JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

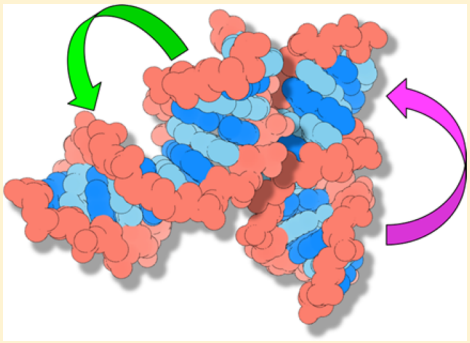
Anisotropy of B-DNA Groove Bending

Ning Ma and Arjan van der Vaart*

Department of Chemistry, University of South Florida, 4202 East Fowler Avenue, CHE 205, Tampa, Florida 33620, United States

Supporting Information

ABSTRACT: DNA bending is critical for DNA packaging, recognition, and repair, and occurs toward either the major or the minor groove. The anisotropy of B-DNA groove bending was quantified for eight DNA sequences by free energy simulations employing a novel reaction coordinate. The simulations show that bending toward the major groove is preferred for non-A-tracts while the A-tract has a high tendency of bending toward the minor groove. Persistence lengths were generally larger for bending toward the minor groove, which is thought to originate from differences in groove hydration. While this difference in stiffness is one of the factors determining the overall preference of bending direction, the dominant contribution is shown to be a free energy offset between major and minor groove bending. The data suggests that, for the A-tract, this offset is largely determined by inherent structural properties, while differences in groove hydration play a large role for non-A-tracts. By quantifying the energetics of DNA groove bending and rationalizing the origins of the anisotropy, the calculations provide important new insights into a key biological process.



INTRODUCTION

While DNA is a stiff molecule, its bending is critical to its biological function.[1] To fit inside the cell, DNA is supercoiled and highly bent by packaging proteins in prokaryotes, and histones in eukaryotes.[2] Bending is also essential for DNA looping, an important regulation mechanism for gene expression that brings together sites that are distant in sequence.[3−5] DNA bending plays a key role in the thermodynamics of protein−DNA binding, and has been shown to strongly modulate binding affinities.[6−8] It is also important for DNA repair, where DNA bending reduces the energetic barrier for base flipping.[9] In recent years, DNA bending has received renewed interest, spurred by the observation that the cyclization of short, 94 base pair DNA strands appeared much more facile than predicted by the worm-like chain model.[10] Since the worm-like chain model[11] well-describes DNA bending at the long length scale (that is, for strands longer than the DNA persistence length of ∼500 Å or ∼150 base pairs),[1,12] these cyclization results were surprising, and subsequently followed by a large number of experimental and theoretical studies. While some of these confirmed the increased flexibility of DNA at the short length scale,[13−21] others saw no anomalous behavior,[22−25] and the debate on the short length scale behavior continues.[7,25]

Here we consider the anisotropy of B-DNA groove bending. B-DNA is the most common and physiologically relevant form of DNA, consisting of a right-handed helix with a pitch of 10 base pairs and two types of grooves: the major groove with a width of ∼22 Å in straight B-DNA, and the minor groove with a width of ∼12 Å (Figure 1a).[26] Bending can proceed toward either groove, but since the grooves are not equivalent, the energetics and resultant shape will depend on the direction of bending. While a few theoretical models incorporated the anisotropy of groove bending,[27−29] with some explicitly favoring bending toward the major groove over bending toward the minor groove,[30−32] or vice versa,[33] direct experimental evidence for the anisotropy is scant. Since the direction of bending is hard to control in AFM, magnetic bead, or other pulling experiments, most experimental information on the anisotropy of DNA bending comes from inferences based on statistical analyses of DNA and protein−DNA structures. An early statistical analysis of 11 structures showed preference for minor groove bending for AG, AT, CG, and CT steps, and major groove bending for AA, GA, GG, TC, TT, and CC steps.[34] A subsequent analysis of 66 B-DNA structures showed a preference for major groove bending for CG, AA and GG dimer steps, a preference for minor groove bending for the GC step, and lesser directional preferences for other steps.[35] An analysis of 86 protein−DNA complexes showed that DNA bends were mostly due to roll angles[36] (the rotation of a base pair along the long axis,[37] Figure 1b), with most rolling at pyrimidine-purine steps and toward the major groove, suggesting favorability of bending toward the major groove.[36] Average rolling toward the major groove was found for all dimer steps in large statistical analysis of protein−DNA structures,[38] while another analysis found a tendency for major groove bending for purine-pyrimidine steps and minor groove bending for pyrimidine-purine steps in B-DNA duplexes.[39] Special attention has been given to A-tracts, DNA sequences with four to six consecutive adenines.[40] Intrinsic curvature of A-tracts was first deduced from anomalously slow

Received: May 18, 2016
Published: July 20, 2016

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

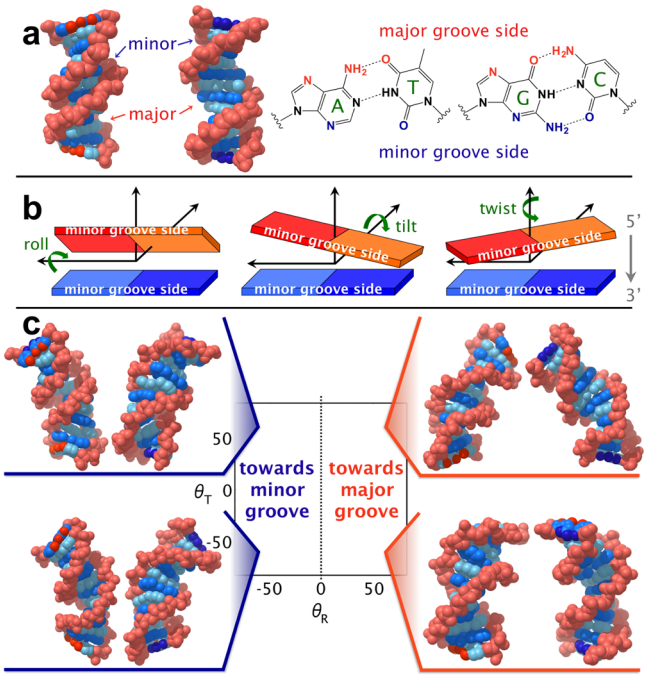


Figure 1. DNA grooves and DNA bending. (a) Major and minor grooves of DNA, and relative orientation of bases with respect to the grooves. (b) Rotational step parameters. (c) DNA configurations at $(\Theta_R, \Theta_T) = (\pm 60°, \pm 60°)$. In (a) and (c), two side views are shown for each conformation. Polar atoms at the major groove side of the first and last base are shown in orange, polar atoms at the minor groove side in dark blue.

electrophoretic mobilities.[41,42] A-tracts have a narrowed minor groove,[40] and gel mobility[43] and solution state studies[44−46] showed a tendency for bending toward the minor groove.

These structural analyses have been complemented by modeling and simulation studies. Early empirical energy models showed that DNA bends toward the grooves rather than the backbone, with a preference for minor groove bending for purine-pyrimidine steps, and preferred major groove bending for pyrimidine-purine steps.[47] This tendency was also found in Monte Carlo simulations of DNA hexamers in implicit solvent.[48] Energy minimizations in implicit water using superhelical symmetry restraints showed a preference for minor groove bending for A_4T_4CG, while T_4A_4CG showed a weak preference for minor bending direction;[49] positive rolls were observed at TA and CG steps and negative roll at AT steps. Energy minimizations in implicit solvent using a screw axis restraint showed a preference for bending toward the minor groove.[50] This preference was strong for the A-tract and the recognition sequence of human and bovine papillomaviruses E2 proteins, but for an alternating AT sequence the tendencies for bending toward the minor and major groove were nearly equal. A molecular dynamics (MD) simulation of the $CGCGA_6CG$ and $CGCA_6GCG$ A-tract sequences showed preferred bending toward the minor groove;[51] the structure, curvature, and solvation of A-tracts has been subject of a number of MD studies.[52−62] A systematic MD study of all tetranucleotide steps of B-DNA in explicit water showed a preference for positive roll angles for pyrimidine-purine and GG steps, and negative roll for GC, GT, AT, and AA steps.[63−66] Significant conformational shifts due to next-nearest neighbor effects were observed, however.[65,66]

While these studies have provided important but somewhat conflicting information, the minimization studies were limited by the exclusion of temperature effects and the implicit representation of the solvent, and the MD studies were limited by sampling. Sampling can be significantly enhanced by the use of biasing techniques,[67,68] which allow for the crossing of energy barriers and sampling of high energy states. Such free energy simulations have been used to study DNA bending without regard to the direction of bending, in both bare[19,20,69,70] and protein-bound DNA,[71−73] while directional bending was studied for the A-tract using a screw axis restraint.[50] The latter study confirmed the preference for minor groove bending, in agreement with the energy minimization studies using the same restraint,[50] but free energy curves were only reported as a function of the overall bending angle and no other sequences were studied.

To gain more insights into the anisotropy of DNA bending, we performed free energy simulation studies of eight DNA dodecamers in explicit water and complemented this information by a statistical analysis of protein−DNA structures. The simulations used a new biasing restraint based on the Madbend definition of the bending angle,[51] which allowed for biasing in a specific direction. The anisotropy in bending was quantified by free energy curves and persistence length analyses, and the origins of the anisotropy were investigated. Marked differences between the A-tract and other sequences were revealed, with a high tendency for bending toward the minor groove for A-tracts and major groove bending for non-A-tract sequences. Our simulations highlight the importance of solvation for preferences in the direction of DNA bending.

METHODS

PDB Analysis. Protein Data Bank structures of all protein−DNA complexes were downloaded, and structures with nicks, gaps, single strands, flipped bases, and modified or damaged nucleotides were removed. To avoid the occurrence of multiple centers of bending, all structures were visually inspected, and only structures in which the protein either bound to the major or to the minor groove were retained, leaving a total of 628 structures (Tables S1 and S2 of the Supporting Information). Reference planes for Madbend[51] analysis were chosen in the center of bending based on visual inspections of the structures.

Simulations. While a number of methods to calculate DNA bending angles have been proposed,[36,50,51,62,74−79] we chose the Madbend procedure[51] for its robustness[80] and ease in distinguishing the direction of bending. According to this procedure DNA bending angles are calculated from roll (ρ), tilt (τ), and twist (Ω) angles. These angles describe the relative orientation of base pairs in a DNA dimer step (i.e., two adjacent base pairs), and are depicted in Figure 1b.[37] Nonzero roll and tilt bend DNA,[36] but as dimer steps are naturally twisted, and the overall bending angle is affected by twist. The DNA bending angle is therefore obtained from the tilt and roll angles while adjusting for twist. Total tilt (Θ_T) and roll (Θ_R) are obtained by rotating the tilt and roll angles through the accumulated twist: $\Theta_T = \Sigma_j(\tau_j \cos \gamma_j + \rho_j \sin \gamma_j)$ and $\Theta_R = \Sigma_j(-\tau_j \sin \gamma_j + \rho_j \cos \gamma_j)$, where the summation is over all dimer steps. γ_j is the accumulated twist from the reference step (N_C): $\gamma_j = -\Sigma_{i=N_C+1}^{j} \Omega_i$ for $j \geq N_C$ or $\gamma_j = \Sigma_{i=1}^{N_C-1} \Omega_i$ otherwise. N_C typically corresponds to the center of the DNA (also here). When N_C is fractional, step $j = \mathrm{int}(N_C)$ is split into two virtual steps, and ρ_j, τ_j, and Ω_j are fractionally distributed over the virtual steps. The DNA bending angle follows from $\varphi = \sqrt{\Theta_T^2 + \Theta_R^2}$, where $\Theta_R > 0$ indicates bending toward the major groove, and $\Theta_R < 0$ bending toward the minor groove (Figure 1c). More details on the Madbend procedure can be found in the original paper.[51]

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

Two-dimensional free energy surfaces (F) as a function of Θ_R and Θ_T were obtained from umbrella sampling[81] simulations. The original definitions of roll, twist and tilt angles require root-mean-square fits with idealized base pairs,[37] which would complicate and slow down the calculation of the biasing forces. Instead, we used our recently developed method to calculate step parameters from local coordinates.[82] This method bypasses overlays, yields analytical forces, shows good correlations with the original definitions (correlation coefficients of 0.998, 0.891, and 0.997 for roll, twist, and tilt, respectively), and is highly efficient. Integration of the two-dimensional free energy surface at constant values of φ gave the one-dimensional free energy profile for DNA bending:

$$F(\varphi) = -kt \ln\left(\int_{\sqrt{\Theta_R{}^2+\Theta_T{}^2}=\varphi} e^{-F(\Theta_R,\Theta_T)/kT}\, d\Theta_R\, d\Theta_T\right)$$

By restricting this integration to positive Θ_R values, the free energy profile of DNA bending toward the major groove was obtained. In a similar way, the free energy profile of DNA bending toward the minor groove was obtained by limiting the integration over negative Θ_R values.

A total of eight double strand DNA (dsDNA) sequences were studied: (1) CGCGAATTCGCG, (2) CGCGCGCGCGCG, (3) CCCTGTTCGGCG, (4) GATTGCGCAATG, (5) GCTATAAA-AGGC, (6) TATCCGCTTAAG, (7) CGTAGATCTACG, and (8) GCGATCGATCGC. Sequence 1 is the Dickerson dodecamer sequence;[83] its center two base pairs are mutated to CG in sequence 2. Sequences 3 and 4 were chosen from sequence similarity clustering of the PDB database and correspond to the most common core sequence that binds protein to the minor groove and bends toward the major groove (sequence 3), or binds the major groove and bends toward the minor groove (sequence 4), flanked by a CG base pair on each terminus. Sequences 5 and 6 correspond to the core DNA in specific examples of the database, with protein binding to the minor and DNA bending toward the major groove in the crystal structure of the human TATA binding protein complexed to DNA (PDB ID 1CDW[84]) for sequence 5; and binding to the major and bending toward the minor groove in the crystal structure of the *Escherichia coli* HipB transcriptional regulator bound to DNA (PDB ID 4YG1[85]) for sequence 6. While sequence 5 is bent toward the major groove in the protein–DNA crystal structure, the sequence is an A-tract, which are known to have a high preference for bending toward the minor groove.[40,43−46] Sequences 7 and 8 are known to maintain a stable B conformation.[86] While only a small number of sequences could be studied due to computational costs, the systems were selected to form a fair representation.

Unbent structures were built with X3DNA,[87] solvated into rectangle TIP3[88] water boxes of 0.15 M KCl, with a solvent layer of 18 Å in each direction. After minimization, the systems were gradually heated from 120 to 300 K over 1 ns, and then equilibrated for 1.2 ns. In these simulations, harmonic restraints with a mass-weighted force constant of 1 kcal/(mol Å^2) were applied to all heavy atoms of the DNA. The restraints were subsequently reduced to 0.5, 0.25, 0.1, and 0.01 kcal/(mol Å^2) in stages of 0.2 ns each, followed by a short unrestrained equilibration. The equilibrated structures were taken as starting points for 2D umbrella sampling simulations using the biasing potential $W = \frac{1}{2}k(\Theta_R - \Theta_R^{des})^2 + \frac{1}{2}k(\Theta_T - \Theta_T^{des})^2$, where Θ_R/Θ_T indicate the instantaneous values of the total roll and tilt angles, $\Theta_R^{des}/\Theta_T^{des}$ the desired values, and k = 0.04 kcal/(mol deg^2). Because of possible fraying, the first and last two base pairs were not included in the biasing. A total of 289 windows were used per sequence, with Θ_R and Θ_T each varying between −80 and 80° in steps of 10°. The first runs started with $\Theta_R^{des}/\Theta_T^{des}$ values closest to the Θ_R/Θ_T values at the end of the unrestrained equilibration; neighboring windows (with Θ_R^{des} 10° higher or lower, or with Θ_T^{des} 10° higher or lower) were run next, and so forth. Each umbrella sampling simulation started with a 0.1 ns equilibration, followed by a 0.5 ns production run. The equilibration end point was also used as the initial conditions for the equilibration of the neighboring windows. Step parameters are only defined with respect to hydrogen bonded base pairs;[37] to ensure hydrogen bonding, a flat bottom harmonic biasing potential was applied to the distance between the purine N_3 and pyrimidine N_1 of the base pairs. The force constant for this potential was 10.0 kcal/(mol Å^2) for distances larger than 3.3 Å, while no force was applied for distances below 3.3 Å. Analysis showed that the flat bottom biasing potential was rarely active and had no effect on the results.

All simulations were performed with an in-house modified version of the CHARMM program[89] that allowed for umbrella sampling in Θ_R and Θ_T. A detailed comparison of the CHARMM-embedded Madbend procedure with X3DNA[87] and Madbend[51] postprocessing is presented in the Supporting Information: results show excellent overall correlation, with a correlation coefficient of 0.973, and an inherent error of 5°. All simulations used the CHARMM 36 force field,[90] which includes the Beglov and Roux parameters for the ions.[91] The leapfrog integrator was used with a time step of 2 fs, snapshots were saved every 2 ps, SHAKE[92] was applied to all covalent bonds involving hydrogen atoms, the temperature was controlled with the Nosé–Hoover thermostat,[93] and long-range electrostatic interactions were handled by the particle mesh Ewald method.[94] For the Lennard–Jones interactions, a shifted potential cutoff of 11 Å was used. All free energy surfaces were calculated by the multistate Bennett acceptance ratio (MBAR) estimator after decorrelation of the trajectories.[95] Errors in the free energies were estimated from the MBAR uncertainty expressions.[95]

While kinking is an important DNA deformation and can lead to strong localized bends,[1,36] the effect of kinking is excluded from our study since the biasing force is distributed over all but the terminal two base pairs.

Persistence Lengths. Persistence lengths (A) were calculated from the one-dimensional free energy profiles of DNA bending by Mazur's procedure.[23] According to this method, the persistence length relates to the curvature of the free energy surface in the following way: $A = -L\frac{\partial \ln P}{\partial(1-\cos\varphi)} = \frac{L}{kT}\frac{\partial F}{\partial(1-\cos\varphi)}$, and can be obtained from fitting $F(\varphi)$ as a linear function of $(1-\cos\varphi)$. Here L indicates the contour length, P the probability distribution of bending angles, k the Boltzmann factor, and T the temperature. Contour lengths were obtained from the distance between the centers of mass of the terminal base pairs of the energy minimized DNA structures. Since DNA was slightly bent at equilibrium in the simulations (see below), fits were taken for bending angles past the equilibrium bending angle.

Hydration Analysis. Water molecules within 5.5 Å from the major groove edge atoms of the bases (Figure 1a) and on the major groove side were considered to be in the major groove, while water molecules within 5.0 Å from the minor groove edge atoms of the bases and on the minor groove side were considered to be in the minor groove. The terminal bases were not included for the hydration and residence time analyses. These definitions were verified graphically, and shown to work well for small and intermediate deviations from the equilibrium bending angle. Larger bending angles were not considered, since these led to significant deviations of the grooves (Figure 1c), which become harder to define.

RESULTS

Typical simulation snapshots are shown in Figure 1c for the CGCGCGCGCGCG sequence; snapshots for the other sequences were similar. At positive Θ_R values, DNA was bent toward the major groove; the major groove was compressed, while the minor groove was widened. Bending toward the minor groove, accompanied by a compression of the minor and widening of the major groove, was observed for negative Θ_R values. Figure 2 shows the two-dimensional free energy surface of DNA bending as a function of Θ_R and Θ_T. Error bars are not shown, but were less than 0.4 kcal/mol in all cases. Contour lines were concentric ovals, with the centers generally offset to Θ_R = 10° and Θ_T = −10°, and principal axes on diagonals. These results indicate that the equilibrium structure of most

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

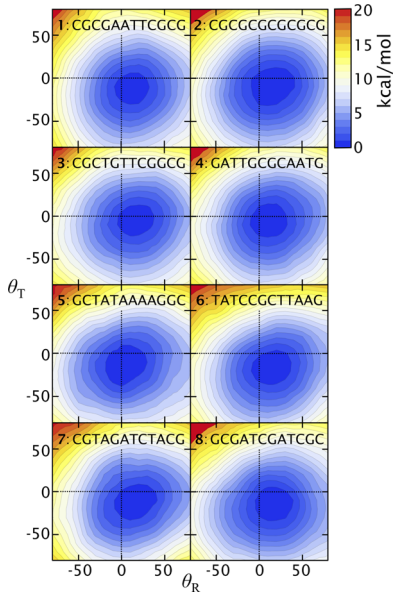


Figure 2. Free energy surfaces of DNA bending as a function of Θ_R and Θ_T. Index and sequence of the various strands as indicated.

sequences is slightly bent toward the major groove. A notable exception was the A-tract sequence, for which the free energy basin had its minimum at $(\Theta_R, \Theta_T) = (0°, -10°)$, indicating a higher preference for minor groove bending than the other sequences.

Integration of the two-dimensional surfaces led to the one-dimensional free energy curves of Figure 3. Shown are the free energy cost of bending toward the minor and major grooves as a function of the overall bending angle (φ), as well as the overall free energy cost of bending irrespective of the direction of bending. For clarity, error bars are not shown, but in all

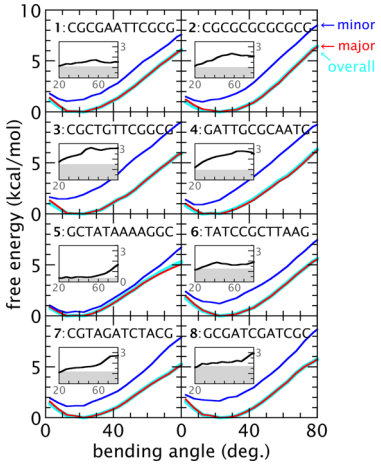


Figure 3. Free energy cost of DNA bending as a function of the overall bending angle ($\varphi = \sqrt{{\Theta_R}^2 + {\Theta_T}^2}$). Curves are shown for bending toward the minor and major grooves, as well as overall bending (i.e., irrespective of the direction). Insets show the difference in free energy for bending toward the minor and major grooves as a function of the bending angle ($\Delta F(\varphi)$). This is shown for bending angles $\geq 20°$ (past the minimum free energy configurations). Shown in gray $\Delta F_{eq} = F_{min}^{minor} - F_{min}^{major}$; the difference between the minimum free energy for minor groove bending and the minimum free energy for major groove bending. Index and sequence of the various strands as indicated.

cases, errors in the free energy were less than 0.2 kcal/mol at small and intermediate bending angles and less than 0.5 kcal/mol at large bending angles. As observed in previous studies,[19,20] the free energy of DNA bending was quadratic in bending angle for small and intermediate bending angles but became linear at large bending angles (>50−60°), indicating a significant deviation from behavior predicted by classical elastic models. In addition, Figure 3 shows several other interesting general features. The free energy curve for bending toward the major groove was generally lower than the curve for bending toward the minor groove (by a few kcal/mol), and consequently the free energy curves for overall bending and bending toward the major groove largely overlap. This observation indicates that bending toward the major groove is preferred. An exception was the A-tract, which showed overlap of the minor and major bending curves (a difference of 0.3 kcal/mol, which is within thermal energy). This is in agreement with experimental data, which shows that A-tract sequences easily bend toward the minor groove.[40,43−46] Integration of the two-dimensional free energy maps (Figure 2) led to a small shift in the location of the minima: while the minima of the two-dimensional map for non-A-tracts corresponded to an overall bending angle of ∼14°, the minima in the one-dimensional map (Figure 3) corresponded to an overall bending angle of ∼20°. The location of the minima indicates that the strands preferred to be slightly bent, by about 20° for major groove and overall bending, and by ∼15° for non-A-tract minor groove bending. While these bending angles seem somewhat large, which may partly be due to the way angles are calculated in the Madbend procedure, and partly be due to the integration, the magnitudes observed in the simulations are in line with what was observed in the PDB database (see below).

Apart from the shift in the position of the minimum, the curves for major and minor groove bending were similar in shape. The black curves in the insets of Figure 3 show how the difference in free energy between minor and major groove bending changes with the overall bending angle; this free energy difference will be denoted by $\Delta F(\varphi) = F^{minor}(\varphi) - F^{major}(\varphi)$. For sequences 1 and 6, $\Delta F(\varphi)$ was nearly constant, and equal to the difference between the minimum free energy of the minor and the minimum free energy of the major groove bending curves. This latter free energy difference will be denoted by $\Delta F_{eq} = F_{min}^{minor} - F_{min}^{major}$, and is shown in gray in the insets of Figure 3. For the A-tract and sequences 7 and 8, $\Delta F(\varphi) \approx \Delta F_{eq}$ for most of the bending (up until ∼60°). For sequence 2 and 3, $\Delta F(\varphi)$ increased slowly with φ, but never exceeded $2\Delta F_{eq}$, while for sequence 4 $\Delta F(\varphi)$ increased slowly to $2.5\Delta F_{eq}$. This shows that $\Delta F(\varphi)$ is largely dominated by ΔF_{eq} for most of the bending.

To further quantify the differences in bending, persistence lengths were calculated. By basing these calculations on three free energy curves, persistence lengths for overall bending as well as bending toward the major and minor groove could be obtained. These results are tabulated in Table 1; the fits are shown in Figure S2 of the Supporting Information. In general, the calculated overall persistence lengths agreed well with the experimental value of ∼500 Å.[1] As shown previously,[20,96] the observed increased flexibility at high bending angles for the short length scale did not imperil agreement with the worm-like chain model for the long length scale, because the Boltzmann probability of high bending is so small. In further agreement with experimental data,[97−99] sequence 2 was the stiffest with

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

Table 1. Calculated Persistence Lengths for Overall Bending and Bending Towards the Major and Minor Grooves

	sequence	$A_{overall}$ (Å)	A_{major} (Å)	A_{minor} (Å)	ΔA (%)[a]
1	CGCGAATTCGCG	538.5	539.5	551.1	2.2
2	CGCGCGCGCGCG	590.6	581.3	625.5	7.6
3	CCCTGTTCGGCG	536.1	533.4	623.7	16.9
4	GATTGCGCAATG	541.7	547.5	659.6	20.5
5	GCTATAAAAGGC	473.1	456.9	533.1	16.7
6	TATCCGCTTAAG	512.0	511.1	506.1	−1.0
7	CGTAGATCTACG	478.6	474.0	594.0	25.3
8	GCGATCGATCGC	557.3	555.7	627.8	13.0

[a] $\Delta A = 100\ (A_{minor} - A_{major})/A_{major}$.

the largest observed persistence length, and the A-tract had the lowest persistence length. In general, the persistence length for bending toward the minor groove was larger than the persistence length for bending toward the major groove, which indicates that bending through a given $\Delta\varphi$ is least costly toward the major groove. However, the relative difference in persistence length was small: 12.7% on average, with a maximum of difference of 25.3%.

Our calculations suggest that for non-A-tract sequences bending toward the major groove is more favorable because of two factors: (1) the free energy offset ΔF_{eq}, which signifies that in equilibrium, DNA is slightly bent toward the major groove, and (2) higher stiffness for bending toward the minor groove than bending toward the major groove. While both factors contribute, our analysis indicates that for most of the bending, the free energy offset is the most important factor. The persistence length for minor groove bending was on average only 12.7% larger than the persistence length for major groove bending, and $\Delta F(\varphi)$ was dominated by ΔF_{eq} for most of the bending. Thus, our data contributes a larger role to the free energy offset: bending toward the major groove is preferred, largely because DNA is slightly bent toward the major groove in equilibrium. The A-tract is different from the other sequences by having a free energy offset $\Delta F_{eq} \approx 0$, with a high tendency to be bent toward the minor groove in equilibrium.

The preference for major or minor groove bending is due to a combination of inherent structural properties of the DNA and solvation. Bending toward the major groove leads to a compression of the major and a widening of the minor groove (Figure 1c). One would therefore expect that major groove bending forces water out of the major and allows more water to enter the minor groove, and that the opposite occurs when DNA is bent toward the minor groove. Figure 4 shows that this was indeed observed in most of the simulations. The figure shows the number of groove waters as a function of the overall bending angle for bending toward the minor or major grooves. Since the grooves become hard to define when DNA becomes significantly bent, values are only shown for bending angles between 10 and 30° (about ±10° from equilibrium). Within error the major grooves gained water when bending toward the major groove was decreased or bending toward the minor groove was increased, while major groove water was lost when bending toward the major groove was increased or bending toward the minor groove was decreased. The opposite occurred for the water in the minor groove. While a few anomalies were seen, notably for CGTAGATCTACG (sequence 7), the hydration numbers generally followed the expected pattern.

Differences in hydration properties of the minor and major grooves are well-established; it is known that minor groove

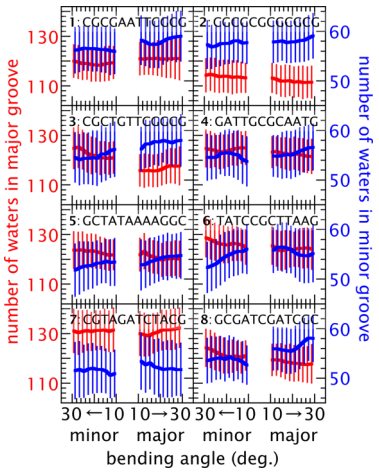


Figure 4. Hydration of the major and minor grooves upon bending. The number of waters in each groove are reported as a function of the overall bending angle, but separated for major and minor groove bending. Number of water molecules in major groove shown in red, minor in blue. The thick lines show averages, the vertical lines show the observed standard deviations. Index and sequence of the various strands as indicated.

water molecules have higher residence times and lower mobilities.[100−107] Higher minor groove residence times were also observed in our simulations. Figure 5 shows the average residence times for water molecules in either groove. These

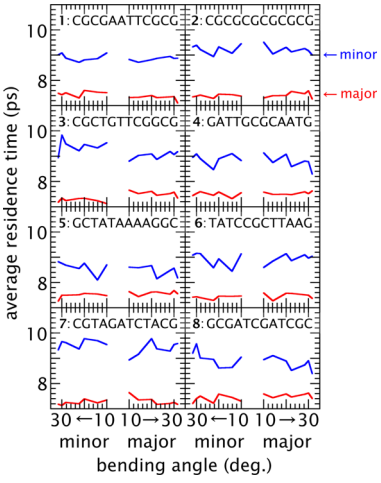


Figure 5. Average residence time of water molecules in the major (red) and minor (blue) grooves, reported as a function of the overall bending angle, but separated for major and minor groove bending directions. Index and sequence of the various strands as indicated.

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

were averaged over all observed residence times in each groove, from the long residence times (tens to hundreds of ps[108−110]) of water molecules deep in the groove, to the short residence times of the waters near the surface of the groove that frequently enter and leave. The lower residence time of water molecules in the major groove (Figure 5) indicates that these waters are easier to displace than water molecules from the minor groove. This was echoed in the average per-water interaction energies between DNA and the groove water molecules, which were more favorable for water in the minor than water in the major groove (Figure S3 of the Supporting Information).

Since water is more easily exchanged from the major than the minor groove, the major groove likely also more easily loses water molecules upon bending. Solvation would therefore favor bending toward the major groove, which releases more easily exchangeable major groove water molecules rather than the more confined minor groove waters. This has two important implications. First, the smaller persistence lengths for major groove bending are likely due to this solvation effect. This also means that bending of the A-tract, for which the persistence length is also smaller for major than for minor groove bending (Table 1), is affected by solvation in the same way as the other sequences; in fact, its larger persistence length in the minor groove direction is consistent with the presence of a low mobility spine of hydration in the minor groove of A-tracts.[108−110] The second implication is that for the non-A-tract sequences the preference for major groove bending is likely largely driven by solvation. In contrast, for the A-tract, the high tendency for minor groove bending is driven by its inherent structural properties (for example, the preference for negative rolls at the A-tract junction),[40] which affect ΔF_{eq}. Solvation plays a role when bending an A-tract away from equilibrium, leading to slightly less resistance in the major groove direction, but at equilibrium a high proportion of A-tracts is bent toward the minor groove.

Aspects of our calculations could be verified by performing a statistical analysis of DNA bending in protein−DNA structures from the Protein Data Bank. Figure 6 shows the distribution of Madbend bending angles for proteins that bend DNA toward the minor and major grooves. The figure shows a higher population at large bending angles for DNAs that bend toward the major groove, which indicates that bending toward the major groove might indeed be less costly. In agreement with the simulations, the distributions also showed a preferred bending angle between 10 and 20° for major groove benders. As observed in the non-A-tract simulations, the preferred bending angle for minor groove benders is shifted to a lower value than that for major groove benders, but the PDB statistics showed a preference between 0 and 10° instead of the 15° found in the simulations. The statistical analyses revealed some other notable aspects as well. While proteins that bind to the major groove indeed tend to bend DNA toward the minor groove, and vice versa,[51] a significant number of exceptions were found, especially for proteins that bind to the major groove. Of the proteins that bind to the major groove, 63% were seen to bend toward the minor groove and 37% toward the major groove. For proteins that bind the minor groove, the vast majority (84%) was observed to bend toward the major groove, while a minority (16%) bent DNA toward the minor groove.

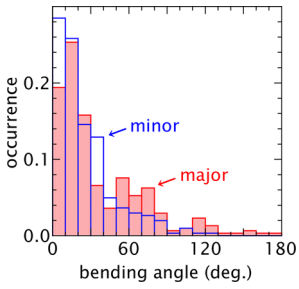


Figure 6. Distribution of DNA bending angles in protein−DNA complexes for proteins that bend DNA toward the major (red) and minor groove (blue).

CONCLUSION

The anisotropy of B-DNA bending toward the major and minor grooves was quantified by a new free energy simulation approach for eight different sequences. The simulations showed that non-A-tract sequences preferably bend toward the major groove, while the A-tract has a high tendency to bend toward the minor groove. An overall tendency for major groove bending was also observed in protein data bank structures of protein−DNA complexes. For non-A-tract sequences, bending toward the major groove is favored because of two factors: a free energy offset ΔF_{eq}, which favors equilibrium structures that are slightly bent toward the major groove, and a smaller stiffness for major groove bending. Both factors are highly affected by differences in groove hydration. The relative ease by which the major groove loses water upon bending reduces resistance toward major groove bending, and is likely the determining factor of the smaller persistence length for major groove bending and the origin of the free energy offset for non-A-tracts. For the A-tract, stiffness toward major groove bending is also smaller than for minor groove bending, but the free energy offset favors equilibrium structures that are bent toward the minor groove. This is likely due to inherent structural properties of the A-tract. Overall, the free energy offset is the dominant factor in determining toward which groove DNA is more easily bent.

Our studies hint at how proteins might use dehydration to aid DNA bending, and imply a role for hydration in selecting the bending angle in protein−DNA complexes.[73] In conlusion, this study provides important new insights into the energetics of DNA bending, which is critical to a large number of life processes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05136.

Discussion of accuracy of the CHARMM-embedded Madbend procedure, Protein Data Bank access codes for the protein−DNA structures used in statistical analysis of DNA bending angles, fits used in calculation of persistence lengths, and average per-water interaction energies between DNA and groove water (PDF)

AUTHOR INFORMATION

Corresponding Author

*avandervaart@usf.edu

Notes

The authors declare no competing financial interest.

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

■ ACKNOWLEDGMENTS

This work was supported by NSF CAREER Award CHE-1007816. Computer time was provided by USF Research Computing, sponsored in part by NSF MRI CHE-1531590.

■ REFERENCES

(1) Peters, J. P.; Maher, L. J. *Q. Rev. Biophys.* **2010**, *43*, 23.
(2) Luijsterburg, M. S.; White, M. F.; van Driel, R.; Dame, R. T. *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 393.
(3) Matthews, K. S. *Microbiol. Rev.* **1992**, *56*, 123.
(4) Saiz, L.; Vilar, J. M. G. *Curr. Opin. Struct. Biol.* **2006**, *16*, 344.
(5) Schleif, R. *Annu. Rev. Biochem.* **1992**, *61*, 199.
(6) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. *Structure* **2000**, *8*, 1015.
(7) van der Vaart, A. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1091.
(8) Hogan, M. E.; Austin, R. H. *Nature* **1987**, *329*, 263.
(9) Ramstein, J.; Lavery, R. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 7231.
(10) Cloutier, T. E.; Widom, J. *Mol. Cell* **2004**, *14*, 355.
(11) Kratky, O.; Porod, G. *Rec. Trav. Chim. Pays-Bas* **1949**, *68*, 1106.
(12) Strick, T.; Allemand, J. F.; Croquette, V.; Bensimon, D. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 115.
(13) Wiggins, P. A.; van der Heijden, T.; Moreno-Herrero, F.; Spakowitz, A.; Phillips, R.; Widom, J.; Dekker, C.; Nelson, P. C. *Nat. Nanotechnol.* **2006**, *1*, 137.
(14) Vafabakhsh, R.; Ha, T. *Science* **2012**, *337*, 1097.
(15) Yuan, C. L.; Chen, H. M.; Lou, X. W.; Archer, L. A. *Phys. Rev. Lett.* **2008**, *100*, 018102.
(16) Weber, G.; Essex, J. W.; Neylon, C. *Nat. Phys.* **2009**, *5*, 769.
(17) Cloutier, T. E.; Widom, J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 3645.
(18) Lankas, F.; Lavery, R.; Maddocks, J. H. *Structure* **2006**, *14*, 1527.
(19) Curuksu, J.; Zacharias, M.; Lavery, R.; Zakrzewska, K. *Nucleic Acids Res.* **2009**, *37*, 3766.
(20) Spiriti, J.; Kamberaj, H.; De Graff, A.; Thorpe, M. F.; van der Vaart, A. *J. Chem. Theory Comput.* **2012**, *8*, 2145.
(21) Taranova, M.; Hirsh, A. D.; Perkins, N. C.; Andricioaei, I. *J. Phys. Chem. B* **2014**, *118*, 11028.
(22) Du, Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5397.
(23) Mazur, A. K. *Phys. Rev. Lett.* **2007**, *98*, 218102.
(24) Mastroianni, A. J.; Sivak, D. A.; Geissler, P. L.; Alivisatos, A. P. *Biophys. J.* **2009**, *97*, 1408.
(25) Vologodskii, A.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2013**, *41*, 6785.
(26) Berg, J. M.; Tymoczko, J. L.; Gatto, J.; Stryer, L. *Biochemistry*, 8th ed.; W. H. Freeman: New York, 2015.
(27) Schellman, J. A. *Biopolymers* **1974**, *13*, 217.
(28) Zhurkin, V. B.; Lysov, Y. P.; Ivanov, V. I. *Nucleic Acids Res.* **1979**, *6*, 1081.
(29) Balaeff, A.; Mahadevan, L.; Schulten, K. *Phys. Rev. E* **2006**, *73*, 031919.
(30) Gromiha, M. M.; Munteanu, M. G.; Gabrielian, A.; Pongor, S. *J. Biol. Phys.* **1996**, *22*, 227.
(31) Munteanu, M. G.; Vlahovicek, K.; Parthasarathy, S.; Simon, I.; Pongor, S. *Trends Biochem. Sci.* **1998**, *23*, 341.
(32) Salari, H.; Eslami-Mossallam, B.; Naderi, S.; Ejtehadi, M. R. *J. Chem. Phys.* **2015**, *143*, 104904.
(33) Eslami-Mossallam, B.; Ejtehadi, M. R. *Phys. Rev. E* **2009**, *80*, 011919.
(34) Yanagi, K.; Prive, G. G.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 201.
(35) Young, M. A.; Ravishanker, G.; Beveridge, D. L.; Berman, H. M. *Biophys. J.* **1995**, *68*, 2454.
(36) Dickerson, R. E. *Nucleic Acids Res.* **1998**, *26*, 1906.
(37) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229.
(38) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11163.
(39) Perez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **2004**, *32*, 6144.
(40) Haran, T. E.; Mohanty, U. *Q. Rev. Biophys.* **2009**, *42*, 41.
(41) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 7664.
(42) Hagerman, P. J. *Biochemistry* **1985**, *24*, 7033.
(43) Zinkel, S. S.; Crothers, D. M. *Nature* **1987**, *328*, 178.
(44) Stefl, R.; Wu, H. H.; Ravindranathan, S.; Sklenar, V.; Feigon, J. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1177.
(45) Barbic, A.; Zimmer, D. P.; Crothers, D. M. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2369.
(46) MacDonald, D.; Herbert, K.; Zhang, X.; Polgruto, T.; Lu, P. *J. Mol. Biol.* **2001**, *306*, 1081.
(47) Ulyanov, N. B.; Zhurkin, V. B. *J. Biomol. Struct. Dyn.* **1984**, *2*, 361.
(48) Wang, D.; Ulyanov, N. B.; Zhurkin, V. B. *J. Biomol. Struct. Dyn.* **2010**, *27*, 843.
(49) Sanghani, S. R.; Zakrzewska, K.; Harvey, S. C.; Lavery, R. *Nucleic Acids Res.* **1996**, *24*, 1632.
(50) Curuksu, J.; Zakrzewska, K.; Zacharias, M. *Nucleic Acids Res.* **2008**, *36*, 2268.
(51) Strahs, D.; Schlick, T. *J. Mol. Biol.* **2000**, *301*, 643.
(52) Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1998**, *281*, 675.
(53) Sprous, D.; Young, M. A.; Beveridge, D. L. *J. Mol. Biol.* **1999**, *285*, 1623.
(54) McConnell, K. J.; Beveridge, D. L. *J. Mol. Biol.* **2001**, *314*, 23.
(55) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981.
(56) Balasubramanian, C.; Ojha, R. P.; Maiti, S. *Biochem. Biophys. Res. Commun.* **2007**, *355*, 1081.
(57) Madhumalar, A.; Bansal, M. *Biophys. J.* **2003**, *85*, 1805.
(58) Mocci, F.; Saba, G. *Biopolymers* **2003**, *68*, 471.
(59) Nikolova, E. N.; Bascom, G. D.; Andricioaei, I.; Al-Hashimi, H. M. *Biochemistry* **2012**, *51*, 8654.
(60) Hamelberg, D.; McFail-Isom, L.; Williams, L. D.; Wilson, W. D. *J. Am. Chem. Soc.* **2000**, *122*, 10513.
(61) Zhu, X.; Schatz, G. C. *J. Phys. Chem. B* **2012**, *116*, 13672.
(62) Lankas, F.; Spackova, N.; Moakher, M.; Enkhbayar, P.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 3414.
(63) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. *Biophys. J.* **2004**, *87*, 3799.
(64) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., 3rd; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89*, 3721.
(65) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 299.
(66) Beveridge, D. L.; Cheatham, T. E., III; Mezei, M. *J. Biosci. (New Delhi, India)* **2012**, *37*, 379.
(67) Spiriti, J.; Kamberaj, H.; van der Vaart, A. *Int. J. Quantum Chem.* **2012**, *112*, 33.
(68) Mitsutake, A.; Mori, Y.; Okamoto, Y. *Methods Mol. Biol.* **2013**, *924*, 153−195.
(69) Spiriti, J.; van der Vaart, A. *J. Phys. Chem. Lett.* **2012**, *3*, 3029.
(70) Sharma, M.; Predeus, A. V.; Mukherjee, S.; Feig, M. *J. Phys. Chem. B* **2013**, *117*, 6194.
(71) Spiriti, J.; van der Vaart, A. *ChemBioChem* **2013**, *14*, 1434.
(72) Chung, Y.-H.; van der Vaart, A. *ChemBioChem* **2014**, *15*, 643.
(73) Barr, D.; van der Vaart, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2070.
(74) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63.
(75) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655.

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958

(76) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. *Nucleic Acids Res.* **2009**, *37*, 5917.
(77) Dickerson, R. E.; Chiu, R. K. *Biopolymers* **1997**, *44*, 361.
(78) Goodsell, D. S.; Dickerson, R. E. *Nucleic Acids Res.* **1994**, *22*, 5497.
(79) Tan, R. K.-Z.; Harvey, S. C. *J. Biomol. Struct. Dyn.* **1987**, *5*, 497.
(80) Barbic, A.; Crothers, D. M. *J. Biomol. Struct. Dyn.* **2003**, *21*, 89.
(81) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(82) Karolak, A.; van der Vaart, A. *J. Comput. Chem.* **2014**, *35*, 2297.
(83) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755.
(84) Nikolov, D. B.; Chen, H.; Halay, E. D.; Hoffman, A.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 4862.
(85) Schumacher, M. A.; Balani, P.; Min, J.; Chinnam, N. B.; Hansen, S.; Vulic, M.; Lewis, K.; Brennan, R. G. *Nature* **2015**, *524*, 59.
(86) Basham, B.; Schroth, G. P.; Ho, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 6464.
(87) Lu, X.-J.; Olson, W. K. *Nat. Protoc.* **2008**, *3*, 1213.
(88) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(89) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.
(90) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8*, 348.
(91) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.
(92) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(93) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
(94) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(95) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
(96) Wiggins, P. A.; Nelson, P. C. *Phys. Rev. E* **2006**, *73*, 31906.
(97) Hogan, M.; LeGrange, J.; Austin, B. *Nature* **1983**, *304*, 752.
(98) Bednar, J.; Furrer, P.; Katritch, V.; Stasiak, A. Z.; Dubochet, J.; Stasiak, A. *J. Mol. Biol.* **1995**, *254*, 579.
(99) Chen, H. H.; Rau, D. C.; Charney, E. *J. Biomol. Struct. Dyn.* **1985**, *2*, 709.
(100) Phan, A. T.; Leroy, J.-L.; Guéron, M. *J. Mol. Biol.* **1999**, *286*, 505.
(101) Liepinsh, E.; Leupin, W.; Otting, G. *Nucleic Acids Res.* **1994**, *22*, 2249.
(102) Denisov, V. P.; Carlström, G.; Venu, K.; Halle, B. *J. Mol. Biol.* **1997**, *268*, 118.
(103) Johannesson, H.; Halle, B. *J. Am. Chem. Soc.* **1998**, *120*, 6859.
(104) Kubinec, M. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 8739.
(105) Liepinsh, E.; Otting, G.; Wüthrich, K. *Nucleic Acids Res.* **1992**, *20*, 6549.
(106) Jana, B.; Pal, S.; Bagchi, B. *J. Phys. Chem. B* **2010**, *114*, 3633.
(107) Saha, D.; Supekar, S.; Mukherjee, A. *J. Phys. Chem. B* **2015**, *119*, 11371.
(108) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535.
(109) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1983**, *163*, 129.
(110) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S. *J. Mol. Biol.* **1992**, *226*, 1161.

DOI: 10.1021/jacs.6b05136
J. Am. Chem. Soc. 2016, 138, 9951−9958