base-flipping motion in the context of this work; nevertheless, to follow the tradition of this topic, base flipping is possibly more appropriate to represent such processes.

ASSOCIATED CONTENT

S Supporting Information

The movie in the supporting information shows the trajectory obtained from the on-the-path generalized ensemble simulation. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 850-645-6884. Fax: 850-644-7244. E-mail: yyang2@fsu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Dr. Alex Mackerell for helpful discussion and suggestions. Funding support from the National Institute of Health (No. GM054403) and the National Science Foundation (No. MCB0919983) is acknowledged. We would also like to thank the Florida State University High Performance Computing (HPC) Center and the Institute of Molecular Biophysics Computing Facility (Dr. Michael Zawrotny) for computing support.

REFERENCES

- (1) Bestor, T. H.; Verdine, G. L. *Curr. Opin. Cell Biol.* **1994**, *6*, 380–389.
- (2) Roberts, R. J. *Cell* **1995**, *82*, 9–12.
- (3) Roberts, R. J.; Cheng, X. D. *Annu. Rev. Biochem.* **1998**, *67*, 181–198.
- (4) Hornby, D. P.; Ford, G. C. *Curr. Opin. Biotechnol.* **1998**, *9*, 354–358.
- (5) Lindahl, T.; Wood, R. D. *Science* **1999**, *286*, 1897–1905.
- (6) Cheng, X. D.; Roberts, R. J. *Nucleic Acids Res.* **2001**, *29*, 3784–3795.
- (7) Fromme, J. C.; Banerjee, A.; Verdine, G. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 43–49.
- (8) Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *7*, 37–65.
- (9) Priyakumar, U. D.; Mackerell, A. D., Jr. *Chem. Rev.* **2006**, *106*, 489–505.
- (10) Gillet, L. C. J.; Scharer, O. D. *Chem. Rev.* **2006**, *106*, 253–276.
- (11) Tubbs, J. L.; Pegg, A. E.; Tainer, J. A. *DNA Repair* **2007**, *6*, 1100–1115.
- (12) Yang, C. G.; Garcia, K.; He, C. *ChemBioChem* **2009**, *10*, 417–423.
- (13) Sanders, D. B.; Wiest, O. J. *Am. Chem. Soc.* **1999**, *121*, 5127–5134.
- (14) Giudice, E.; Varnai, P.; Lavery, R. *ChemPhysChem* **2001**, *2*, 673–677.
- (15) Giudice, E.; Lavery, R. *Acc. Chem. Res.* **2002**, *35*, 350–357.
- (16) Banavali, N. K.; MacKerell, A. D. *J. Mol. Biol.* **2002**, *319*, 141–160.
- (17) Fuxreiter, M.; Luo, M.; Jedlovszky, P.; Simon, I.; Osman, R. *J. Mol. Biol.* **2002**, *323*, 823–834.
- (18) Huang, N.; Banavali, N. K.; MacKerell, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68–73.
- (19) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434–1443.
- (20) Hagan, M. F.; Dinner, A. R.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13922–13927.
- (21) Cheng, X. L.; Kelso, C.; Hornak, V.; de Los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2005**, *127*, 13906–13918.
- (22) Hart, K.; Nystrom, B.; Ohman, M.; Nilsson, L. *RNA* **2005**, *11*, 609–618.
- (23) Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16194–16198.
- (24) Priyakumar, U. D.; MacKerell, A. D. *J. Chem. Theor. Comput.* **2006**, *2*, 187–200.
- (25) Bouvier, B.; Grubmüller, H. *Biophys. J.* **2007**, *93*, 770–786.
- (26) Mura, C.; McCammon, J. A. *Nucleic Acids Res.* **2008**, *36*, 4941–4955.
- (27) Hu, J.; Ma, A.; Dinner, A. R. *Proc. Natl. Acad. Sci.* **2008**, *105*, 4615–4620.
- (28) Venditti, V.; Clos, L. II; Niccolai, N.; Butcher, S. E. *J. Mol. Biol.* **2009**, *391*, 894–905.
- (29) Song, K.; Campbell, A. J.; Bergonzo, C.; de Los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Chem. Theor. Comput.* **2009**, *5*, 3105–3113.
- (30) Zheng, H.; Cai, S.; Ding, Y.; Tang, K.; Kropachev, Y.; Zhou, L.; Wang, S.; Geacintov, N.; Zhang, Y.; Broyde, S. *Chem. Res. Toxicol.* **2010**, *23*, 1868–1870.
- (31) Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A. P.; Simmerling, C. *J. Am. Chem. Soc.* **2011**, *133*, 14504–14506.
- (32) Keepers, J. W.; Kollman, P. A.; Weiner, P. K.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5537–41.
- (33) Chen, Y. Z.; Mohan, V.; Griffey, R. H. *J. Biomol. Struct. Dyn.* **1998**, *15*, 765–777.
- (34) Giudice, E.; Varnai, P.; Lavery, R. *Nucleic Acids Res.* **2003**, *31*, 1434–1443.
- (35) Ma, A.; Dinner, A. R. *J. Phys. Chem.* **2005**, *109*, 6769–6779.
- (36) Basner, J. E.; Schwartz, S. D. *J. Am. Chem. Soc.* **2005**, *127*, 13822–13831.
- (37) Pineda, E. D.; Antoniou, D.; Schwartz, S. D. *J. Phys. Chem. B* **2010**, *114*, 15985–15990.
- (38) Rosta, E.; Woodcock, H. L.; Brooks, B. R.; Hummer, G. *J. Comput. Chem.* **2009**, *30*, 1634–1641.
- (39) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. *Nucleic Acids Res.* **2013**, *41*, 33–43.
- (40) Ensing, B.; Laio, A.; Parrinello, M. *J. Phys. Chem.* **2005**, *109*, 6676–6687.
- (41) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (42) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
- (43) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
- (44) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- (45) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (46) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- (47) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904–2914.
- (48) Babin, V.; Roland, C.; Sagui, C. *J. Chem. Phys.* **2008**, *128*, 134101.
- (49) Zheng, L. Q.; Chen, M. G.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.
- (50) Zheng, L. Q.; Chen, M. G.; Yang, W. *J. Chem. Phys.* **2009**, *130*, 234105.
- (51) Zheng, L. Q.; Yang, W. *J. Chem. Theor. Comput.* **2012**, *8*, 810–823.
- (52) Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375–380.
- (53) Henkelman, G.; Uberuaga, B. P.; Jonsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- (54) E, W. N.; Ren, W. Q.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
- (55) Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R. *Theor. Chem. Acc.* **2003**, *109*, 140–148.
- (56) E, W. N.; Ren, W. Q.; Vanden-Eijnden, E. *J. Phys. Chem.* **2005**, *109*, 6688–6693.
- (57) Brokaw, J. B.; Haas, K. R.; Chu, J. W. *J. Chem. Theor. Comput.* **2009**, *5*, 2050–2061.
- (58) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. *J. Chem. Phys.* **2006**, *125*, 024106.
- (59) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2007**, *446*, 182–190.
- (60) Pan, A. C.; Sezer, D.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 3432–3440.
- (61) Chen, M. E.; Yang, W. *J. Comput. Chem.* **2009**, *30*, 1649–1653.
- (62) Zhu, F. Q.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19814–19818.
- (63) Huang, H.; Post, C. B. *J. Phys. Chem. B* **2012**, *116*, 11046–11055.
- (64) Lelievre, T.; Rousset, M.; Gabriel, S. *J. Chem. Phys.* **2007**, *126*, 134111.
- (65) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313.

- (66) Min, D. H.; Chen, M. E.; Zheng, L. Q.; Jin, Y. H.; Schwartz, M. A.; Sang, Q. X. A.; Yang, W. *J. Phys. Chem.* **2011**, *115*, 3924–3935.
- (67) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (68) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Calfisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Feig; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (70) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (71) MacKerell, A. D., Jr.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (72) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (73) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (74) Nōse, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (75) Nōse, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055–1076.
- (76) Bolhuis, P. G.; Dellago, C.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5877–5882.