Conserved Patterns in Backbone Torsional Changes Allow for Single Base Flipping from Duplex DNA with Minimal Distortion of the Double Helix

Nilesh K. Banavali,*,† Niu Huang,‡ and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, 20 Penn St, Baltimore, Maryland 21201

Base flipping is a structural mechanism common to many DNA processing and repair enzymes. Changes in the local backbone torsions that occur during base flipping and the effect of environment on their behavior are of particular interest in understanding different base flipping mechanisms. In the present study, structures sampled during umbrella sampling molecular dynamics (MD) simulations of base flipping in aqueous and protein-bound environments, carried out with two different MD simulation strategies, are analyzed to find the most significant phosphodiester backbone distortions in the vicinity of the flipping base. Torsional sampling on the 5′ side of the flipping base during flipping through the major groove shows similarities to the torsional sampling on the 3′ side during flipping through the minor groove and vice versa. In differing environments, this behavior varies only marginally. These compensating torsional changes in the DNA backbone on 5′ and 3′ sides of the flipping base limit overall distortion of the DNA double helix during single base flipping. Rotameric intermediate states observed during base flipping are identified and postulated to be metastable states implicated in both large-scale structural changes and functional effects of chemical modifications in DNA.

Introduction

Hydrogen bonds between paired bases (e.g., Watson–Crick interactions) help to stabilize the DNA double helix, but they also partially obscure the individual stacked bases from the environment. Although the chemical identity of the bases can be established through major or minor groove interactions, for catalysis of chemical modification or repair of the bases, structural distortion away from the duplex state is required. Base flipping is one such commonly used structural mechanism seen in multiple crystal structures of DNA–protein complexes where the flipped state of a single base is stabilized in the enzyme active site.[1–5] The details of the intermediates leading from the base stacked state to the final flipped state are difficult to characterize experimentally, because they involve metastable structures that are not significantly populated. Chemical substitutions that trap intermediates by making these usually transient states more stable than either the standard double helical state or the fully flipped state can sometimes highlight their detailed structural features,[6–8] but these still represent only snapshots along the flipping pathway.

Umbrella sampling molecular dynamics (MD) simulations have been used to investigate the energetic basis of the structural distortion of DNA associated with DNA base flipping in aqueous solution.[9–11] This approach allows for visualization of the base flipping intermediates not typically accessible to experimental methods. The relative free energy differences between base stacked and flipped states obtained through these investigations show near-quantitative agreement with experiments probing base pair dynamics.[9,12–15] In the presence of a suitable protein environment, such as that provided by the cofactor-bound cytosine-5-methyltransferase from *Hha*I (**M**.*Hha*I), base flipping can be facilitated through destabilization of all states (including the original Watson–Crick paired stacked state) relative to the completely flipped state, through a balance between protein interactions and solvent exclusion.[16] Recognition of specific DNA sequences by **M**.*Hha*I is related to the sequence-dependent variability in protein facilitation of the base flipping process with the cognate sequence being the one that experiences the greatest extent of such facilitation.[17]

In the present study, analysis of all torsions in the local backbone region between the two sugars on either side of the flipping base is performed to identify the specific patterns in their distortions that occur during flipping. This focus on internal degrees of freedom of DNA is necessary to identify the details of the structural pathway utilized for base flipping. The energetic aspects of the interactions with the environment[16–18] and with the extended neighboring DNA sequence itself,[9] while included explicitly in the MD simulations, have been addressed before and are not actively considered in the present analysis. The flipping of a central (underlined) cytosine base in the **M**.*Hha*I cognate sequence GTCAG<u>C</u>GCATGG has been studied[16] using MD studies in four different environmental conditions: (a) in aqueous solution, (b) in the "open" binary complex with **M**.*Hha*I in which the catalytic loop is in an extended, inactive conformation,[19] (c) in the "closed" binary complex with **M**.*Hha*I in which the catalytic loop is in its closed, active conformation,[20] and (d) in the closed ternary complex with both the enzyme and cofactor present.[20] In addition, flipping of the cytosine and its base-paired guanine was studied using a different simulation methodology in aqueous solution.[9] The goal of the present study

* Corresponding authors. E-mail: nilesh.banavali@cornell.edu; alex@outerbanks.umaryland.edu. Tel: 410-706-7442. Fax: 410-706-5017.

† Current address: Department of Physiology and Biophysics, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021. Tel: 212-746-4237, Fax: 212-746-8690.

‡ Current address: Department of Biopharmaceutical Sciences, School of Pharmacy, UCSF, 600 16th Street, Suite N474E, San Francisco, CA 94143.

is to identify the torsional sampling occurring during base flipping in different environments and to determine the commonalities of these torsional changes. To fulfill this goal, the structures gathered during these MD simulation studies are analyzed to identify patterns of torsional sampling occurring during base flipping and to examine the influence of the environment on these patterns.

Motivation for this work came partly from an early study by Olson and co-workers on the mechanics of distortion of DNA to accommodate binding of carcinogenic compounds.[21] The study explored local torsional changes required to ensure closure of extended DNA sequences with bulges. It was suggested that simultaneous and anticorrelated variations of certain rotations around bonds whose bond vectors were approximately parallel to the helix axis would allow partial base flipping in DNA while preserving the direction of the helical backbone. Using a different nomenclature for backbone torsions than the presently accepted one,[22] that study also indicated that the torsion pairs (in the present nomenclature): γ and 3′-α or 5′-ϵ and ϵ on opposite (5′ vs 3′) sides of the flipping base along the helical axis can mediate such base flipping. Our present MD simulation results support their prediction of compensating distortions on opposite (5′ vs 3′) sides of the flipping base, while indicating that the backbone torsion changes involved are not simple pairwise correlations. Instead, they involve sampling of complex combinations of alternate minima in backbone torsion energy surfaces, mostly centered around the phosphodiester moiety, including some minima not typically populated in isolated DNA crystal structures.

Methods

The MD simulation methods used to obtain the structures analyzed in the present study have been described before[9,16,17] and are only briefly summarized here. The program *CHARMM*[23] was used to perform all calculations with the all-hydrogen protein[24] and nucleic acid[25,26] force field parameters, the TIP3P water model,[27] and published sodium parameters.[28]

A total of six different umbrella sampling MD simulations were analyzed in this study. Five of these simulations studied the flipping of the cytosine in (a) aqueous solution with periodic boundary conditions and particle mesh Ewald (PME)[29] electrostatics, (b) aqueous solution with stochastic boundary conditions, (c) the presence of **M**.*Hha***I** in the open state (binary, open) with stochastic boundary conditions, (d) the presence of **M**.*Hha***I** in the closed state (binary, closed) with stochastic boundary conditions, and (e) the presence of **M**.*Hha***I** with (*S*)-adenosylhomocysteine cofactor in the closed state (ternary, closed) with stochastic boundary conditions. The sixth simulation studied the flipping of the base-paired guanine in aqueous solution with periodic boundary conditions and PME electrostatics.

The **M**.*Hha***I**−(*S*)-adenosylmethionine binary complex crystal structure[19] (PDB ID: 1HMY) was used to model the open binary complex. The closed binary and ternary complex structures were modeled on the basis of the **M**.*Hha***I**−DNA−(*S*)-adenosylhomocysteine crystal structure[20] (PDB ID: 5MHT). A pseudodihedral linkage between the four center-of-mass points created by the four non-hydrogen atom aggregates: (a) flipping cytosine base, (b) sugar attached to the flipping cytosine base, (c) sugar 3′ to the flipping base position on the same strand, and (d) base-pair 3′ to the flipping base position was the one-dimensional reaction coordinate describing base flipping. The 360° range of this center-of-mass pseudodihedral restraint governing base flipping was spanned by generating 72 windows which were each sampled in the four different environments. The intermedi-

ate windows in the protein environment were prepared by least-squares fitting separately generated flipped DNA intermediate structures onto the crystal structure of bound DNA excluding the flipping base, followed by minimization to remove bad contacts. Stochastic boundary simulations with a 35 Å TIP3P water sphere with randomly distributed sodium ions were carried out for all four environments.

Periodic boundary PME constant temperature and pressure (NPT)[30] base flipping umbrella sampling MD simulations were also carried out for the same sequence in a cubic preequilibrated solvent box with dimensions 54 × 54 × 54 Å^3 with enough randomly distributed sodium ions to neutralize the system. Periodic boundary conditions were used with the images generated using the CRYSTAL module in *CHARMM*.[31] To use an integration time step of 0.002 ps, the fast motion of covalent bonds with hydrogen atoms was constrained using SHAKE.[32] Long-range electrostatic interactions were treated with the PME approach[29] with a B-spline order of 4 and a fast Fourier transform grid of one point per Å and a real-space Gaussian-width κ of 0.3 Å^{-1}. Real-space and Lennard-Jones (LJ) interaction cutoffs of 10 Å were used with nonbond interaction lists maintained and heuristically updated out to 12 Å.

These umbrella sampling simulations were performed for both cytosine and its base-paired partner, guanine, flipping separately out of the otherwise stacked, equilibrated B-form double helix. In the case of flipping of the guanine base, the atoms defining the pseudodihedral restraint were varied such that the four non-hydrogen atom aggregates were (a) flipping guanine base, (b) sugar attached to the flipping guanine base, (c) sugar 5′ to the flipping base position on the same strand, and (d) base pair 5′ to the flipping base position.

Analysis of the backbone torsion parameters was carried out using a version of the FREEHELIX program[33] modified to read *CHARMM* DNA trajectories. Time frames for analysis were saved at 1 ps intervals in the umbrella sampling MD simulations. The sampling of torsion pairs such as α−β, γ−δ, or ϵ−ζ obtained through the FREEHELIX analysis was distributed into two-dimensional (2D) density histograms binned at 10° intervals in each torsion. These 2D density distributions were then converted into 2D percentage occupancy histograms by multiplying each bin by 100 such that a cumulative percentage value of 100 was obtained if the values in each bin over the entire 2D torsional space were integrated. The maximum percentage occupancy of each bin in this histogram was cut off at a maximum of 5% to allow visualization of the regions of 2D torsional space showing less than 1% occupancy. The minor groove flipping pathway was defined as the sampling that occurred within the pseudodihedral flipping coordinate umbrella sampling windows between 10° and 180°; all the other windows were classified as the major groove pathway. The main aim of this 2D analysis is to identify the sampling of every torsion with respect to every other torsion, for all torsions intimately involved in base flipping. All figures are made using the OPENDX program, version 4.2.0 (http://www.opendx.org), and the Gnu Image Manipulation Program (GIMP), version 1.2 (http://www.gimp.org).

Results and Discussion

Dihedral Distortion Patterns. According to the standard nomenclature of nucleic acids[22] partially depicted in Figure 1A, the torsional degrees of freedom for each base position are described by the backbone torsions α, β, γ, δ, ϵ, ζ, and χ and the set of sugar torsions collectively represented by the pseudorotational angle (P) and the pseudorotation amplitude.

Dihedral Distortions during Base Flipping

J. Phys. Chem. B, Vol. 110, No. 22, 2006 **10999**
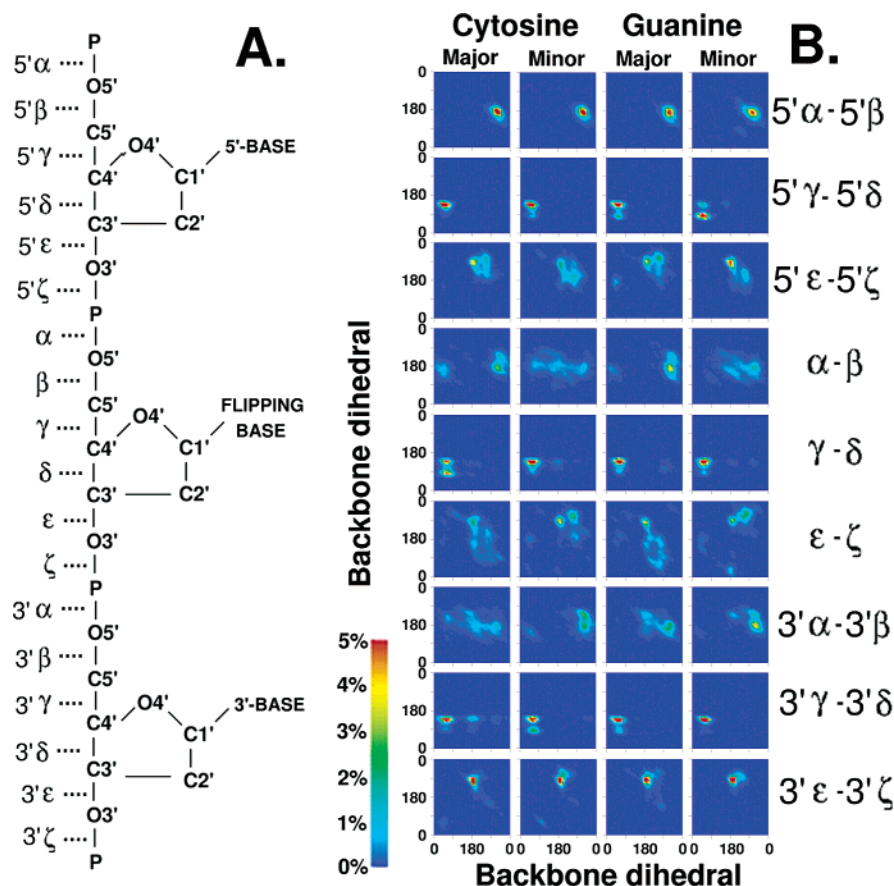
Figure 1. (A) Depiction of nomenclature for backbone torsions, (B) Percentage occupancy of 2D torsional space during restrained umbrella sampling base flipping NPT MD simulations with PME electrostatics treatment. Two columns on left labeled "Cytosine" show sampling during cytosine base flipping; two columns on right labeled "Guanine" show sampling during guanine base flipping. "Major" and "Minor" refer to sampling through major and minor groove pathways, respectively. Labels on right indicate the torsion pairs plotted in that row, for example, $3'\alpha$-$3'\beta$ refers to the row of panels with α torsion on the *X*-axis and β torsion on the *Y*-axis for the base position $3'$ to the flipping base. Color bar shows color gradation of percentage occupancy; all dihedral values are in degrees.

When taken together, these degrees of freedom can almost completely describe the internal geometry of a single DNA base position. Other descriptors, such as helicoidal parameters,[34] are necessary to characterize the geometric complexity of the DNA double helix created by multiple combinations of different internal geometries among individual base positions. These helicoidal parameters are not considered here, because the present study is focused on understanding the detailed mechanics behind the localized nature of the base flipping distortion in the context of the rotatable bonds in the DNA backbone.[9] The only parameters analyzed are, therefore, the ones providing the best description for the restriction of the flipping distortion to the continuous DNA backbone stretch between the two base positions on either side of the flipping cytosine or guanine base shown in Figure 1A. Such a localization was also indicated in a survey of experimental flipped DNA structures bound to **M**.*Hha***I**.[35]

Single base flipping can occur through either the major groove or the minor groove of the DNA double helix.[9] The behavior of the backbone dihedrals analyzed separately for these two pathways is shown in Figure 1B. The $5'\alpha$, $5'\beta$, $5'\gamma$, $3'\epsilon$, and $3'\zeta$ torsions, as visualized along the appropriate axis of the 2D histograms, sample primarily B-form conformations,[25] supporting the observation that the two bordering sugar moieties limit the range of the distortion of the DNA backbone during base flipping.[9]

The majority of the remaining rotable bonds between the sugar moieties surrounding the flipping base sample a wide range of conformations. The exceptions are the γ–δ and $3'\gamma$-$3'\delta$ pairs, that show only marginal flexibility. The changes that do occur in the δ torsion, with values shifting from $130°$ to $80°$, are due to changes in sugar pucker from C2′-endo to C3′-endo forms.[36] Sampling of C3′-endo sugar pucker occurs in both groove pathways for both flipping bases; however, the most substantial sampling for the C3′-endo sugar pucker is seen during guanine flipping through the minor groove for the $5'$-sugar moiety and cytosine flipping through the major groove for the sugar moiety attached to the flipping cytosine itself.

The dihedrals that show the greatest distortion during flipping are those closely associated with the two phosphodiester groups surrounding the flipping base. This class includes the $5'\epsilon$, $5'\zeta$, α, β, ϵ, ζ, $3'\alpha$, and $3'\beta$ torsions. Each of these torsions show different sampling behavior for the major versus minor groove pathways. However, the specific patterns observed during cytosine flipping also occur during guanine flipping in both minor and major groove pathways. Comparing individual panels from the first and third columns and the second and fourth columns for the $5'\epsilon$-$5'\zeta$, α–β, ϵ–ζ, and $3'\alpha$-$3'\beta$ torsion pairs in Figure 1B illustrates these similarities. This pattern occurs despite the differences in both sequence context and atom selections used in the pseudodihedral restraint for cytosine versus guanine flipping, suggesting that the sampling patterns in backbone torsions during single base flipping are intrinsically conserved.

There is another interesting pattern of similarities between sampling for torsions on one side ($5'$ or $3'$) of the base flipping

11000 *J. Phys. Chem. B, Vol. 110, No. 22, 2006*

Banavali et al.

through the major groove pathway and the corresponding torsions on the opposite side of the base flipping through the minor groove pathway. This can be visualized by comparing the following panels for both cytosine and guanine base flipping in Figure 1B: the major groove $5'\epsilon$-$5'\zeta$ with the minor groove ϵ-ζ, the major groove ϵ-ζ with the minor groove $5'\epsilon$-$5'\zeta$, the major groove α-β with the minor groove $3'\alpha$-$3'\beta$, and the major groove $3'\alpha$-$3'\beta$ with the minor groove α-β. For these pairs, it is evident that rotameric states sampled on the 5′ side of the flipping base in the major groove pathway are sampled on the 3′ side of the flipping base in the minor groove pathway, and vice versa. Separate analysis of the minor groove and major groove pathways can thus identify the rotameric patterns sampled by the phosphodiester backbone on opposite (5′ vs 3′) sides of a flipping base while moving through the two helical grooves. Altered sampling of a set of torsions on one side (5′ or 3′) of the flipping base is balanced by compensating alterations in sampling of a similar set of torsions on the opposite side. Moreover, the alterations are specific with respect to the side (i.e., 5′ or 3′) on which the torsions are located and the groove through which the base flips. For example, a greater range of the $3'$-α torsion is sampled during base flipping through the major groove, while a greater range of the α torsion is sampled during base flipping through the minor groove.

Torsions on Opposite (5′ vs 3′) Sides of Flipping Base. The previous section indicated that certain torsions show different behavior depending on the helical groove through which base flipping occurs. This pattern of behavior repeats itself on the opposite (5′ vs 3′) sides of the flipping base, during flipping through opposite DNA grooves. To further probe this pattern, 2D percentage occupancy histograms of identical torsion pairs, separated by one nucleotide unit (e.g., α and $3'\alpha$), are shown in Figure 2. In addition, the cytosine and guanine base flipping results from the PME periodic boundary simulations are also compared to cytosine base flipping carried out under stochastic boundary (SB) conditions in aqueous solution.

A systematic comparison of the SB and the PME periodic boundary MD simulations for cytosine base flipping is beyond the scope of this study. In terms of relative computational cost, the SB approach allows the simulation to be run about 3 times faster than the PME periodic boundary approach for the isolated DNA system. The SB approach[37] is necessarily more approximate, since long-range electrostatic effects are ignored through atom-based truncation, and Langevin dynamics determines the behavior of the solvent at the boundary of the enclosed spherical system. The PME approach with periodic boundary conditions takes into account long-range electrostatic effects and allows diffusion of solvent molecules in and out of the primary cell, but imposes an artificial periodicity in its behavior.[29] The choice of SB conditions in a previous study[16] was a tradeoff between accuracy and computational cost that was considered adequate because of the similarities between the base flipping free energy profiles for isolated DNA in aqueous solution obtained using both approaches. In the present study, the additional criteria considered for comparison between the SB and PME periodic boundary approaches are the detailed conformational behavior of the local torsions during base flipping.

Analysis of Figure 2 again shows the γ and δ torsions to be relatively rigid, mostly occupying either one or two specific conformations. The α, β, ϵ, and ζ torsions, on the other hand, occupy many different conformational states, consistent with the more flexible nature of the phosphodiester moiety,[38] with the major and minor groove pathways showing distinctly
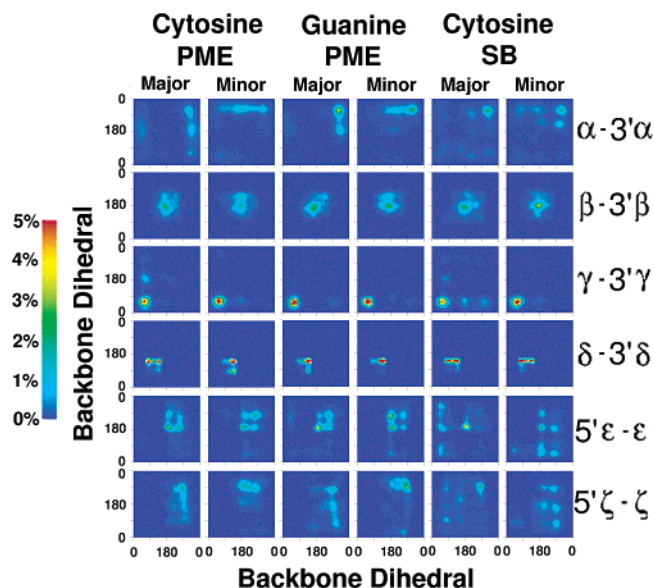
Figure 2. Percentage occupancy of 2D torsional space during restrained umbrella sampling MD simulations of cytosine and guanine base flipping shown as follows: Two columns at left labeled "Cytosine" and two columns in center labeled "Guanine" show sampling for cytosine and guanine base flipping, respectively, both carried out with PME electrostatics and periodic boundary conditions; two columns on right labeled "SB" (for stochastic boundary) show sampling for cytosine base flipping carried out with SB conditions. "Major" and "Minor" labels refer to sampling through the major and minor groove pathways, respectively. Labels on right indicate torsions plotted in each row, for example, α-$3'\alpha$ refers to the row of panels with α torsion for the flipping base position on the X-axis and α torsion for the 3′ base position on the Y-axis. Color bar shows color gradation of percentage occupancy. All dihedral values are in degrees.

different behavior for each torsion pair (See Supporting Information for a more detailed analysis of pairwise sampling among these phosphodiester torsions on opposite (5′ vs 3′) sides of the flipping base).

The most noticeable patterns are the greater range of sampling for (a) the $3'\alpha$ torsion in the major groove pathway and the α torsion in the minor groove pathway, and (b) the ζ torsion in the major groove pathway and the $5'\zeta$ torsion in the minor groove pathway, for both cytosine and guanine base flipping in the periodic boundary system MD simulations. The torsional sampling for the more approximate SB simulations of cytosine base flipping is spread out over a larger range for the flexible torsions. Nevertheless, a similar pattern of a wider range of sampling of the $3'\alpha$ torsion during major groove flipping and a wider range of sampling of the α torsion during minor groove flipping can be discerned. As stated in the Introduction, Olson and co-workers anticipated anticorrelated behavior in specific backbone torsion pairs on opposite (5′ vs 3′) sides of DNA bulges.[21] From Figure 2, the essential insight is that flipping does not seem to occur due to correlated (or anticorrelated) transitions in specific torsion pairs on opposite (5′ vs 3′) sides of the flipping base, but rather due to adjustments in specific combinations of phosphodiester torsions on opposite (5′ vs 3′) sides of the flipping base in patterns that compensate for energetic costs associated with the individual dihedral distortions. The distortions seen here are consistent with the distortions observed in the survey of flipped base structures (involving 5′-ϵ, γ, ζ, and 3′-β torsions) with the same cognate sequence bound to **M**.*Hha***I**.[35] These patterns of distortions appear to allow the DNA duplex to maintain an approximately canonical conformation beyond the portion of the strand involved in the localized flipping transition itself.
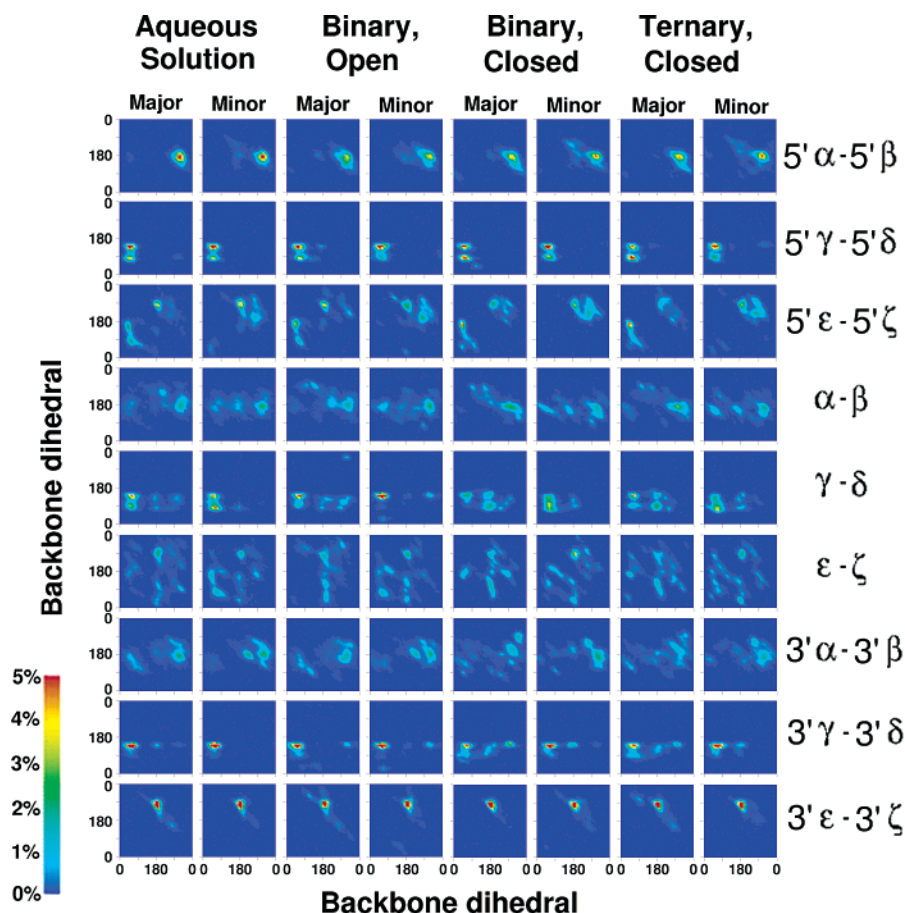
Dihedral Distortions during Base Flipping

J. Phys. Chem. B, Vol. 110, No. 22, 2006 **11001**

Figure 3. Percentage occupancy of 2D torsional space during restrained umbrella sampling SB MD simulations of cytosine base flipping in four environments shown as follows: Two columns on left labeled "Aqueous Solution" show sampling in aqueous solution; two columns in center left labeled "Binary, Open" show sampling in the binary, open loop **M**.*Hha***I**−DNA complex; two columns in center right labeled "Binary, Closed" show sampling in the binary, closed loop **M**.*Hha***I**−DNA complex; two columns on right labeled "Ternary, Closed" show sampling in the ternary, closed loop **M**.*Hha***I**−DNA−(*S*)-adenosylhomocysteine complex. "Major" and "Minor" labels refer to sampling through major and minor groove pathways, respectively. Labels on right indicate torsion pairs plotted in that row, for example, $5'\alpha$-$5'\beta$ refers to the row of panels with α torsion on the *X*-axis and β torsion on the *Y*-axis for the base position $5'$ to the flipping base. Color bar shows color gradation of percentage occupancy; all dihedral values are in degrees.

Effect of Environment. DNA in aqueous solution is in a homogeneous environment that imposes minimal asymmetry on its dynamic behavior. When bound to a DNA-binding protein that specifically facilitates base flipping, the stabilization of intermediate flipped states due to specific DNA−protein interactions could, in principle, change the conformational pathway utilized for base flipping entirely as compared to aqueous solution. In Figure 3, percentage-occupancy 2D torsional space histograms are shown for SB base flipping MD simulations with (a) DNA alone in aqueous solution (labeled "Aqueous Solution"), (b) DNA bound to **M**.*Hha***I** with the catalytic loop in the open conformation (labeled "Binary, Open"), (c) DNA bound to **M**.*Hha***I** with the catalytic loop closed around the DNA (labeled "Binary, Closed"), and (d) DNA bound to **M**.*Hha***I** with bound (*S*)-adenosylhomocysteine cofactor and the catalytic loop also closed around the DNA (labeled "Ternary, Closed").

As mentioned before, there is a tendency for the SB MD simulation results to show sampling of a greater conformational range than those obtained using PME periodic boundary MD simulations. The clearest example is the much wider range of values sampled in the $\epsilon-\zeta$ torsion pair seen by comparing the PME periodic boundary MD simulation results in Figure 1B with the SB results in Figure 3. This increased sampling in the SB simulations may be associated with the different strategy for obtaining starting conformations for the individual umbrella

sampling windows, the use of atom truncation versus PME for the treatment of long-range electrostatic interactions, and/or edge effects associated with the SB approach.[16] Some of the trends, however, do remain consistent with those observed in the PME simulations, e.g., the $\alpha-\beta$ histograms in Figures 1 and 3 show a tendency for a smaller range of torsional sampling during major groove flipping as compared to minor groove flipping.

Comparison of the torsional space sampling during flipping in the presence of **M**.*Hha***I** shows both similarities and differences as compared to flipping in aqueous solution. The sampling of the $5'\alpha$-$5'\beta$ torsion pair is marginally increased in the binary and ternary complexes, presumably due to interactions with the protein. The $5'\gamma$-$5'\delta$ and $3'\gamma$-$3'\delta$ torsion pairs maintain their relative rigidity for the most part, while the $\gamma-\delta$ torsion pair for the flipping base has increased sampling in the closed binary and ternary complexes, occupying alternate minimum-energy conformations not seen in the aqueous solution MD simulations. The other torsional degrees of freedom between the two bordering sugar moieties sample a large conformational range but maintain their distinct behavior in the major and minor groove pathways in all four environments to a significant extent. The major groove $\alpha-\beta$ and the minor groove $3'\alpha$-$3'\beta$ torsional distributions provide good examples of this trend. Even with drastic changes in the environment of the DNA, the patterns of the local dihedral distortions associated with base flipping through specific grooves remain similar to those in aqueous

TABLE 1: 2D Backbone Torsional States around the Underlined Flipping Cytosine in the Sequence GTCAGCGCATGG or the Underlined Flipping Guanine in the Paired Opposite Strand Sequence CCATGCGCTGAC that Have Substantial Occupancy in the MD Simulations with Periodic Boundary Systems and PME Electrostatics

flipping base	position	groove	$(\alpha, \beta)^a$	$(\gamma, \delta)^a$	$(\epsilon, \zeta)^a$
cytosine	5′	major	$300,170_{A,B}$	$50,140_B$	$190,270_{BI}$ $250,280_{g^-,g^-}$
		minor	$300,170_{A,B}$	$50,140_B$	$190,250_{BI}$ $250,200_{BII}$
	flipping	major	$300,170_{A,B}$ $320,230_{g^-,t}$	$60,140_B$ $60, 80_A$	$190,260_{BI}$
		minor	$300,170_{A,B}$	$60,140_B$	$190,270_{BI}$ $250,290_{g^-,g^-}$
	3′	major	$300,180_{A,B}$	$60,140_B$ $180,140_{t,t}$	$190,260_{BI}$
		minor	$300,230_{g^-,t}$ $300,180_{A,B}$	$60,140_B$ $60, 90_A$	$190,260_{BI}$
guanine	5′	major	$300,170_{A,B}$	$60,140_B$	$190,270_{BI}$ $250,290_{g^-,g^-}$
		minor	$290,170_{A,B}$	$60, 90_A$ $50,140_B$	$180,270_{BI}$ $250,220_{g^-,t}$
	flipping	major	$300,170_{A,B}$	$60,140_B$ $60, 90_A$	$190,260_{BI}$ $250, 60_{g^-,g^+}$
		minor	$300,180_{A,B}$ $190,150_{t,t}$	$60,140_B$	$190,270_{BI}$ $250,300_{g^-,g^-}$
	3′	major	$300,170_{A,B}$	$60,140_B$ $60, 80_A$	$190,260_{BI}$
		minor	$300,180_{A,B}$ $300,230_{g^-,t}$	$60,140_B$	$190,260_{BI}$

^{a} All values are in deg, estimated in 10° intervals. Note that individual torsion pair states on the same row but in different columns are not implied to occur simultaneously. Subscripts A and B refer to conformations occurring in the A or B forms of DNA, respectively; subscripts BI and BII refer to BI and BII conformations of the $\epsilon-\zeta$ torsion pair, respectively. Subscript labels g^+, t, and g^- are assigned to values in the ranges 0° through 120°, 120° through 240°, and 240° through 360°, respectively.

solution for the SB MD simulations of cytosine base flipping. Although this could be attributed to the initial configuration of the DNA in all SB simulations being the same, the change in environment clearly does not result in large-scale conformational pathway readjustments.

Alternate Metastable Rotameric States. The metastable states that are involved in the observed patterns of backbone dihedral distortions are indicated in Tables 1 and 2. These states are identified by locating bins in the 2D torsional percentage occupancy histograms shown in Figure 1B and Figure 3 that show the greatest occupancy among any island cluster of bins with an occupancy greater than 1%. This criteria eliminates regions of space that are sampled less than 1% of the time in each 10 × 10° region of 2D torsional space. This selection process is expected to filter out transition-state conformations, leaving only metastable intermediate states.

Analysis of Tables 1 and 2 shows that the usual local minima seen in A-form and B-form DNA[25] are present for all torsion pairs in both pathways and in all environments. However, a variety of additional local minima not seen in canonical structures are also sampled. Specific examples are t, t values for the $\alpha-\beta$ and $\gamma-\delta$ torsion pairs, and g^-, g^- for the $\epsilon-\zeta$ torsion pair. Specific alternate local conformations of torsions during base flipping identified here are consistent with alternate local energy minima in potential energy surfaces calculated using quantum mechanical (QM) calculations of the same torsions in representative model compounds.[39,40] The present analysis along with the QM data indicate that these alternate conformations may be useful in mediating important distortions in DNA, even though they may not be substantially represented in isolated DNA crystal structure distributions.[39]

Conclusions

Structural analysis of umbrella sampling MD simulations of base flipping indicate that the localized distortion of DNA during base flipping involves changes in multiple torsions in the vicinity of the flipping base. The perturbed torsions lie on both the 5′ and 3′ sides of the flipping base and are mostly associated with the phosphodiester moiety. The ability to manipulate the conformational behavior of the phosphodiester moiety through specific charged amino acid side chain interactions may be critical to achieving these distortions in the presence of nucleic acid-binding proteins. Specific torsion combinations define the intermediate configurations that are sampled during flipping through minor or major groove pathways. The work of Olson and co-workers[21] predicted these torsion combinations to be composed of one specific torsion on the 5′-side of a DNA bulge and another specific torsion on the 3′-side of a DNA bulge. The present results are consistent with that general concept; however, the torsion combinations involved in base flipping do not show the anticipated highly correlated pairwise changes. The torsional changes involved instead are complex combinations of canonical and alternate minima in specific individual backbone torsion degrees of freedom on opposite (5′ vs 3′) sides of the flipping base.

The pseudodihedral restraint used to induce base flipping, the simulation methodology, sequence effects, and environmental effects, all show discernible but limited effects on the dihedral distortion details. The effect of DNA sequence or flipping base identity on the pattern of dihedral distortions is not systematically addressed in this work. The flipping of a guanine surrounded by a different sequence context in aqueous solution does, however, show similar dihedral distortions as those occurring during cytosine flipping in aqueous solution. This suggests that sequence may influence the energetics of the flipping process through intramolecular and intermolecular nonbonded interactions[9] but may not substantially affect the structural backbone dihedral distortion pathways used in base flipping. The predicted local backbone transitions, though identified for base flipping only, may also be applicable to other

TABLE 2: 2D Backbone Torsional States around the Underlined Flipping Cytosine in the Sequence GTCAGCGCATGG Showing Substantial Occupancy in SB MD Simulations in Different Environments[a]

position	groove	aqueous solution (α, β)	(γ, δ)	(ϵ, ζ)	binary, open (α, β)	(γ, δ)	(ϵ, ζ)
5′	major	$300,180_{A,B}$	$50,140_B$ $50, 80_A$	$190,260_{BI}$ $40,160_{g^+,t}$ $60, 80_{g^+,g^+}$	$300,170_{A,B}$	$60,140_B$ $60, 80_A$	$190,260_{BI}$ $40,170_{g^+,t}$ $60, 80_{g^+,g^+}$
	minor	$300,170_{A,B}$	$50,140_B$ $50, 80_A$	$190,260_{BI}$ $260,200_{BII}$	$300,170_{A,B}$	$60,140_B$ $60, 90_A$	$180,260_{BI}$ $260,200_{BII}$
flipping	major	$300,170_{A,B}$ $220,240_{t,t}$	$60,140_B$ $60, 90_A$	$190,270_{t,g^-}$	$300,180_{A,B}$ $220,180_{t,t}$ $140,270_{t,g^-}$	$60,140_B$ $60, 90_A$ $270,110_{g^-,t}$	$190,260_{BI}$ $190, 50_{t,g^+}$ $190,110_{t,g^+}$
	minor	$300,170_{A,B}$	$60, 80_A$ $60,140_B$	$190,270_{BI}$ $40,160_{g^+,t}$ $70, 80_{g^+,g^+}$	$300,170_{A,B}$	$50, 80_A$ $60,140_B$	$190,260_{BI}$ $60, 80_{g^+,g^+}$ $100,190_{g^+,t}$ $180,130_{t,t}$ $180, 50_{t,g^+}$
3′	major	$300,170_{A,B}$	$60,140_B$	$190,270_{BI}$	$300,170_{A,B}$	$60,140_B$	$190,260_{BI}$
	minor	$300,180_{A,B}$ $230,180_{t,t}$	$60,140_B$	$190,260_{BI}$	$300,190_{A,B}$ $220,190_{t,t}$	$60,140_B$ $300,140_{g^-,t}$	$190,260_{BI}$

position	groove	binary, closed (α, β)	(γ, δ)	(ϵ, ζ)	ternary, closed (α, β)	(γ, δ)	(ϵ, ζ)
5′	major	$290,170_{A,B}$	$60, 80_A$ $50,140_B$	$40,160_{g^+,t}$ $200,270_{BI}$	$290,170_{A,B}$	$50, 90_A$ $60,140_B$	$40,170_{g^+,t}$ $200,270_{BI}$
	minor	$290,180_{A,B}$	$60,140_B$ $60, 90_A$	$190,260_{BI}$ $260,220_{g^-,t}$	$290,180_{A,B}$	$60,140_B$ $60, 90_A$	$190,260_{BII}$ $260,220_{g^-,t}$
flipping	major	$270,170_{A,B}$	$60,140_B$ $180, 90_{t,g^+}$ $180,150_{t,t}$	$180,260_{BI}$ $180,110_{t,g^+}$ $260,170_{BII}$ $50,170_{g^-,t}$	$280,170_{A,B}$	$60,140_B$ $180, 90_{t,g^+}$ $180,140_{t,t}$ $60, 90_A$	$180,120_{t,g^+}$ $40,170_{g^+,t}$
	minor	$300,160_{A,B}$ $40,160_{g^+,t}$	$60, 90_A$ $60,130_B$	$190,270_{BI}$ $180, 40_{t,g^+}$ $80,190_{g^+,t}$	$300,160_{A,B}$ $50,160_{g^+,t}$ $160,160_{t,t}$	$70, 70_A$ $50,130_B$	$190,270_{BI}$ $180, 40_{t,g^+}$ $180,270_{BI}$
3′	major	$320,270_{g^-,g^-}$ $300,200_{A,B}$	$60,140_B$ $280,150_{g^-,t}$ $160,100_{t,g^+}$ $60, 90_A$	$190,270_{BI}$	$280,210_{g^-,t}$ $180,190_{t,t}$	$60,140_B$ $280,140_{g^-,t}$ $60,100_A$	$190,270_{BI}$
	minor	$300,170_{A,B}$ $260,240_{g^-,t}$	$60,140_B$ $180,140$	$190,270_{BI}$	$300,170_{A,B}$ $250,250_{g^-,g^-}$	$60,140_B$	$190,270_{BI}$

[a] All values are in deg, estimated in 10° intervals. Note that individual torsion pair states on the same row but in different columns are not implied to occur simultaneously. Subscripts A and B refer to conformations occurring in the A- or B-forms of DNA, respectively; subscripts BI and BII refer to BI and BII conformations of the ϵ-ζ torsion pair, respectively. Subscript labels g^+, t, and g^- are assigned to values in the ranges 0° through 120°, 120° through 240°, and 240° through 360°, respectively.

distortions of DNA. If base flipping events occur in close conjunction repeatedly over neighboring base pairs, they can ultimately lead to DNA unwinding.[41,42] The present observations are therefore potentially useful in probing large-scale DNA distortions involved in the crucial biological functions of replication and transcription.

The identification of conserved distortions in DNA can advance the understanding of the biological effects of introduction of chemical modifications in the nucleic acid backbone. These modifications, utilized in the design of antisense oligonucleotides, can affect backbone torsions involved in base flipping without changing their sequence-dependent behavior.[35,43] The ability or inability of backbone-modified DNA analogues to attain intermediate rotameric states required in DNA distortion pathways can be tested with *ab initio* calculations on model compounds and MD simulation studies. Knowledge of common metastable substates sampled during DNA distortions can thus be linked to the impact of chemical modifications on the dynamic behavior of modified nucleic acid analogues. The connection between specific chemical modifications and the resulting change in dynamic behavior can be further exploited to modulate interaction of the modified analogues with nucleic acid binding proteins.

Supporting Information Available: Six tables showing linear correlation coefficients for all pairs of torsions analyzed, one table showing linear correlation coefficients between δ and pseudorotation angle that illustrates what the value of the correlation coefficient should be for two torsions that are truly correlated. Five figures showing the configurations visited in multiple torsion pairs on opposite (5′ vs 3′) sides of the flipping base in different MD simulation systems and environments, a text description of the observations in these five figures, and one figure illustrating the five different MD simulation systems used in the calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.

(2) Mol, C. D.; Hosfield, D. J.; Tainer, J. A. *Mutat. Res.* **2000**, *460*, 211–229.

(3) Roberts, R. J. *Cell* **1995**, *82*, 9–12.

(4) Roberts, R. J.; Cheng, X. D. *Annu. Rev. Biochem.* **1998**, *67*, 181–198.

(5) Stivers, J. T. *Prog. Nucleic Acid Res. Mol. Biol.* **2004**, *77*, 37–65.

(6) van Aalten, D. M. F.; Erlanson, D. A.; Verdine, G. L.; Joshua-Tor, L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11809–11814.

(7) Verdine, G. L.; Norman, D. P. G. *Annu. Rev. Biochem.* **2003**, *72*, 337–366.

(8) Horton, J. R.; Ratner, G.; Banavali, N. K.; Huang, N.; Choi, Y.; Maier, M. A.; Marquez, V. E.; MacKerell, A. D., Jr.; Cheng, X. D. *Nucleic Acids Res.* **2004**, *32*, 3877–3886.

(9) Banavali, N. K.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2002**, *319*, 141–160.

(10) Varnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272–7273.

(11) Priyakumar, U. D.; MacKerell, A. D., Jr. *Chem. Rev.* **2006**, *106*, 489–505.

(12) Gueron, M.; Kochoyan, M.; Leroy, J. L. *Nature (London)* **1987**, *328*, 89–92.

(13) Varnai, P.; Canalia, M.; Leroy, J. L. *J. Am. Chem. Soc.* **2004**, *126*, 14659–14667.

(14) Priyakumar, U. D.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2006**, *2*, 187–200.

(15) Priyakumar, U. D.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2006**, *128*, 678–679.

(16) Huang, N.; Banavali, N. K.; MacKerell, A. D., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 68–73.

(17) Huang, N.; MacKerell, A. D., Jr. *J. Mol. Biol.* **2005**, *345*, 265–274.

(18) Huang, N.; MacKerell, A. D., Jr. *Philos. Trans. R. Soc. London, Ser. A* **2004**, *362*, 1429–1460.

(19) Cheng, X. D.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J. *Cell* **1993**, *74*, 299–307.

(20) O'Gara, M.; Roberts, R. J.; Cheng, X. *J. Mol. Biol.* **1996**, *263*, 597–606.

(21) Olson, W. K., Marky, N. L., Srinivasan, A. R., Do, K. D., Cicariello, J. *Molecular Basis of Cancer: Part A*; Alan R. Liss, Inc: New York, 1985.

(22) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(24) MacKerell, A. D., Jr.; Bashford, D.; Bellott, R. L.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(25) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.

(26) MacKerell, A. D., Jr.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.

(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(28) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

(29) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(30) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(31) Field, M. J.; Karplus, M. *CRYSTAL: program for crystal calculations in CHARMM*; Harvard University: Cambridge, MA, 1992.

(32) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(33) Dickerson, R. E. *Nucleic Acids Res.* **1998**, *26*, 1906–1926.

(34) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.

(35) Wang, P. Y.; Brank, A. S.; Banavali, N. K.; Nicklaus, M. C.; Marquez, V. E.; Christman, J. K.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2000**, *122*, 12422–12434.

(36) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819–840.

(37) Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159–181.

(38) Banavali, N. K.; MacKerell, A. D., Jr. *J. Am. Chem. Soc.* **2001**, *123*, 6747–6755.

(39) Foloppe, N.; MacKerell, A. D., Jr. *J. Phys. Chem.* **2000**, *102*, 6669–6678.

(40) Foloppe, N.; MacKerell, A. D., Jr. *Biophys. J.* **1999**, *76*, 3206–3218.

(41) Patel, P. H.; Suzuki, M.; Adman, E.; Shinkai, A.; Loeb, L. A. *J. Mol. Biol.* **2001**, *308*, 823–837.

(42) Schneider, T. D. *Nucleic Acids Res.* **2001**, *29*, 4881–4891.

(43) Banavali, N. K.; MacKerell, A. D., Jr. *Nucleic Acids Res.* **2001**, *29*, 3219–3230.