A Molecular Dynamics Study of Slow Base Flipping in DNA using Conformational Flooding

Benjamin Bouvier and Helmut Grubmüller
Theoretical and Computational Biophysics Department, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

ABSTRACT Individual DNA bases are known to be able to flip out of the helical stack, providing enzymes with access to the genetic information otherwise hidden inside the helix. Consequently, base flipping is a necessary first step to many more complex biological processes such as DNA transcription or replication. Much remains unknown about this elementary step, despite a wealth of experimental and theoretical studies. From the theoretical point of view, the involved timescale of milliseconds or longer requires the use of enhanced sampling techniques. In contrast to previous theoretical studies employing umbrella sampling along a predefined flipping coordinate, this study attempts to induce flipping without prior knowledge of the pathway, using information from a molecular dynamics simulation of a B-DNA fragment and the conformational flooding method. The relevance to base flipping of the principal components of the simulation is assayed, and a combination of modes optimally related to the flipping of the base through either helical groove is derived for each of the two bases of the central guanine-cytosine basepair. By applying an artificial flooding potential along these collective coordinates, the flipping mechanism is accelerated to within the scope of molecular dynamics simulations. The associated free energy surface is found to feature local minima corresponding to partially flipped states, particularly relevant to flipping in isolated DNA; further transitions from these minima to the fully flipped conformation are accelerated by additional flooding potentials. The associated free energy profiles feature similar barrier heights for both bases and pathways; the flipped state beyond is a broad and rugged attraction basin, only a few kcal/mol higher in energy than the closed conformation. This result diverges from previous works but echoes some aspects of recent experimental findings, justifying the need for novel approaches to this difficult problem: this contribution represents a first step in this direction. Important structural factors involved in flipping, both local (sugar-phosphate backbone dihedral angles) and global (helical axis bend), are also identified.

INTRODUCTION

Base flipping is an elementary and localized deformation of a double-stranded DNA fragment, initiated by the rupture of the Watson and Crick hydrogen bonds at a target basepair and followed by the swiveling of one of the bases to an extrahelical position, where it subsequently becomes exposed to chemical attack from its environment. Repair and modification enzymes, for example, take advantage of this availability to operate on the base itself. In fact, it is assumed that these enzymes facilitate the flipping mechanism upon binding. Cytosine-5 methyltransferase and uracil glycosylase are two such enzymes, whose actions have been studied both theoretically and experimentally (1–7). X-ray diffraction structures of these enzymes in complex with base-flipped DNA strands have been published (8,9). Experimental evidence suggests two different opening mechanisms: uracil glycosylase facilitates the flipping of its target uracil through the major groove of the DNA double strand (5), whereas cytosine-5 methyltransferase favors a minor groove pathway for its target cytosine (2) (although theoretical results (7,10) supporting major groove flipping have recently been published).

Interestingly, base flipping does not require the assistance of an enzyme. It has been proven to occur spontaneously in solution, by means of NMR studies monitoring the increase in the imino proton exchange rate when the base becomes accessible to the solvent (11–14). Base flipping was found to occur on a millisecond timescale, compatible with the energy penalty involved with the breaking of Watson and Crick hydrogen bonds and the unstacking of the target base relative to its intrastrand neighbors. Novel experimental techniques (15,16) based on measuring the time-resolved fluorescence emission of a probe chromophore binding to the extrahelical bases could yield complementary data in the near future. In particular, additional experimental evidence on the relative stability of the flipped state of DNA, as well as possible structural characterizations thereof, is highly desirable.

Base flipping also plays an important role as an elementary step in complex biological processes. For instance, it is supposed to be involved in the separation of strands that precedes the DNA replication and transcription processes, where it stabilizes intermediate protein-DNA complexes (17–19). Additionally, it is thought to induce modifications to the mechanical properties of the double helix, promoting the selectivity of enzyme-DNA recognition (20–25). Very recently, an example of sequence-independent recognition was reported (26), revealing the key role of base flipping in the recognition and exchange of sequence-degenerate single-stranded genetic material between bacteria.

Due to its ubiquitous functional relevance, an atomistic understanding of the base flipping mechanism, together with

Submitted June 20, 2006, and accepted for publication April 6, 2007.

Address reprint requests to H. Grubmüller, Tel.: 49-551-201-2301; E-mail: hgrubmu@gwdg.de

Editor: Tamar Schlick.

0006-3495/07/08/770/17 $2.00

doi: 10.1529/biophysj.106.091751

its driving forces and energetics, should yield insights into a very diverse array of biological processes related to protein-DNA interaction and recognition. It is also intrinsically valuable for the understanding of the structural and mechanical properties of double-stranded DNA and its dynamical behavior.

Due to the relatively short lifetimes of intermediate conformations, experimental studies of base flipping have only been able to characterize the endpoints of the base-flipping process (i.e., the closed and fully flipped states), or, more recently, relatively stable intermediates (27) corresponding to local free energy minima. Theoretical studies can provide a very helpful complement by simulating the entire flipping pathway and proposing relevant intermediate states and transitions in structural parameters. Another incentive for these studies is the assessment of the relative stability of the different structures involved, as well as the calculation of the free energy barriers that separate them. Finally, from a more practical perspective, the comparison between experimental and theoretical results should provide a tough test case for the methods and force fields employed.

However, the timescale on which the base flipping process occurs (millisecond or longer) is far out of the reach of state-of-the-art classical molecular dynamics (MD) simulations. This is especially true for free energy calculations, since the high-energy conformations which will rarely—if ever—be sampled during a simulation are those that have the highest weight in the partition function. Their omission therefore results in unconverged free energy profiles.

To overcome this problem, theoretical studies on base flipping have so far employed the umbrella sampling procedure (28) together with the weighted histogram analysis method (29,30). Within this framework, a variable (also referred to as an order parameter) is defined which describes the progression along the flipping pathway. Multiple MD simulations are then carried out inside windows at different values of this parameter (enforced by a harmonic restraining potential) spanning the entire transition pathway. The bias introduced by this umbrella potential is subsequently removed by an iterative procedure using data from all the windows, yielding an approximation of the unbiased population density necessary to compute the free energy profile along the reaction coordinate. Most prominent among the applications of this methodology to base flipping in DNA are works by the groups of Lavery (22,31–36) and MacKerell (27,37–41). Among these, a number of studies (27,38,39) compare the energetics of the flipping motion in the DNA-enzyme complex of cytosine-5 methyltransferase to those of DNA alone.

Within the umbrella-sampling framework, the choice of the order parameter dictates to a large extent the resulting reaction pathway and energetics. This is not an issue for systems where the choice of this parameter is unambiguous. Unfortunately, for the specific problem of base flipping, several coordinates are conceivable. In fact, the two major theoretical studies to date have adopted different angular coordinates to this end, corresponding to slightly different assumptions on the pathway and the degrees of freedom involved. While these studies concur on a number of similar trends, they have not been able to provide an unambiguous flipping mechanism, especially beyond the initial breaking of the Watson and Crick hydrogen bonds, nor to clearly identify the factors triggering base flipping. Obtaining additional information is crucial to complement and validate experimental data, especially since the experimental side of the problem is also subject to incertitudes. Recent NMR results (C. Griesinger and H. Schwalbe, 2006, private communication) provide evidence for the insufficiency of the commonly accepted two-state model (closed and flipped, the latter only exchanging protons with the solvent) on which experimental free energy estimations are based. The existence of local minima on the free energy surface, their relative energetic penalty compared to the closed state, and the nature of the fully flipped state, thus benefit from further investigation.

This article presents and discusses an alternate theoretical approach, using principal component analysis coupled with conformational flooding. We attempt to derive the global coordinates relating to flipping from the dynamics of a B-DNA duplex in its closed conformation, under as few assumptions as possible, using principal component analysis. By combining the relevant principal components, we define global coordinates, each of which describes the initial flipping of a particular base through one of the two grooves. Applying conformational flooding along these coordinates during molecular dynamics simulations allows the system to overcome the initial free energy barrier on a short timescale. Subsequent flooding simulations reveal the entire pathway, for each base and each groove, in a stepwise manner. We then apply restraints along the aforementioned global coordinates to compute the free energy profiles along each possible pathway. Finally, we compare our results to the existing theoretical and experimental literature, highlighting our model's advantages and possible shortcomings.

METHODS

Principal components analysis

To determine the linear collective coordinates that encapsulate the most significant global motions of the DNA, principal component analysis (PCA) was applied to the configuration space ensemble sampled by MD simulations, as outlined below.

Defining $\mathbf{x} = (x_1, x_2, \ldots, x_N)^t$ as the $3N$-dimensional vector of the N nuclear positions of the molecule in Cartesian space, any linear collective coordinate $\mathbf{c}$ takes the form

$$\mathbf{c} = \mathbf{Q}(\mathbf{x} - \bar{\mathbf{x}}), \tag{1}$$

where $\mathbf{Q}$ is a real orthonormal square matrix of rank $3N$, and $\bar{\mathbf{x}}$ a $3N$-dimensional vector representing the origin of the coordinate. In the framework of principal component analysis, $\mathbf{Q}$ is calculated by diagonalizing the covariance matrix $\mathbf{C}$ of the nuclear dynamics, obtained from the ensemble $E = \{\mathbf{x}^{\{k\}}\}_{k\in[1\ldots n]}$ of n configurations $\mathbf{x}$ generated by the MD simulation,

$$\mathbf{C} = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) \cdot (\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathrm{t}} \rangle = \mathbf{Q}^{\mathrm{t}} \mathbf{\Lambda} \mathbf{Q}, \tag{2}$$

where the angular brackets denote averages over the ensemble E. The origin $\overline{\mathbf{x}}$ (Eq. 1) is the average structure over this ensemble,

$$\overline{\mathbf{x}} = \langle \mathbf{x} \rangle = \frac{1}{n} \sum_{\mathrm{k}=1}^{\mathrm{n}} \mathbf{x}^{\{\mathrm{k}\}}. \tag{3}$$

$\mathbf{\Lambda}$ is the eigenvalue matrix

$$\mathbf{\Lambda} = \left(\delta_{\mathrm{ij}} \lambda_{\mathrm{i}}^2 \right)_{\mathrm{i,j} \in (1..3\mathrm{N})^2}, \tag{4}$$

the eigenvalue λ_{i} measuring the extension of the ensemble E along the direction of the corresponding eigenvector $\mathbf{q}_{\mathbf{i}} = (\mathbf{Q}_{1\mathrm{i}} \mathbf{Q}_{2\mathrm{i}} \dots \mathbf{Q}_{3\mathrm{Ni}})^{\mathrm{t}}$. Thus, PCA identifies collective motions associated with the largest variance and lowest frequencies as the first few eigenvectors, assuming that these are sorted in the order of decreasing λ_{i}. High-frequency low-amplitude collective modes that do not contribute much to the fluctuations within ensemble E, on the contrary, are associated with small eigenvalues and form the last few eigensolutions.

The fact that the PCA eigenmodes are sorted by relevance to the global atomic motion is very useful for selecting an ensemble of m so-called essential coordinates, $m < 3N$, that optimally span this global motion. This is discussed further in the next section, from the standpoint of conformational flooding.

Conformational flooding

The principles of conformational flooding are explained in much detail in the original article (42), to which the reader is referred. Successful applications to protein dynamics (43) or the prediction of unimolecular reactions in small systems (44) have also been published. Hence, the method will only be briefly summarized, here and in Fig. 1. It is based on the construction of an approximate local free energy landscape for the system, using relevant essential collective coordinates typically obtained from a PCA analysis. This landscape is supposed to be multiharmonic, i.e., its projection along each essential coordinate is a Gaussian function of that coordinate. From this approximation, a local repulsive flooding potential is built and added into the Hamiltonian for nuclear motion. It destabilizes a given educt conformation of the system without affecting the surrounding barriers or the other possible product conformations lying beyond, effectively accelerating the associated conformational changes from educt to product. Therefore, an MD simulation using such a modified Hamiltonian is likely to reveal conformational changes which would normally only occur at a much longer timescale using the unmodified nuclear Hamiltonian.

Here, the inverse covariance matrix from a PCA is used to construct a multivariate Gaussian approximation $\rho(\mathrm{x})$ of the configuration space density

$$\rho(\mathbf{x}) = \frac{1}{Z} \exp\left[-\frac{1}{2} (\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathrm{t}} \mathbf{C}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle) \right], \tag{5}$$

with the associated approximate partition function

$$Z(\mathbf{x}) = \int \exp\left[-\frac{1}{2} (\mathbf{x} - \langle \mathbf{x} \rangle)^{\mathrm{t}} \mathbf{C}^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle) \right] d^{3\mathrm{N}}\mathbf{x}. \tag{6}$$

At this point, $m < 3N - 6$ essential coordinates c_{j}, $1 < j < m$, are selected from the PCA eigenvectors, disregarding the three rigid-body rotations and the three translations of the entire system. Using the approximate expression for the configuration space density (Eq. 5) and integrating over the remaining (unessential) degrees of freedom yields the approximate harmonic free energy landscape,

$$G^{\mathrm{qh}}(\mathbf{c}) = -\frac{1}{2} k_{\mathrm{B}} T \sum_{\mathrm{j}=1}^{\mathrm{m}} \lambda_{\mathrm{j}} c_{\mathrm{j}}^2, \tag{7}$$

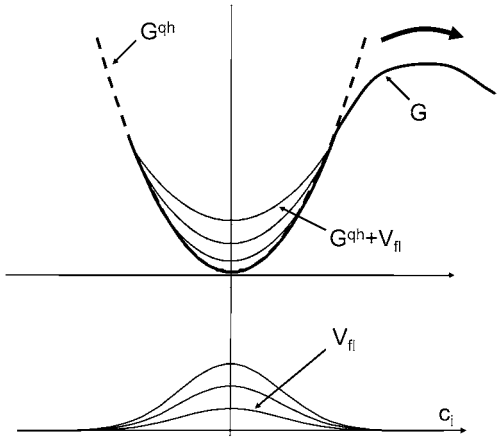

FIGURE 1 Action of the flooding potential. The lower part shows the flooding potential V_{fl}, with different curves corresponding to different values of the flooding strength E_{fl}. The upper part shows the free energy landscape G along a generalized coordinate c_{i} (*bold line*), its local quasiharmonic approximation G^{qh} (*bold dashes*) and the flooded landscape once the potential has been applied (*thin lines*, different E_{fl} values). Inclusion of the flooding potential within the force field effectively lowers free energy barriers and accelerates conformational transitions.

with $\mathbf{c} = \{c_{\mathrm{j}}\}$ defined as in Eq. 1. This approximation is obviously only valid in the region of configurational space spanned by the educt state and does not describe the barrier separating it from other conformers of the system. The flooding potential is then expressed as

$$V_{\mathrm{fl}}(x) = E_{\mathrm{fl}} \exp\left[-\frac{k_{\mathrm{B}} T}{2\omega} \sum_{\mathrm{j}=1}^{\mathrm{m}} \lambda_{\mathrm{j}} c_{\mathrm{j}}^2 \right]. \tag{8}$$

As can be inferred from Eq. 8, the flooding potential is a function of the m essential degrees of freedom c_{j} only: the corresponding force acts on these degrees of freedom exclusively, leaving the rest unaltered. Although PCA eigenvectors, or linear combinations thereof, were used as essential modes in this study, other choices (such as normal modes (42,44)) are possible. The essential modes are generally the ones that contribute the most to the nuclear dynamics: in the case of PCA, they are associated with the m largest eigenvalues λ_{i}. By increasing the value of m, the level of approximation can be made finer; conversely, by reducing the value of m or selecting a relevant ensemble of coordinates c_{j}, the action of the flooding potential can be focused into selected subspaces, ignoring degrees of freedom, which are known to be unrelated to the problem under study. The force deriving from the flooding potential, initially computed in the subspace spanned by the coordinates c_{j}, is subsequently translated into Cartesian space and added to the force field. At this point, the action of the flooding potential can be limited spatially, by letting the corresponding force act on a reduced subset of atoms only rather than on the entire set of atoms considered for the PCA. This is useful to limit the applied bias to regions of a macromolecule where it is actually needed, especially when the ensemble of atoms considered for PCA is very large.

E_{fl} is a scaling parameter referred to as the flooding strength; its action may be seen in Fig. 1. The destabilization free energy introduced by the flooding potential increases with increasing flooding strength, and decreases with increasing m. The parameter ω, controlling the width of the flooding potential, is increased above its canonical value of $\omega = E_{\mathrm{fl}}$ to compensate for partially insufficient sampling in the ensemble E, as described below (see Designing a Flooding Potential; see also Table 2).

The flooding potential introduced in Eq. 8 is repulsive, i.e., it expels the system from a given region and prohibits it from returning there. By negating it, one obtains an attractive potential that can be used, for instance,

as a restraint along one or several PCA eigenmodes. A harmonic restraining potential can be constructed in a very similar fashion, with the eigenvalues λ_j acting as spring constants (Eq. 10 below). Such a potential was applied to the determination of free energy profiles in a manner similar to that of the umbrella sampling and the weighted histogram analysis method, described in the next subsection.

Umbrella sampling and the weighted histogram analysis method

Umbrella sampling (US) (28) and the weighted histogram analysis method (WHAM) (29,30) are well-established recipes to extract free energy profiles from MD simulations. Obtaining a converged profile requires good sampling of the high-energy conformations of the system under study, only achieved at thermodynamical equilibrium on the associated free energy surface at the simulated temperature. These conditions are generally not met in practice, but can be obtained locally on a modified free energy landscape using known biasing potentials which favor the sampling of high-energy regions. The subsequent correction of the bias provides an approximation of the actual free energy profile. The reader is referred to the original works for a more detailed explanation of these techniques; in this study, they have been used to provide an enhanced (non-Boltzmann) sampling of the high-energy conformations along the flipping pathway, as revealed by conformational flooding.

A total of W individual simulations (windows) were performed, adding to the nuclear Hamiltonian $\hat{H}_0$ a window-dependent restraining potential $\hat{V}_i$, $i \in [1 \dots W]$,

$$\hat{H}_i(\mathbf{x}) = \hat{H}_0(\mathbf{x}) + \hat{V}_i(\mathbf{x}). \tag{9}$$

The restraining potentials $\hat{V}_i, i \in [1 \dots W]$ are harmonic and defined similarly to inverted flooding potentials,

$$\hat{V}_i(\mathbf{x}) = \frac{1}{2} k_{harm} c_i(\mathbf{x})^2, \tag{10}$$

where

$$c_i(\mathbf{x}) = \left(\mathbf{x} - \mathbf{x}_i^{ref}\right)^t \mathbf{q} \tag{11}$$

is the projection of the trajectory on the relevant optimal flipping mode $\mathbf{q}$ (see Reweighting the PCA Eigenvectors), which depends on the base and pathway. The origin of the projection is a conformation $\mathbf{x}_i^{ref}$ on the flipping pathway obtained from the flooding simulations, chosen such that the sampling distribution is shifted along the pathway from one window to the next, with the W windows spanning the entire pathway. k_{harm} is the harmonic force constant. A harmonic potential was judged more suited to the umbrella sampling method than an inverted Gaussian flooding potential, essentially for its behavior at large values of c_i. However, as for flooding, the main motivation is that the restraint can be applied along one or a combination of relevant global coordinates only.

Each simulation window employed its own reference structure $\mathbf{x}_i^{ref}$, providing sufficient sampling of the region of configuration space around this point. At each window, the system was restrained along the generalized flipping coordinate $\mathbf{q}$ but unrestrained along all remaining nonessential degrees of freedom. Applying the WHAM framework to the specific problem under study yields the expression of the free energy f_i of the system in the neighborhood of $\mathbf{x}_i^{ref}$ (29),

$$\exp(-f_i/k_BT) = \sum_{k=1}^{W} \sum_{t=1}^{n_k} \frac{\exp\left[-\frac{1}{k_BT}\sum_{j=1}^{W} V_{j,t}^{(k)}\right]}{\sum_{m=1}^{W} n_m \exp\left[f_m - \frac{1}{k_BT}\sum_{j=1}^{W} V_{j,t}^{(k)}\right]}, \tag{12}$$

where n_i is the number of snapshots generated by the i^{th} simulation window and $V_{j,t}^{(k)}$ the value of the restraining potential $\hat{V}_j$ (Eq. 10) at the t^{th} snapshot of the k^{th} window. This self-consistent expression was iterated to convergence over the set of simulation windows, resulting in a free energy profile or potential of mean force.

Geometrical quantification of base flipping

To provide a geometrical measure of base flipping along an MD trajectory or the projection thereof, the flipping angle previously defined by Bernet et al. (31) and sketched in Fig. 2 was employed.

The local helical axis at the target basepair is computed by calculating the average of the two vectors joining the C1′ atoms of the bases on either side of the target base in each strand, and obtaining the perpendicular projection of this vector with respect to the C1′-C1′ vector of the target basepair. The plane P whose normal is the local axis and contains the C1′-C1′ vector is thus defined. The flipping angle is then calculated between the C1′-C1′ vector and the projection onto P of the glycosidic bond of the target base. It measures close to 55° for canonical B-DNA, and exhibits variations of ±10° during an MD simulation at 300 K. Positive (respectively, negative) variations of the angle denote flipping of the target base through the major (respectively, minor) groove of the helix. This definition separates the actual flipping motion toward the grooves from the buckling of the target basepair, as the latter has no effect on the angle due to the projection onto the plane P.

Optimizing a linear combination of PCA eigenvectors to describe flipping

To reproducibly trigger the onset of base flipping on tractable timescales, it proved necessary to build a flooding potential acting on a single generalized coordinate, defined as an optimal linear combination of PCA eigenmodes; this is explained in Reweighting the PCA Eigenvectors. The generalized coordinate $\mathbf{X}_{opt}$ was written as

$$\mathbf{X}_{opt} = \sum_{\text{PCA eigenmodes i}} \lambda_i c_i \mathbf{q_i} \text{ subject to } \sum_{\text{PCA eigenmodes i}} c_i^2 = 1, \tag{13}$$

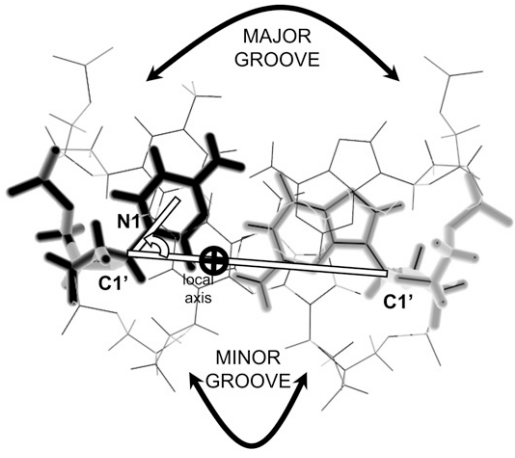

FIGURE 2 Definition of the angle (31) used to quantify the flipping motion. The target cytosine is shown as solid sticks, its partner guanine as shaded sticks, and the two neighboring basepairs that define the local helical axis (*perpendicular to the plane of the figure*) as shaded lines. The flipping angle of the target base is symbolized by the thick open lines and arrow (refer to text for details).

where the $\mathbf{q}_i$ and λ_i are the PCA eigenvectors and eigenvalues, respectively, and the coefficients c_i are the unknowns. For any given set $\{c_i\}$, the reference trajectory used for PCA is projected onto $\mathbf{X}_{opt}$, and the difference in flipping angle between the two extremes of the projection is computed. This difference is the cost function: $\{c_i\}$ sets that maximize it are solutions to the problem.

This optimization task with nonlinear constraints was solved using the nonlinear programming techniques in Computational Details. Since no straightforward analytical expression of the cost function is available, a numerical derivative scheme was used to provide conjugate gradient information.

COMPUTATIONAL DETAILS

All DNA structure parameters used in this work conform to their respective definitions in reference (45). Watson and Crick hydrogen bonds were supposed to exist if the corresponding donor-acceptor distance was lower than 3.5 Å and the deviation to linearity of the donor-hydrogen-acceptor angle was $<30°$.

All simulations were performed using the AMBER99 (46) force field implemented within the GROMACS 3.2.1 (47,48) simulation package. The velocity Verlet algorithm with a 2 fs timestep was used to integrate the equations of motion. Periodic boundary conditions were applied. The particle-mesh Ewald approach (49) with a 10 Å direct space cutoff distance was employed for the long-range electrostatic interactions. Temperature and pressure couplings were treated with the method of Berendsen (50). The translation motion of the center of mass of the DNA was removed at every time step. The length of the bonds involving hydrogen atoms were constrained using LINCS (51). DNA strands were built using NAB (52). Structural parameters of DNA strands were computed with CURVES (53).

The model 13-mer DNA duplex employed here, (dG-(dA-dG)$_{12}$)·(dA-(dG-dA)$_{12}$), was built in canonical B-DNA geometry. It was then neutralized using 24 Na^+ counterions and solvated with 6742 TIP3P water molecules in an octahedral box. Solvent and counterions were relaxed by energy minimization and subsequently equilibrated for 100 ps at 300 K and constant volume; the DNA molecule was kept frozen in these two steps. Harmonic restraints of 5 kcal/mol^{-1} Å^{-1} on all DNA atoms were then applied for 500 ps, and progressively removed over 1 ns, at a constant pressure of 1 bar. The coupling parameters for the temperature and pressure algorithms were relaxed to their final values of 1.0 ps and 0.5 ps, respectively, over 500 ps. The system was then simulated for 5 ns before production runs were begun. The equilibrated simulation box had a face-to-face dimension of 65 Å. A total of 200,000 configurations from a 20-ns trajectory of the system were used for the PCA. They will be referred to as ''reference set'' hereafter. Overall translation and tumbling were removed by fitting each structure of the reference set to a set of atoms whose relative geometry stays relatively unchanged throughout: the backbone atoms of nucleosides 2–5 and 9–12 in each strand (nucleosides that are neither capping nor involved in the central basepair triplet considered for flipping). The PCA analysis was performed on the entire helix except the capping basepairs. The flooding potentials used to accelerate base flipping were built on combinations of these PCA eigenmodes, but the action of the flooding force was restricted to the backbone atoms of the flipping base. The removal of overall translation and tumbling, necessary for the correct computation of the flooding potential, was performed as for the PCA analysis.

Umbrella sampling was performed over 100 windows for each base and pathway. The reference structures for each window ($\mathbf{x}_i^{ref}$ in Eq. 11) were taken at regular time intervals along the pathway obtained from the corresponding flooding simulations. The trajectories were projected on the relevant optimal flipping mode for each base and pathway, as defined in Reweighting the PCA Eigenvectors to Describe Base Flipping ($\mathbf{q}$ in Eq. 11). Harmonic force constants of 200 kcal/mol Å^{-2} were found to ensure proper conformational overlap between windows. The starting atomic position and velocities for window $n+1$ were obtained by equilibrating the ending structure of window n for 200 ps, using the restraining potential of window $n+1$. The simulation time of the production trajectory for each window was 2 ns, amounting to 200 ns over the entire pathway. The WHAM binning was done versus the flipping angle θ to facilitate comparison with previous studies.

Nonlinear constrained optimizations were performed using the general nonlinear programming routines from the WNLIB software library (54). Numerical derivatives of the cost function were computed using a two-point scheme with extrapolation to higher orders (55). Since the optimization techniques employed are local in nature, 150 optimizations starting from random guesses were performed for each base to ensure a global exploration of the cost function landscape.

RESULTS AND DISCUSSION

Relationship between PCA eigenvectors and base flipping

Attempting to relate the base-flipping mechanism to one or a combination of PCA eigenvectors provides answers to two crucial questions: 1), whether the yet-to-be-defined generalized coordinate associated with flipping is sampled to some extent during a molecular dynamics simulation of the educt state (in which case it will be detected by PCA); and 2), whether a coupling exists between the low-frequency collective motions of the DNA stack (identified as most significant by PCA) and the flipping motion, the latter being much more localized.

To answer these questions, PCA analyses were carried out on two subsets of atoms of the reference set. The first subset comprised the entire DNA duplex with the exclusion of the capping basepairs, the second was restricted to the central triplet of basepairs that is used to compute the opening angle (Geometrical Quantification of Base Flipping). As expected, in both cases the first few eigenvectors correspond to large-scale collective motions, most importantly stack bending (two orthogonal modes, vectors 1 and 2) and groove breathing (vectors 3 and 4).

All structures from the reference set were projected onto each of the PCA eigenvectors. The difference in flipping angle computed between the two extremes of the projection indicates how much each individual PCA eigenmode contributes to the flipping during a typical MD simulation at room temperature. Fig. 3 compares this amount (shown as *bars*) to the corresponding PCA eigenvalue which, once divided by the sum of all eigenvalues, quantifies the fraction of atomic motion described by the mode during the same simulation. In the figure, this information is represented as the cumulative percentage of atomic motion over the PCA eigenmodes, which are ordered by decreasing eigenvalues (*line plots*). The plots show that taking into account the first few eigenvectors for each set suffices to account for a large amount of the atomic motion initially present in the reference set (15 eigenvectors for 80%, approximately). However, no such thing can be asserted for the DNA flipping motion of interest here, as demonstrated by the variation of the flipping angle along each mode. Although most eigenmodes do describe some fluctuation of the angle, none can be said to be directly related to it. Rather, the flipping motion appears to be smeared-out over a

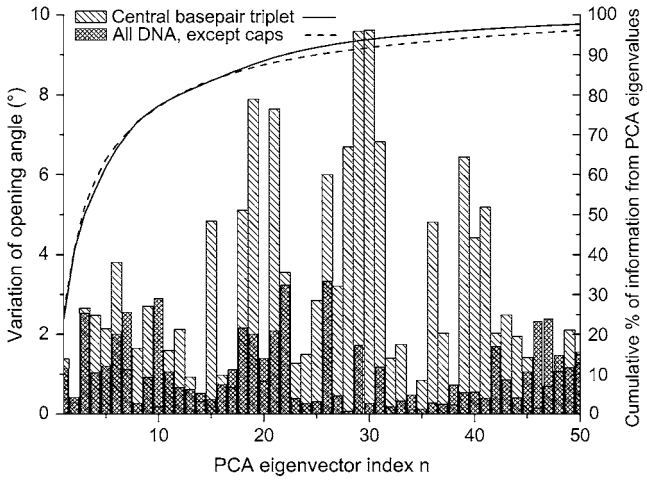

FIGURE 3 Results of the PCA analysis on a set of configurations taken from a 20 ns MD simulation. Two subsets of atoms are considered: the DNA duplex without the capping basepairs, and the central basepair triplet. The two curves represent the cumulative amount of atomic motion retained by taking into account the first n PCA modes only. The bar plots denote the absolute value of the variation of the flipping angle along each eigenmode.

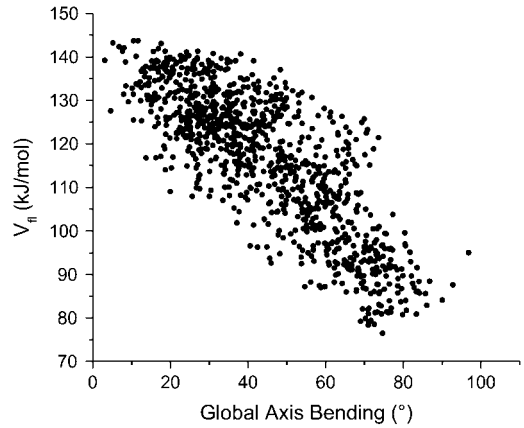

FIGURE 4 A flooding potential built upon the first five PCA eigenvectors, with $E_{fl} = 35$ kcal/mol, induces exaggerated bending of the DNA helix. Each point represents a snapshot from a flooding simulation and is associated with a global helical bend angle and a flooding potential value; these two parameters can be seen to exhibit opposite variations.

large number of modes, the most relevant of which feature indices from 15 to 45. In particular, the first 15 eigenvectors mentioned above are crucial for the global motion but have a modest impact on the flipping angle. It is worth mentioning that eigenmodes with very high indices that contribute very little to the overall motion also do not affect the flipping angle (indices above 80 in the present case, not shown in Fig. 3). However, the relationship to flipping decreases much more slowly with the eigenvector index than the eigenvalue.

To summarize, the PCA eigenvectors of the DNA helix can be grouped into three categories: the first 14 determine the large-scale helical motion, and have the highest weights in the flooding potential (Eq. 8); modes with indices comprised between 15 and 50 are more localized and relate optimally to base flipping, although none has a dominant influence over it; and, finally, the variance of the reference set along the remaining eigenvectors is so small that these are, in effect, negligible.

Thus, the PCA eigenvalues cannot be used as a suitable set of weights for building a flooding potential selectively inducing base flipping. This was confirmed by a series of short flooding simulations (<5 ns), acting on different test sets of PCA eigenvectors with varying flooding strengths. Including the first five eigenmodes in the flooding potential results in exaggerated bending motions eventually causing strand dissociation. Fig. 4 shows the opposite variation of flooding potential and axis bend for such a flooding simulation; as can be seen, extremely bent structures are achieved. In some cases, these were seen to exhibit spontaneous base flipping, but at bending angles of 80° or more this was deemed irrelevant to the biological process under study. This finding confirms the already stated relationship between axis bend and base flipping, further discussed in Structural Aspects of Base Flipping. However, since inducing reasonably low axis bend angles via flooding is not sufficient to accelerate base flipping to timescales compatible with MD simulations, other coupling factors probably come into play to provide a sufficient lowering of the free energy barrier. In turn, restricting the action of the flooding potential to the central basepair triplet has no noticeable effect on the dynamics of the DNA and was not observed to induce flipping.

These results provide an answer to the two questions raised above. First, no direct coupling between large-scale low-frequency deformations of the DNA helix and the occurrence of base flipping is observed during a typical MD simulation. Second, a large number of collective modes are weakly coupled with the flipping of the target base, but none can be said to dominate it.

Reweighting the PCA eigenvectors to describe base flipping

To promote the influence of modes with relatively small eigenvalues but a high contribution to flipping, an optimal generalized coordinate describing base flipping was defined as a linear combination of the PCA eigenmodes. This is effectively a basis change, from the PCA eigenmode basis to a new set in which base flipping can be described using one mode only. It can also be considered a reweighting of the PCA eigenmodes, from the PCA eigenvalues measuring each mode's relative contribution to the global atomic motion to a different set of weights that take into account the mode's contribution to flipping. The expression of the optimal coordinate as a function of the coefficients for the linear expansion c_i is given by Eq. 13; the actual values of c_i are determined using the local optimization procedure described in Optimizing a Linear Combination of PCA Eigenvectors to Describe

Flipping and the section called Computational Details. As mentioned in these sections, 150 local optimizations for each base, starting from random weight guesses, were performed to ensure a global exploration of the cost function landscape. Fig. 5 shows the average and standard deviation, over the 150 optimizations, of the squared weights of each PCA eigenvector (c_i^2 in Eq. 13) for each of the two bases under study. Also shown is the corresponding eigenvalue. Although optimizations starting from different random guesses may converge to different local minima, only a small number of PCA modes seem to be involved in each case. These are encountered, for each base, between indices 5 and 15 approximately: although they are relevant to the global motion and have rather large associated eigenvalues, they are not related to the low-amplitude stack deformations described earlier (first five eigenvectors). This explains the previously discussed failure to derive an adequate flooding potential directly from the first few eigenvectors. In fact, the PCA eigenvectors involved in the optimized modes may be seen as the best compromises between a high contribution to the global motion (high eigenvalue) and a tight relationship to variations in the flipping angle.

Solutions for the flipping of cytosine involve mainly three PCA eigenmodes (modes 6, 8, and 12). In the case of guanine, the number of relevant eigenmodes is significantly higher, but modes number 4–7, 11, and 12 can be considered paramount. A description of the nature of these PCA modes may be found in Table 1, and movies are available as Supplementary Material.

The aforementioned modes feature common trends that help explain their relationship to base flipping. The bending of the helix around the target base helps to reduce the stabilizing stacking interactions, lowering the barrier to flipping. The same can be said for the local increase in the rise parameter between the target base and its intrastrand neighbors. Transitions in the ε and ζ backbone dihedral angles, either on the 3′ or the 5′ side of the target base, increase the slide and shift distances between the target base and either its 3′ or 5′ neighbor, further reducing the stacking interactions. Interstrand distortions at the target basepair reduce the strength of the Watson and Crick hydrogen bonds, facilitating their breaking during the initial phase of the flipping motion. Deviation of structural parameters such as tip, buckling, and propeller from their canonical values contribute to this. Finally, broadening and narrowing of the grooves play an important role, since they determine the pathway that the flipping base is likely to adopt. In particular, the widening of the sterically encumbered minor groove is required for the flipping base to pass through it; this has been confirmed by recent experimental work (56).

The fact that these modes need to be combined to produce an effective flooding potential reflects the inherent complexity of the mechanism and the consequent difficulty of defining an order parameter to account for it. As can be seen in Table 1, many structural degrees of freedom, both local and global, are simultaneously involved, and the balance between the effect of these global and local variables appears quite subtle. The PCA modes used in this study provide a straightforward description of both the local and global factors involved in flipping, in contrast to the purely local definition adopted in previous theoretical works. The question of whether such an approach imposes more bias on the system is discussed in more detail in Free Energy Profiles of Base Flipping.

Designing a flooding potential

From the set of local solutions to the optimization problem described in Reweighting the PCA Eigenvectors to Describe Base Flipping (150 per base), four individual flooding potentials were constructed, each selectively expelling one of the two possible target bases (cytosine or guanine) through one of the grooves (major or minor), as now explained.

Since high-index PCA eigenvectors have almost no contribution to the overall motion (small λ_i in Eq. 13), their weights in the solution vector can vary with almost no impact

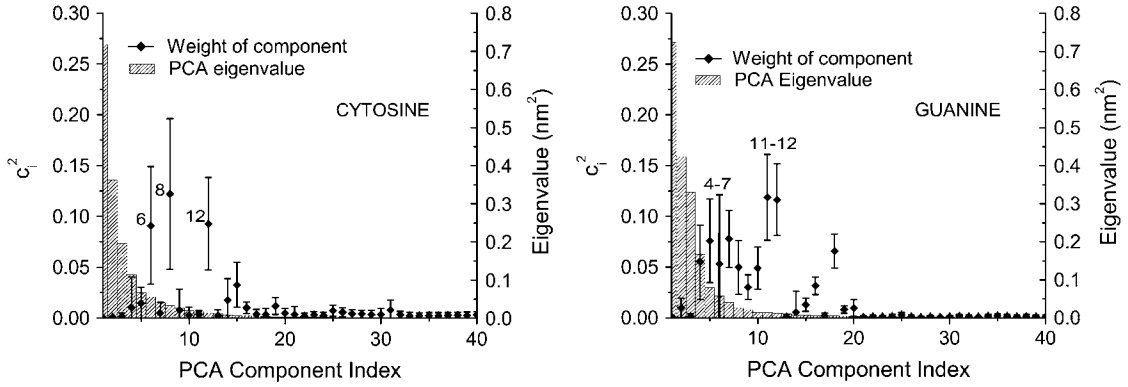

FIGURE 5 Synthetic results of 150 optimizations of the optimal flipping coordinate (see text), for cytosine (*left*) and guanine (*right*). The diamonds represent the average squared weight (over the 150 optimizations) of each PCA eigenvector within this coordinate (Eq. 13), the error bars the associated standard deviation. The corresponding PCA eigenvalues are shown as bars.

TABLE 1 Structural characterization of the most relevant PCA eigenmodes combined to obtain the optimized flipping coordinate, for the two bases under study (dC and dG)

Base	Index	Description
dC	6	Major groove narrows, minor groove widens. Target basepair tip (θ) increases.
	8	Bending of helix at target basepair, away from target base. Target base moves toward major groove. Local increase of rise around target base. Transitions in ε and ζ dihedrals.
	12	Bending of helix at target basepair, away from target base. Target base moves toward minor groove. Increase of rise between target base and 3′ neighbor. Transitions in 5′ neighbor ε and ζ dihedrals.
dG	5	Increase of tip at target basepair. Moderate buckling and propeller at target basepair. Transition in target base ε and ζ dihedrals. Increase of slide and shift with respect to 3′ neighbor.
	7	Antisymmetric movement of backbone: minor groove widens on 5′ side of target base and narrows on 3′ side. Major groove width shows opposite variation. Target base moves toward major groove. Transition in 5′ neighbor ε and ζ dihedrals.
	11	Similar to 7. Target base moves toward minor groove. Transition in 5′ neighbor ε and ζ dihedrals.
	12	Bending of helix at target basepair, away from target base. Minor groove widens around target basepair. Moderate propeller at target basepair.

on either the cost function or the global mode itself. On the contrary, important differences in the weights of the most relevant PCA components discussed above probably distinguish two different minima on the cost-function energy landscape, i.e., two potentially different flipping pathways. However, even the weights of these important modes can vary slightly between solutions describing the same mode, because of the action of the normalization constraint (Eq. 13). Hence, it is necessary to group the local minima found during the optimizations into classes corresponding to actually different flipping pathways.

For cytosine, this process yielded two solutions only. Constructing flooding potentials from each and applying either of them in a simulation resulted in reproducible base flipping, respectively through the major or the minor groove. For guanine, the situation is less clear-cut, since the collective motions that maximize the fluctuation of the flipping angle consist of a more complex combination of PCA eigenvectors. As before, however, two solutions could be isolated that result in reproducible base-flipping through each of the grooves.

Fig. 6 depicts the normalized contributions of the first 20 PCA eigenvectors to each of these four collective modes (Eq. 13). Interestingly, PCA modes associated with groove breathing (mode 6 for the flipping of cytosine modes, and 12 for guanine) have very different weights and opposite signs in the major and minor groove flipping pathways and can be considered discriminating factors in this respect.

As mentioned in Computational Details, each of the four flooding potentials acts selectively on one of the optimal flipping modes, and on that coordinate only, and the deriving force is applied exclusively to the backbone atoms of the flipping base. In this fashion, both the actual flipping and its triggering effects (local influence of bending and groove breathing) are simultaneously solicited, acting in a concerted fashion that probably plays an important role in minimizing the free energy barrier to flipping. The choice of the flooding potential applied during a simulation determines the target base and the flipping pathway followed. Table 2 lists the flooding strength and width parameters used in the flooding potentials. These values reproducibly accelerate the occurrence of base flipping to within 2 ns of simulation time.

Local minima on the free energy landscape

As mentioned in Conformational Flooding, the flooding potential locally destabilizes the system in its initial configuration (educt), facilitating egress from, and preventing return to, the corresponding attraction basin on the free energy surface. The action of the flooding potential decays rapidly outside the educt well, allowing free relaxation of the system to another attraction basin (product).

Applying the flooding potentials discussed above to the model DNA strand under study expels the system from the well corresponding to the initial closed-state geometry. However, the flipping base does not necessarily complete its opening motion during a typical simulation, but rather

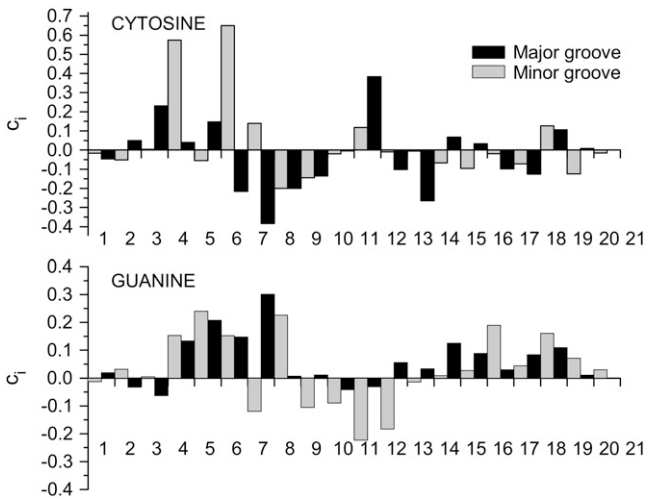

FIGURE 6 Contribution of the 20 first PCA eigenvectors to the four optimal generalized flipping coordinates (one per base and groove). The normalized weights of the eigenvectors in each case are shown as histograms (*shaded bars* for the minor groove pathway, *solid* for the major groove pathway).

TABLE 2 Parameters used for the flooding simulations, as defined in Eq. 8

	Global minimum			Local minima		
Base	E_{fl} (kcal/mol)	ω (kcal/mol)	m	E_{fl} (kcal/mol)	ω (kcal/mol)	m
dC	29	95	1	24	48	10
dG	36	167	1	24	48	10

E_{fl} scales the entire flooding potential, ω controls its width, and m is the number of global coordinates on which it acts.

remains in a number of intermediate states, between fully open and fully closed, that are stable at least on the nanosecond timescale and correspond to local minima on the free energy surface. This was concluded from series of 5 ns flooding simulations (20 for each base and pathway, with different starting velocities and closed-state structures).

Each local minimum is uniquely characterized by its well bottom, which is obtained by averaging the relevant part of the trajectory. Five such local minima were found in the case of cytosine, four for the flipping of guanine (Fig. 7).

For cytosine, all encountered local minima were located along the minor groove pathway, and can be loosely categorized into three different classes. The first class corresponds to structures where the target base, having broken its Watson and Crick pairing, remains sandwiched in the minor groove (Fig. 7 *e*). Further along the minor groove opening pathway, the target base interacts with the backbone of the opposing strand, spanning the minor groove which dilates around it (Fig. 7, *b–d*). Once the base has exited the minor groove, it may fold back and come in favorable interaction with the backbone of its own strand, on the 5′ side (Fig. 7 *a*).

Guanine, on the other hand, features three local minima on the major groove side, and one on the minor groove pathway. The latter (Fig. 7 *g*) is similar in nature to the sandwiched class previously mentioned in the case of cytosine. This structure is a good example of the rotation of the base around its glycosidic bond that seems to characterize the minor groove flipping pathway of guanine; the following paragraphs (see Local Minima on the Free Energy Landscape; and see The Flipped States) discuss this in more detail. Of the three major groove minima, one (Fig. 7 *f*) is encountered early on along the flipping pathway and can be considered to belong to the sandwiched class; the two remaining minima (Fig. 7, *h* and *i*) feature guanine residues that have exited the major groove and have folded back onto the backbone on the 5′ side. Stable conformations, where the target base spans the major groove, were not found for either base. In all cases, the remaining basepairs were not found to be disrupted by the flipping of the target base. In some conformations (Fig. 7, *b* and *i*) the former partner of the target base was found to interact with one of the neighboring basepair on the 3′ or 5′ side, compensating for the lost Watson and Crick interactions with the target.

These local minima feature major structural deviations to the fully flipped crystal structures (8,9), and correspond to metastable structures of the DNA which have not, to the best of our knowledge, been experimentally characterized yet. However, they probably have important implications in solution, where they could help interpret the experimentally determined opening rate and imino proton exchange equilibrium constant. Indeed, in the local minima described above, the imino proton of guanine is in contact with water molecules, either from the bulk solvent or from individual water molecules acting as bridges between former Watson and Crick hydrogen-bond partner atoms, as already mentioned for intermediate structures in previous studies (33). An extension of this work, comparing results from NMR and MD simulations in more detail, is currently under way. A recent study (27) has also revealed the crystal structure of an intermediate state (albeit with an abasic target nucleotide) flipped halfway through the major groove, in the presence of the HhaI methyltransferase enzyme, suggesting a possible role of the semi-flipped local minima for enzyme-assisted base flipping.

To let the flipping motion proceed to the fully flipped state, each of the aforementioned local minima was filled with a

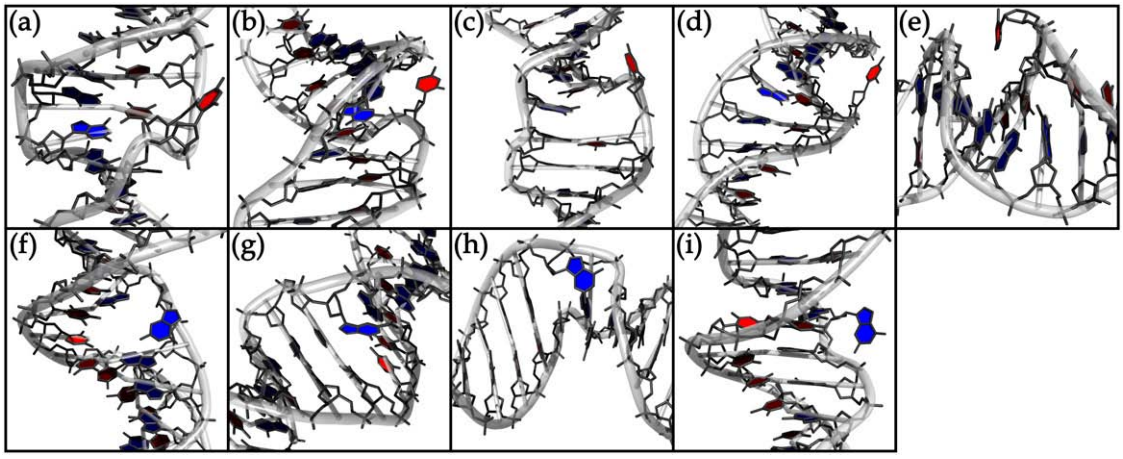

FIGURE 7 Structures of the local free energy minima encountered along the flipping pathways. Refer to text for details.

flooding potential, constructed from molecular dynamics simulations where the base remains in the corresponding local minimum. The first 10 PCA modes from these simulations, along with the corresponding PCA eigenvalues, were employed in each case. Enhancing sampling in such a 10-dimensional essential subspace was found to efficiently prevent the base from getting trapped in the local minimum, while not requiring presuppositions on the pathway followed. The flooding strengths applied were chosen so as to expel the target base from the minimum in less than a nanosecond. They turned out to be much smaller than those required to drive the base out of the closed-state global minimum (Table 2), reflecting the shallowness of the local minima compared to the global closed-state basin (for the global minimum the flooding strength only acts on one global coordinate, while for the local minima it is split between 10 modes).

Flooding the free energy wells of the global closed state and the local minima simultaneously in a simulation resulted in complete and reproducible base flipping. The pathway followed (minor or major groove) was determined by the initial flooding potential used (see Designing a Flooding Potential).

Free energy profiles of base flipping

The use of specific flooding potentials allowed reproducible flipping of each base of the G-C basepair in either groove, and served to define the flipping pathway followed by the system in each case. Free energy profiles along each path were obtained using umbrella sampling and the WHAM method (see Umbrella Sampling and the Weighted Histogram Analysis Method; and see Computational Details), to enhance the sampling of the high-energy regions. Comparing the corresponding flipping barriers is expected to provide valuable insights into the occurrence of flipping phenomena in biological systems, and the timescales on which they occur.

Fig. 8 shows the free energy profiles thus obtained, for the flipping of the guanine and cytosine bases through each of the grooves, as a function of the flipping angle. The initial free energy barrier to flipping occurs within a range of 20–25° around the initial closed-state flipping angle value of ~55°. The quadratic nature of the free energy profiles in this region, and the associated free energy increase of 6–10 kcal/mol, are in agreement with previous studies (33). For higher absolute variations of the flipping angle, the free energy surface exhibits a rather complex topology featuring broad minima only 2 kcal/mol higher in energy than the initial closed state. These results differ significantly from the previously published theoretical studies, which report either a linearly increasing (33) or rugged plateaulike (37) free energy profile. In the present work, the barrier to flipping is clearly defined since it is followed by a series of deep local minima, and can be directly related with the breaking of the Watson and Crick hydrogen bonds. The intervals of the flipping angle inside which each of these three hydrogen bonds exist are superimposed on Fig. 8

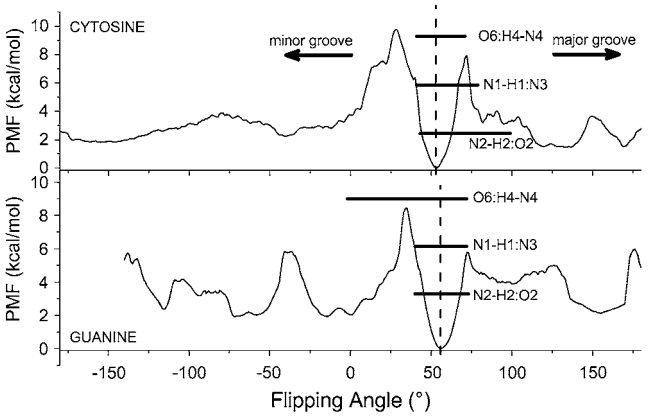

FIGURE 8 Free energy profiles associated with the flipping of cytosine (*top*) and guanine (*bottom*) through each of the grooves, as a function of the flipping angle. Closed-form DNA conformations occur for values at ~55° (*dashed line*). A positive (negative) variation of the angle denotes flipping through the major (minor) groove. The superimposed thick lines denote intact Watson and Crick hydrogen bonds between guanine and cytosine.

as thick horizontal lines, based on the geometrical criterion defined in Computational Details. While the three hydrogen bonds rupture simultaneously for the flipping of cytosine through the minor groove or the flipping of guanine through the major groove, one bond (N2-H2:O2 for cytosine and O6:H4-N4 for guanine) is retained much longer than the others when the flipping occurs through the opposite groove. This bond is the most distant from the glycosidic linkage where the flipping motion originates, and is thus the last to break. For guanine flipping through the minor groove, the O6:H4-N4 bond survives a variation of >50° of the flipping angle.

The barrier to flipping is significantly higher for the minor groove pathway for both bases, consistent with the increased sterical hindrance that has to be overcome. Surprisingly, despite the fact that guanine is bulkier than cytosine, the free energy barriers associated with the flipping of both bases through the minor groove are quite similar, whereas the major groove flipping barrier for guanine is actually lower than that of cytosine. The flipping of guanine through the minor groove was not seen to distort the groove more than is the case for cytosine: with its H2 atom interacting with the phosphorus atom on nucleotide $n+2$ (in $5' \rightarrow 3'$ sense), the base rotates around the backbone and exits the stack. The dihedral angle χ, which quantifies the rotation of the base out of the plane perpendicular to the helical axis, can be used to differentiate between the two mechanisms (Fig. 9). In the case of cytosine, the value of χ remains unmodified during the flipping through either groove; only the flipped structure exhibits major variations in χ. In contrast, the flipping of guanine through either groove is accompanied by a progressive increase (major groove) or decrease (minor groove) in χ. The guanine base departs from the plane it was initially forming with its Watson and Crick partner and creates favorable interactions with the

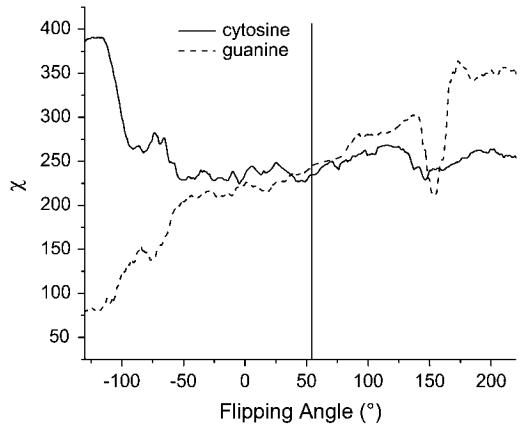

FIGURE 9 Variation of the χ dihedral angle during the flipping of the guanine and cytosine bases. The thick vertical line materializes the average closed-state conformation.

backbone of its own strand, at the same time minimizing the sterical barrier to flipping. This may tentatively explain the low free energy barriers observed for guanine, especially through the minor groove.

The free energy barriers to base flipping determined in this work are in good agreement with the corresponding value of 9 kcal/mol measured in NMR experiments (57). It should be stressed that the experimental result is only valid under the assumption of a very simple two-state model (closed and flipped, the latter only permitting the exchange of the imino proton); on the contrary, the existence of local minima discussed in Local Minima on the Free Energy Landscape, suggests a more complex mechanism. The aforementioned agreement could hence be partly fortuitous. In addition, in our work a clear definition of the barrier is rendered possible by the low-energy region that we have found to lie beyond it. As this is in contrast with previous theoretical studies, and for lack of experimental evidence for a very stable flipped state, the caveats formulated in the next paragraph concerning this part of the profile should also be considered to apply to the barrier itself.

The ulterior part of the flipping pathway is characterized by a very complex free energy surface. This is also the part where existing works diverge by the largest extent. The main finding of this study is that a very stable, albeit broad and not precisely defined, local minimum (or rather ensemble of local minima) exists in this area. The previously cited studies did not identify such a minimum: while shallow minima may be found, especially for the flipping of guanine (37), they are on the order of a few kcal/mol deep and much higher in energy (10 (33) to 20 (37) kcal/mol above the closed state). This is an important discrepancy, which we now address in more detail.

The experimental population ratio between the closed and flipped conformations corresponds to a free energy difference of 6–7 kcal/mol (11), which is markedly higher than our figure of 2 kcal/mol. Two reasons can be invoked. First, our free energy profile might suffer from artifacts, for instance caused by the multiatom restraint potential employed. While we do not think this is the case, for reasons detailed below, we cannot completely rule it out, and other valid methodological points (sampling, force-field issues) could also apply. Second, the experimentally detected flipped state population (able to exchange its imino proton with water) might in fact represent a subset of the possible flipped conformations only. Interest for this hypothesis was renewed following recent NMR results (C. Griesinger and H. Schwalbe, 2006, private communication) in which the lines of exchangeable imino protons of the flipping bases were seen to be broadened by 40–60 Hz compared to the line width of exchangeable protons ($<$10 Hz). This can be explained either by a flipped state population higher than 0.001%, or by the existence of a third state of unspecified structure but verifying the following properties: 2–3 kcal/mol higher in energy than the closed state, separated from the latter by a barrier of $\sim$16 kcal/mol, and unable to exchange protons with the solvent. We stress that these preliminary experimental results are not presented to justify our free energy profile: although conformations where the flipped base is in contact with the helical backbone (e.g., Fig. 7 *h*) could possibly fulfill these conditions, there is at present no convincing evidence that this is effectively the case. We mention these results solely to outline the fact that both the experimental and the theoretical characterizations of base flipping are challenging and to stress the necessity of additional research, especially employing novel theoretical or experimental approaches as these become available.

As previously mentioned, our profile also differs from previous umbrella sampling studies by a sizeable barrier to closing, where these works find this barrier to be almost nonexistent. During our unbiased MD simulations of the flipped states (5 $\times$ 10 ns for each base, with varying starting structures), the flipped base was never seen to reintegrate the helix. While this is in part due to an unfavorable entropic effect reminiscent of protein folding (as discussed further down in The Flipped States, the plasticity of the flipped state is very high), it also fits the presence of a quantitative barrier to closing.

Finally, we shortly discuss the conceptual difference of our global-mode restraint compared to the angular restraint employed in previous studies. Even though we impose our restraints only on the backbone atoms of the flipping base, the number of affected atoms is higher in our scenario. However, as demonstrated with PCA, these atoms actually move in a concerted manner along a generalized coordinate, rather than independently from one another. Our restraint operates on this generalized degree of freedom only, leaving the others untouched; up to a basis change, our approach is identical to traditional umbrella sampling. Naturally, it is only valid under the assumption that the pathway can be described, in a piecewise manner, by a collection of linear global coordinates (combinations of PCA eigenvectors), which is difficult to verify

in practice. However, we think that including relatively slow modes (bending, breathing) in the restraint should be regarded as a positive point: as previously stated (22) and in agreement with our findings, these modes clearly participate in the flipping, and are probably quite difficult to equilibrate at each window if left out of the umbrella coordinate, unless very extensive windows are computed.

While this work offers new input on the pathways and free energy profiles associated with base flipping, it actually raises more questions on the topic than it is able to answer. It is our hope that it will provide incentive for applying novel theoretical methods to this difficult problem, which still remains beyond the scope of state-of-the-art computational methods.

Structural aspects of base flipping

Closer analysis of the backbone dihedral angles reveals modified behaviors or transitions for the flipping nucleotide and its 3′ and 5′ neighbors, compared to typical B-DNA values. Most angles do not show reproducible transitions directly related to the flipping phenomenon, but rather increased fluctuations and occasional transitions as a consequence of it. On the other hand, some dihedral angles exhibit reproducible transitions during certain phases of the opening process. The values of these parameters have been extracted from the flipping simulations and binned by flipping angle values, in Fig. 10.

The flipping of both guanine and cytosine can be related to three dihedral angles on the backbone of the target base: ε, ζ, and γ (Fig. 10 *a*). Transitions in these parameters are especially noticeable for the major groove pathway; the behavior of ζ especially seems exclusively related to this mechanism. As a matter of fact, the flipping of a base through the minor groove requires more extensive structural modifications of the helix and involves backbone dihedral transitions at the 3′ and 5′ neighbors of the target base; this can be seen in Fig. 10, *b*, *c* (5′ side), and *d* (3′ side). The 5′ ε for the minor

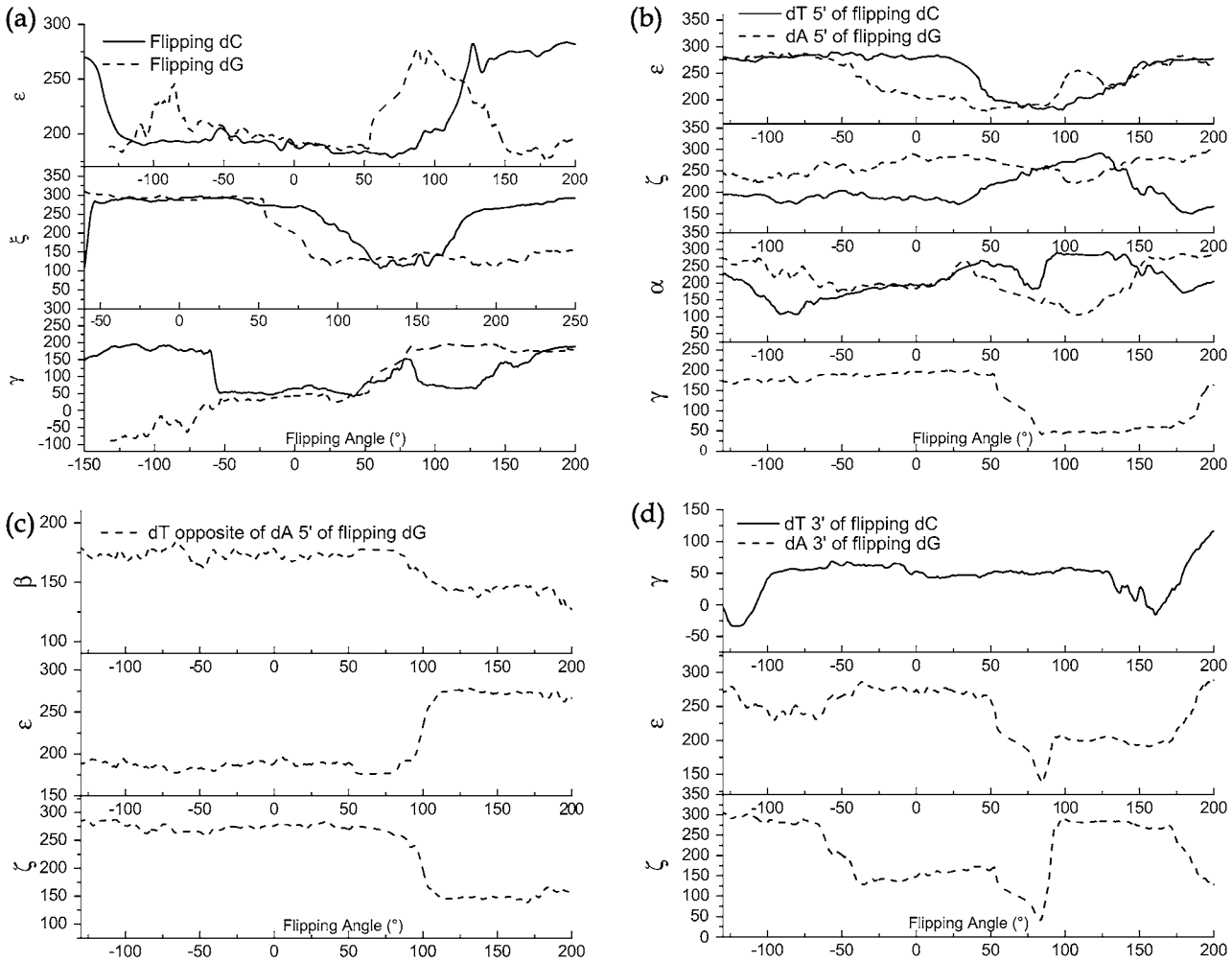

FIGURE 10 Values of relevant backbone dihedral angles as a function of the flipping angle. (*a*) ε, ζ, and γ dihedral angles of the flipping nucleoside; (*b*) ε, ζ, α, and γ of the nucleoside 5′ of the flipping base; (*c*) β, ε, and ζ of the deoxythymidine residue opposite the deoxyadenosine residue, 5′ of the flipping guanine; and (*d*) γ, ε, and ζ of the nucleoside 3′ of the flipping base. See text for details.

groove flipping of cytosine, and the 3′ γ for guanine are especially relevant. The 3′ ζ dihedral, proposed in previous works (58,59) as a good monitor of the flipping process, was found to be reproducibly involved in the flipping of the guanine base only (Fig. 10 *d*).

In most cases, the backbone conformation of the opposing strand was not found to undergo major or reproducible changes during flipping. This is especially remarkable for the former Watson and Crick partner of the flipping base. Only for the flipping of guanine through the major groove were there found transitions in the β, ε, and ζ dihedrals at the adenosine opposite the 5′ neighbor of the target.

The phase of the sugar puckers in the region affected by flipping has previously been hinted at as possibly relevant (33). However, in this study, no direct linkage could be found between the two phenomena. The puckering of the sugar for the flipped base undergoes more frequent transitions between south and north-type conformations than is the case in the closed state (south-type pucker characteristic of B-DNA), but this appears as a consequence of flipping rather than a cause for it. The major groove pathways were found to be more prone to such transitions than the minor groove ones, for either base.

Axis bend has already been related to base flipping in previous studies, both theoretical (22) and experimental (21,56,60). Relationship Between PCA Eigenvectors and Base Flipping mentions simulations where, after inducing extreme helical bending (>80°) using a flooding potential, occurrences of spontaneous base flipping were seen. This is obviously related to the reduction of the π-stacking interactions that stabilize the closed helix (61). Inducing a less drastic helical bend using a milder flooding potential could possibly reduce the timescale of flipping to computationally tractable values, although this has not been investigated further here. The relationship of flipping and bending is shown in Fig. 11. Both pathways, but most importantly the minor groove one, involve quantitative bending of the helix. This should be interpreted both as a cause and a consequence of flipping: bending favors flipping by reducing stacking between basepairs, while the flipped base introduces a locus of flexibility in the helix, facilitating further bending. Hence, fully flipped states display a wide variety of bending angles whose average has no physical meaning, and is therefore not shown in Fig. 11.

Finally, an attempt to relate other local or helical parameters to the flipping of the target base has shown that very few of these parameters exhibit reproducible behaviors beyond the onset of the mechanism. This onset was found to be most conveniently described using a combination of the buckling (κ), opening (σ), and shear (S_x) parameters (Fig. 12), for absolute variations of the flipping angle <50° around the closed-state value. It is likely that building a flooding potential on a combination of these three elementary motions would provide an alternative approach to induce flipping in DNA helices.

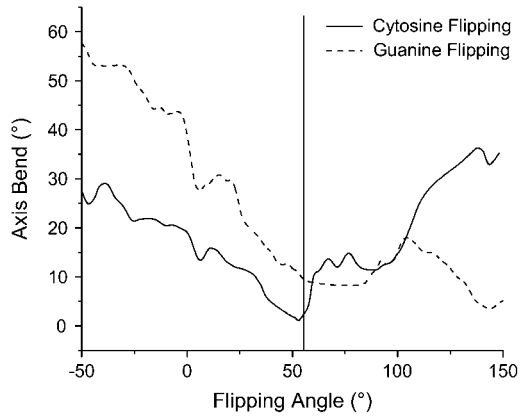

FIGURE 11 Global axis bending angle of the DNA helix as a function of the flipping angle. The thick vertical line materializes the average closed-state conformation.

The flipped states

The two possible flipping pathways for each base (major or minor groove) do not lead to identical flipped states. This is already apparent from the evolution of the backbone dihedral angles during the flipping motion (Fig. 10). Further evidence is provided by the RMSD between the reoptimized endpoint structures for each pathway; for the backbone atoms of the flipping base and its two intrastrand neighbors, this amounted to 2.14 Å for cytosine and 2.30 Å for guanine. This points at the insufficiency of the (2π-periodic) flipping angle θ to discriminate important features in the open states, and stresses the necessity for additional structural parameters. During subsequent MD simulations on the flipped states, however, conformational transitions between endpoint structures were observed on the nanosecond timescale.

As previously stated, the analysis of the local structural parameters of the DNA helix during and after the flipping process revealed increased fluctuations and transitions as

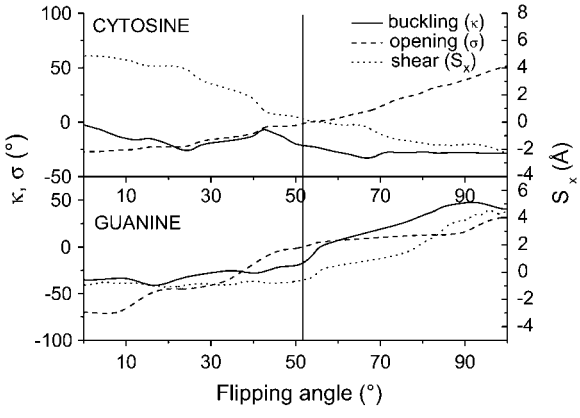

FIGURE 12 Variation of the buckling, opening (°) and shear (A) parameters as a function of the flipping angle. The thick vertical line materializes the average closed-state conformation.

compared to typical B-DNA simulations. Not unlike the unfolded state of proteins, the flipped conformation of DNA is quite polymorphic and characterized by high entropy. The associated structural heterogeneity contributes to the rough topology of the free energy surface in the corresponding region. Notably, frequent A- to B-DNA-type transitions are seen in the backbone structural parameters and sugar puckers. Even though this conformational freedom is restricted to the flipped base and its nearest neighbors, it also impacts global structural parameters such as helical bend or groove dimensions. As discussed, it is very likely that this plasticity complicates the experimental characterization of the flipped state by structural techniques, as well as the assessment of its relative stability.

The rotation of the base around the glycosidic bond, generally referred to as the χ dihedral angle, displays frequent transitions in the flipped state. In this paragraph, rather than computing χ, we measure the angle between the normal vector to the base's plane and the local helical axis, which is 0° for the closed state. From the combined population density of MD simulations on the flipped states (50 ns in total for each base), a free energy profile along this angle was computed and is plotted, for each of the bases, in Fig. 13. The position of the global minimum differs significantly between cytosine, which tends to rotate perpendicularly to its initial plane in the closed state, and guanine, which adopts an intermediate position characterized by a 20° rotation from its initial orientation. Cytosine also features a very shallow local minimum at this angle. The barrier to rotation in each case is quite low (typically 1.5 kcal/mol) and is easily overcome during room-temperature MD simulations on the nanosecond timescale.

SUMMARY AND CONCLUSION

Base flipping in DNA is a slow process that requires enhanced sampling techniques of conformational space to be studied theoretically. In this work, we combine principal component analysis and conformational flooding to this effect. Adding an artificial destabilizing potential centered on the initial closed-state B-DNA structure accelerates the egress of the system from this attraction basin into tractable time spans. Because a suitable flooding potential could not be built directly from a straightforward application of principal component analysis, the potential was tailored to act on a flipping coordinate identified as an optimal linear combination of the PCA modes that contribute the most to base-flipping. Four individual flooding potentials were thus constructed, driving either of the two bases studied reproducibly out of the stack, through either of the grooves. The number of PCA eigenmodes having non-negligible weights in the optimal linear combinations was found to be quite small. These modes describe both global motions (helical bending, groove breathing) and more localized deformations (increase of rise, propeller, or buckling around the central triplet of basepairs) that lower the energetic penalty to flipping by reducing the stacking and Watson and Crick hydrogen bonding interactions which stabilize the closed state.

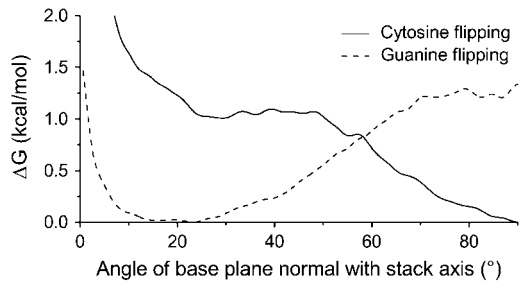

FIGURE 13 Free energy profile for the orientation of the flipped base with respect to the helical axis; at 0° the base plane is perpendicular to the helical axis.

Important structural aspects of base flipping were observed and discussed. First, local minima on the free energy surface were identified. It is likely that these are relevant to base flipping in isolated DNA, and should be included in the model used to interpret the NMR data measuring the flipping timescales and free energy barriers. Although more evidence is clearly desirable, the popular two-state model for base flipping appears oversimplified in the light of both this work and recent experimental results.

Certain structural parameters, such as backbone dihedral angles, were also found to be involved in the flipping mechanism; increased variations and more frequent transitions were found to characterize most of the remaining structural parameters. The importance of helical bend was also confirmed and structurally quantified.

Finally, the free energy profiles associated with base flipping were computed. Most notable in this respect is the existence of a broad but deep attraction basin associated with the flipped state, in contradiction with previous theoretical studies; we hope that additional evidence from the experimental side will help pinpoint the origin of this apparent discrepancy. The flipped state itself was found to exhibit a large structural heterogeneity, actually consisting of a wide array of interconverting conformations not unlike the unfolded state of a protein. Conformational transitions involve local structural parameters in the backbone around the flipped base, but also global ones like helical bend. The barrier to flipping was found to be of similar height for both bases and both possible pathways, a fact attributed to the rotation of the base around the glycosidic bond which facilitates its egress through the minor groove. This work also provides evidence that the reverse process (i.e., the flipping of the base back to its Watson and Crick position) is hampered by a sizeable free-energy barrier.

The ability of the conformational flooding method to accelerate slow transitions and reveal the associated pathway, already put to the test in protein systems, plays a central role in the current study. Here, however, one additional step is

required subsequent to the usually employed principal component analysis, namely the combining of the PCA eigenvectors. This became necessary due to the localized nature of the conformational transition under study, as compared to large-scale transitions more frequently observed in proteins. This generalization not only allowed us to focus the flooding potential on motions of smaller amplitude, but also provided valuable information about the very weak couplings that exist between the large-amplitude motions of the DNA helix and the local variables related to flipping, eventually yielding the optimal combination of modes maximizing this coupling. The small number and the nature of the modes thus combined also shed light on the mechanism behind base flipping. However, the flipside of the coin is that the additional step introduces a bias in the study, since the optimization attempts to maximize the variation of an arbitrarily chosen flipping angle (θ) during a simulation. Nonetheless, we deem this bias to be smaller than the one introduced by using this angle directly to define the entire flipping pathway, especially since the flipping angle is better suited to describe the onset of the flipping motion than to account for its latter part. Using a combination of flooding potentials, each with a local action on conformational space, should provide a better approximation of the actual pathway of base flipping.

The current methodology could nevertheless be improved in two major directions. First, dealing with a highly complex free energy surface featuring numerous local minima requires using multiple flooding potentials. Finding the right parameters for these by trial-and-error can rapidly become tedious, while suboptimal choices for these parameters will increase the impact of unwanted nonequilibrium effects on the simulations. For these cases, an adaptive scheme (e.g., similar in spirit to that proposed by (62)) could in principle fill up each major attraction basin on the landscape with a correctly parameterized flooding potential, yielding a completely flat free energy surface where all conformations are equiprobable. For base flipping, this would result in the system going back and forth several times between fully closed and fully open states during a typical simulation, providing an optimal sampling of the entire pathway. This could not be achieved here: once it has left the initial closed conformation, the system samples an extended conformational space but was never seen (and could not be made using flooding) to return to the fully closed state. As discussed in The Flipped States, the fully flipped state has a high entropy. The related region of the free energy surface is quite complex and probably consists of an important number of very shallow minima inside a very broad attraction basin. Manually compensating for such a complex landscape using flooding potentials proved unfeasible in practice. An adaptive procedure to retrieve accurate flooding strengths and widths from simulations would hence extend the scope of the method to even more complex and interesting biological problems. A possible implementation of such a framework was recently presented by Fukunishi et al. (63).

The second direction along which methodological improvements would be worthwhile is the fact that flooding is currently performed along a combination of PCA eigenvectors, which are linear by definition. Driving a system along a curvilinear coordinate (such as the one involved in flipping) currently requires the use of several flooding potentials, placed along the trajectory. A single flooding potential along a curvilinear coordinate would be a great improvement, both on theoretical and practical grounds, and would result in a novel approach to the determination of the mechanism and free energy barriers of enzyme-assisted base flipping, also taking into account the conformational transition of the enzyme itself.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors express their gratitude to Richard Lavery and Christian Griesinger for fruitful discussions. Additional thanks go to C. Griesinger for providing access to his unpublished NMR results.

REFERENCES

1. Daujotyte, D., S. Serva, G. Vilkaitis, E. Merkiene, D. Venclovas, and S. Klimasauskas. 2004. HhaI DNA methyltransferase uses the protruding Gln^{237} for active flipping of its target cytosine. *Structure*. 12: 1047–1055.
2. Klimasauskas, S., S. Kumar, R. J. Roberts, and X. D. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell*. 76:357–369.
3. Klimasauskas, S., and R. J. Roberts. 1995. Disruption of the target G-C base-pair by the HhaI methyltransferase. *Gene*. 157:163–164.
4. Klimasauskas, S., T. Szyperski, S. Serva, and K. Wuthrich. 1998. Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution. *EMBO J*. 17:317–324.
5. Slupphaug, G., C. D. Mol, B. Kavli, A. S. Arvai, H. E. Krokan, and J. A. Tainer. 1996. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature*. 384: 87–92.
6. Stivers, J. T., K. W. Pankiewicz, and K. A. Watanabe. 1999. Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry*. 38:952–963.
7. Luo, J., and T. C. Bruice. 2005. Low-frequency normal mode in DNA HhaI methyltransferase and motions of residues involved in the base flipping. *Proc. Natl. Acad. Sci. USA*. 102:16194–16198.
8. O'Gara, M., J. R. Horton, R. J. Roberts, and X. D. Cheng. 1998. Structures of Hhal methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol*. 5:872–877.
9. Parikh, S. S., G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn, and J. A. Tainer. 2000. Uracil-DNA glycosylase-DNA substrate and product structures: conformational strain promotes catalytic efficiency by coupled stereoelectronic effects. *Proc. Natl. Acad. Sci. USA*. 97:5083–5088.
10. Huang, N., N. K. Banavali, and A. D. MacKerell. 2003. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. USA*. 100:68–73.
11. Gueron, M., and J. L. Leroy. 1995. Studies of basepair kinetics by NMR measurement of proton exchange. *In* Nuclear Magnetic Resonance And Nucleic Acids. Academic Press, San Diego, CA.

12. Leijon, M., and A. Graslund. 1992. Effects of sequence and length on imino proton-exchange and basepair opening kinetics in DNA oligonucleotide duplexes. *Nucleic Acids Res.* 20:5339–5343.

13. Leroy, J. L., M. Kochoyan, T. Huynhdinh, and M. Gueron. 1988. Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J. Mol. Biol.* 200:223–238.

14. Snoussi, K., and J. L. Leroy. 2001. Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry.* 40:8898–8904.

15. Neely, R. K., D. Daujotyte, S. Grazulis, S. W. Magennis, D. T. F. Dryden, S. Klimasauskas, and A. C. Jones. 2005. Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Res.* 33:6953–6960.

16. O'Neil, L. L., and O. Wiest. 2005. A selective, noncovalent assay for base flipping in DNA. *J. Am. Chem. Soc.* 127:16800–16801.

17. Lyakhov, I. G., P. N. Hengen, D. Rubens, and T. D. Schneider. 2001. The P1 phage replication protein RepA contacts an otherwise inaccessible thymine N3 proton by DNA distortion or base flipping. *Nucleic Acids Res.* 29:4892–4900.

18. Schneider, T. D. 2001. Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucleic Acids Res.* 29:4881–4891.

19. Sommer, N., R. Depping, M. Piotrowski, and W. Ruger. 2004. Bacteriophage T4 α-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. *Biochem. Biophys. Res. Commun.* 323:809–815.

20. Fuxreiter, M., M. Luo, P. Jedlovszky, I. Simon, and R. Osman. 2002. Role of base flipping in specific recognition of damaged DNA by repair enzymes. *J. Mol. Biol.* 323:823–834.

21. Hopkins, B. B., and N. O. Reich. 2004. Simultaneous DNA binding, bending, and base flipping—evidence for a novel M. EcoRI methyltransferase-DNA complex. *J. Biol. Chem.* 279:37049–37060.

22. Ramstein, J., and R. Lavery. 1988. Energetic coupling between DNA bending and basepair opening. *Proc. Natl. Acad. Sci. USA.* 85:7231–7235.

23. Stivers, J. T. 2004. Site-specific DNA damage recognition by enzyme-induced base flipping. *In* Progress in Nucleic Acid Research and Molecular Biology, Vol 77. Academic Press, San Diego, CA.

24. Keepers, J. W., P. A. Kollman, P. K. Weiner, and T. L. James. 1982. Molecular mechanical studies of DNA flexibility—coupled backbone torsion angles and base-pair openings. *Proc. Natl. Acad. Sci. USA.* 79:5537–5541.

25. Malta, E., G. F. Moolenaar, and N. Goosen. 2006. Base flipping in nucleotide excision repair. *J. Biol. Chem.* 281:2184–2194.

26. MacDonald, D., G. Demarre, M. Bouvier, D. Mazel, and D. N. Gopaul. 2006. Structural basis for broad DNA-specificity in integron recombination. *Nature.* 440:1157–1162.

27. Horton, J. R., G. Ratner, N. K. Banavali, N. Huang, Y. Choi, M. A. Maier, V. E. Marquez, A. D. MacKerell, and X. D. Cheng. 2004. Caught in the act: visualization of an intermediate in the DNA base flipping pathway induced by Hhal methyltransferase. *Nucleic Acids Res.* 32:3877–3886.

28. Torrie, G. M., and J. P. Valleau. 1977. Non-physical sampling distributions in Monte Carlo free energy estimation—umbrella sampling. *J. Comput. Phys.* 23:187–199.

29. Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* 13:1011–1021.

30. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1995. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.* 16:1339–1350.

31. Bernet, J., K. Zakrzewska, and R. Lavery. 1997. Modeling basepair opening: the role of helical twist. *Theochem. J. Mol. Struct.* 398:473–482.

32. Giudice, E., and R. Lavery. 2003. Nucleic acid basepair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc.* 125:4998–4999.

33. Giudice, E., P. Varnai, and R. Lavery. 2003. Basepair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* 31:1434–1443.

34. Giudice, E., P. Varnai, and R. Lavery. 2003. Basepair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* 31:2703.

35. Giudice, E., P. Varnai, and R. Lavery. 2001. Energetic and conformational aspects of A·T base-pair opening within the DNA double helix. *ChemPhysChem.* 2:673–677.

36. Ramstein, J., and R. Lavery. 1990. Basepair opening pathways in B-DNA. *J. Biomol. Struct. Dyn.* 7:915–933.

37. Banavali, N. K., and A. D. MacKerell. 2002. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* 319:141–160.

38. Huang, N., and A. D. MacKerell. 2005. Specificity in protein-DNA interactions: energetic recognition by the (cytosine-C5)-methyltransferase from HhaI. *J. Mol. Biol.* 345:265–274.

39. MacKerell, A. D. 2003. Computational studies of base flipping in DNA alone and bound to the cytosine-5-methyltransferase from Hhal. *Abstr. Papers Am. Chem. Soc.* 226:U431–U431.

40. Priyakumar, U. D., and A. D. MacKerell. 2006. Base flipping in a GCGC containing DNA dodecamer: a comparative study of the performance of the nucleic acid force fields, CHARMM, AMBER, and BMS. *J. Chem. Theor. Comput.* 2:187–200.

41. Priyakumar, U. D., and A. D. MacKerell. 2006. Computational approaches for investigating base flipping in oligonucleotides. *Chem. Rev.* 106:489–505.

42. Grubmuller, H. 1995. Predicting slow structural transitions in macromolecular systems—conformational flooding. *Phys. Rev. E.* 52:2893–2906.

43. Schulze, B. G., H. Grubmuller, and J. D. Evanseck. 2000. Functional significance of hierarchical tiers in carbonmonoxy myoglobin: conformational substates and transitions studied by conformational flooding simulations. *J. Am. Chem. Soc.* 122:8700–8711.

44. Muller, E. M., A. de Meijere, and H. Grubmuller. 2002. Predicting unimolecular chemical reactions: chemical flooding. *J. Chem. Phys.* 116: 897–905.

45. Diekmann, S. 1989. Definitions and nomenclature of nucleic-acid structure parameters. *J. Mol. Biol.* 205:787–791.

46. Wang, J. M., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–1074.

47. Berendsen, H. J. C., D. Vanderspoel, and R. Vandrunen. 1995. GROMACS—a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* 91:43–56.

48. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. (Online).* 7:306–317.

49. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an *N*.Log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089–10092.

50. Berendsen, H. J. C., J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

51. Hess, B., H. Bekker, H. J. C. Berendsen, and J. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18: 1463–1472.

52. Macke, T. J., and D. A. Case. 1998. Modeling unusual nucleic acid structures. *In* Molecular Modeling Of Nucleic Acids, ACS Symposium Series. 379–393.

53. Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids—conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

54. Nailor, W., and B. Chapman. WNLIB. http://www.willnaylor.com/wnlib.html

55. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1988. Numerical Recipes in C. Cambridge University Press, New York.

56. Horton, J. R., X. Zhang, R. Maunus, Z. Yang, G. G. Wilson, R. J. Roberts, and X. D. Cheng. 2006. DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.* 34:939–948.

57. Dornberger, U., M. Leijon, and H. Fritzsche. 1999. High basepair opening rates in tracts of GC basepairs. *J. Biol. Chem.* 274:6957–6962.

58. Chen, Y. Z., V. Mohan, and R. H. Griffey. 1998. Effect of backbone ζ-torsion angle on low energy single base opening in B-DNA crystal structures. *Chem. Phys. Lett.* 287:570–574.

59. Chen, Y. Z., V. Mohan, and R. H. Griffey. 1998. The opening of a single base without perturbations of neighboring nucleotides: a study on crystal B-DNA duplex d(CGCGAATTCGCG)(2). *J. Biomol. Struct. Dyn.* 15:765–777.

60. Su, T. J., M. R. Tock, S. U. Egelhaaf, W. C. K. Poon, and D. T. F. Dryden. 2005. DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.* 33:3235–3244.

61. Nakano, S., Y. Uotani, K. Uenishi, M. Fujii, and N. Sugimoto. 2005. DNA base flipping by a basepair-mimic nucleoside. *Nucleic Acids Res.* 33:7111–7119.

62. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA.* 99:12562–12566.

63. Fukunishi, Y., Y. Mikami, and H. Nakamura. 2003. The filling potential method: a method for estimating the free energy surface for protein-ligand docking. *J. Phys. Chem. B.* 107:13201–13210.