

DNA Bending through Large Angles Is Aided by Ionic Screening

Justin Spiriti,[†] Hiqmet Kamberaj,^{‡,§,||,⊥,#} Adam M. R. de Graff,^{§,||,⊥,#} M. F. Thorpe,^{§,||} and Arjan van der Vaart^{*,†}

[†]Department of Chemistry, University of South Florida, 4202 E Fowler Ave. CHE 205, Tampa, Florida 33620, United States

[‡]Faculty of Technical Sciences, International Balkan University, Skopje, Republic of Macedonia

[§]Center for Biological Physics, Arizona State University, Tempe, Arizona 85287, United States

^{||}Department of Physics, Arizona State University, P.O. Box 871504, Tempe, Arizona 85287, United States

[⊥]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, United States

Supporting Information

ABSTRACT: We used adaptive umbrella sampling on a modified version of the roll angle to simulate the bending of DNA dodecamers. Simulations were carried out with the AMBER and CHARMM force fields for 10 sequences in which the central base pair step was varied. On long length scales, the DNA behavior was found to be consistent with the worm-like chain model. Persistence lengths calculated directly from the simulated structures and indirectly through the use of sequence-independent coarse-grained models based on simulation data were similar to literature values. On short length scales, the free energy cost of bending DNA was found to be consistent with the worm-like chain model for small and intermediate bending angles. At large angles, the bending free energy as a function of the roll angle became linear, suggesting a relative increase in flexibility at larger roll angles. Counterions congregated on the concave side of the highly bent DNA and screened the repulsion of the phosphate groups, facilitating the bending.

INTRODUCTION

The inherent flexibility of DNA is an important part of many biological processes. Most transcription factors and other DNA-binding proteins recognize their target sequences by specific contacts between amino acids and DNA bases (direct readout) as well as by recognizing the flexibility of the sequence (indirect readout).¹ Moreover, the binding of DNA to a protein often involves substantial bending,² which has important consequences for the binding thermodynamics.³ The bending of DNA is also important to its packaging. In eukaryotes, DNA is wound tightly around nucleosomes.⁴ The tightness of the DNA wrapping can be modulated by chemical modifications to the histones; this epigenetic modification is thought to be an important control mechanism for DNA expression.⁵

The most successful model for the inherent flexibility of DNA is the worm-like chain (WLC) model,⁶ which describes the inherent flexibility of a polymer in terms of the persistence length (PL). On long length scales, this model describes DNA well, and the PL of DNA has been measured to be about 150 base pairs or 500 Å by a variety of methods.⁷ The ability of the WLC model to describe DNA on short length scales is less clear; experimental studies have yielded conflicting results. DNA cyclization,⁸ atomic force microscopy,⁹ small-angle X-ray scattering,¹⁰ and a combination of fluorescence resonance energy transfer and small-angle X-ray scattering¹¹ all provide evidence that DNA is more flexible on short length scales than predicted by the WLC model. On the other hand, other experiments using many of the same techniques failed to reproduce these findings.¹²

There is also debate about the nature of the conformation of DNA on smaller scales when it is significantly bent. Early

proposals included significant bending at a single base pair (type I kink) through 90°¹³ or 45°.¹⁴ Similar kinking is observed in the nucleosome core particle, although with smaller angles.^{4b} Later, it was suggested that bending also involves localized base flipping or melting of the DNA (forming a type II kink).¹⁵ There is evidence that localized DNA melting increases flexibility,¹⁶ but this may not be enough to explain the observations mentioned above.¹⁷ In addition, transient Hoogsteen base pairs have recently been observed by NMR.¹⁸ Structures of DNA containing type I and type II kinks, taken from crystal structures of protein–DNA complexes, are shown in Figure 1.

There appear to be two major forces leading to the observed stiffness of DNA: the electrostatic repulsion between phosphate groups on the backbone and the stacking interaction between adjacent base pairs.¹⁹ Experiments provide contradictory results on the question of which of these forces is most important. If the electrostatic repulsion between phosphate groups is an important force, it will be screened by counterions,²⁰ so the distribution of counterions around DNA will be expected to have an impact on the flexibility of DNA. Several experimental studies of asymmetric neutralization of the phosphate backbone²¹ have shown that this neutralization increases the flexibility of DNA. On the other hand, the PL of DNA has been found to be sequence dependent,²² and the “meroduplex” (a single strand of DNA base-paired with free bases) has been found to have similar rigidity to double-stranded DNA.²³

Received: March 1, 2012

Published: April 30, 2012



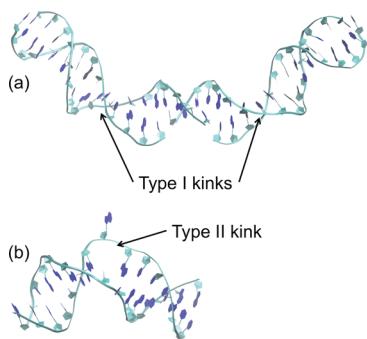


Figure 1. Structure of type I and type II kinks in DNA, as seen in protein–DNA crystal structures. Bound proteins are not shown for clarity. (a) Type I kinks in DNA bound to the *E. coli* catabolite activator protein (PDB code 1JS9).⁶⁷ (b) Type II kink in DNA bound to human 8-oxoguanine DNA glycosylase (PDB code 1EBM).⁶⁸

In principle, molecular dynamics simulations can resolve these issues, since simulations permit the direct observation of conformational changes in DNA at the atomic scale.²⁴ Some of the previously reported efforts to simulate DNA flexibility and dynamics have been hampered by insufficient sampling, however. A molecular dynamics simulation study of small DNA minicircles obtained evidence for occasional kink formation in DNA, but did not obtain enough kinks for statistical characterization due to insufficient sampling.²⁵ Systematic studies of all 136 DNA tetranucleotides have characterized the sequence dependence of DNA flexibility^{24a–c} but did not sample significant overall deviations from the B-DNA structure for any of the 39 strands simulated. Some of these studies^{24a–c,26} may also have been hampered by issues involving the treatment of backbone dihedrals in the AMBER force fields, which were solved with the introduction of the AMBER parambsc0 force field.²⁷ Umbrella sampling simulations of DNA have been performed by Zacharias et al. using a screw-axis reaction coordinate corresponding to the overall bending angle.^{26,28} This coordinate is defined for a conformation of a whole DNA strand, by creating two “handle” vectors based on averaging vectors from the local reference frames of the four base pairs on each end. This technique allowed determination of the free energy surface of DNA in terms of the magnitude and direction of DNA bending, and the simulations showed that the free energy of bending is quadratic in the bending angle for small bending angles but eventually becomes linear.²⁸ The physical reasons for the observed linearity were left unexplained, however.²⁸

There have also been attempts to separate the electrostatic and nonelectrostatic contributions to DNA flexibility using simulations of DNA. For example, the increased flexibility of asymmetrically neutralized DNA has been confirmed in simulations.^{21c} In addition, comparison of the persistence length of fully charged and neutralized DNA simulated using both coarse-grained and atomistic molecular dynamics suggests that electrostatic and nonelectrostatic interactions make approximately equal contributions to the persistence length.²⁹ It is not clear how accurately a coarse-grained force field for DNA can capture the complex interactions between the phosphate backbone, water, and ions; moreover, the contributions of electrostatic and nonelectrostatic interactions to the persistence length need not be additive.

Studies of the structures of protein–DNA complexes show that most of the DNA bending flexibility is due to changes in

the roll angle.³⁰ Consequently, in this work, we modify the calculation of the roll angle for use as a reaction coordinate for adaptive umbrella sampling³¹ and apply this simulation method to study DNA flexibility. In contrast to the coordinate used by Zacharias, our coordinate is more localized and biases fewer atoms. It is thus better suited for studying DNA bending in protein–DNA complexes, where the bending usually involves only a few specific base pair steps. In addition, our coordinate opens up the possibility of determining free energy surfaces for multiple roll angles using a multidimensional adaptive umbrella sampling method. A drawback of the method is its inability to study base flipping or DNA melting, due to the fact that roll angles become ill-defined for nonbase paired steps. This is not a problem for most protein–DNA complexes, since most complexes do not involve melting or base flipping. For bare DNA, the method will yield a conservative estimate of the flexibility since it excludes the possible role of type II kinks.

In this first application, we enhanced the sampling of one roll angle for bare DNA strands. We studied 10 possible dinucleotide fragments using this technique, by embedding each dinucleotide in the center of a symmetric DNA dodecamer. In addition, each sequence was simulated using the AMBER parambsc0 force field²⁷ as well as the CHARMM force field³² to permit a comparison between the force fields. The long-range behavior of DNA observed in our simulations is consistent with the WLC model, as indicated by PLs calculated directly from the simulations and from a simple coarse-grained model constructed using data from the simulations. Although our computational framework precluded the assessment of the effect of type II kinks on DNA flexibility, we obtained substantial bending and observed a shift of the dependence of the free energy on roll angle from quadratic to linear for large angles, suggesting a relative increase of DNA flexibility at high roll angles. We also studied the distribution of ions around strongly bent DNA and found that cations congregate on the concave side of the DNA. Our findings show that DNA becomes more flexible on short length scales when highly bent and that this change does not require melting or bubble formation of the DNA. As DNA bends, cations from the surrounding solution gather on the concave side, screening the repulsion between phosphate groups and reducing the resistance to further DNA bending.

METHODS

Definition of the Pseudoroll Angle. Several programs have been developed for the analysis of nucleic acid structures and computation of base pair step parameters. These include FREEHELIX,³⁰ CEHS,³³ CURVES,³⁴ Curves+,³⁵ NUPARM,³⁶ RNA,³⁷ CompDNA,³⁸ and 3DNA.³⁹ The differences between these programs have been reviewed elsewhere.⁴⁰ The definition of roll angle used in this study is similar to the definitions embedded in these programs, especially FREEHELIX,³⁰ but has been modified to make it more suitable for use as a reaction coordinate for adaptive umbrella sampling. The mathematical definition has been simplified in order to make it much quicker and easier to calculate the derivatives with respect to the Cartesian coordinates of the atoms, which are needed to calculate the forces due to the biasing potential. The new definition of roll angle will be called the “pseudoroll angle” to distinguish it from the roll angle as defined in nucleic acid analysis programs.

The pseudoroll angle $\rho_{i,i+1}$ between base pairs i and $i+1$ is given by $\cos \rho_{i,i+1} = \mathbf{a}_1 \cdot \mathbf{a}_2$, where

$$\begin{aligned} \mathbf{a}_1 &= \frac{\mathbf{P}_i \times \mathbf{L}_{i,i+1}}{|\mathbf{P}_i \times \mathbf{L}_{i,i+1}|} \\ \mathbf{a}_2 &= \frac{\mathbf{P}_{i+1} \times \mathbf{L}_{i,i+1}}{|\mathbf{P}_{i+1} \times \mathbf{L}_{i,i+1}|} \end{aligned} \quad (1)$$

In this equation, \mathbf{P}_i is the normal vector to base pair i and is defined by

$$\mathbf{P}_i = \frac{\mathbf{P}_R + \mathbf{P}_Y}{|\mathbf{P}_R + \mathbf{P}_Y|} \quad (2)$$

where \mathbf{P}_R is the unit normal vector to the purine within the base pair, which is perpendicular to the plane that passes through atoms N3, C6, and N7 of the purine ring. \mathbf{P}_Y is the unit normal vector to the pyrimidine within the base pair and is perpendicular to the plane that passes through atoms C2, C4, and C6 of the pyrimidine ring. $\mathbf{L}_{i,i+1}$ is the unit vector along the line between the centers of mass of the atoms on the first strand that are part of base pairs i and $i+1$ and the atoms on the second strand that are part of the same base pairs. The sign of $\rho_{i,i+1}$ is the same as the sign of $\mathbf{P}_i \cdot (\mathbf{P}_{i+1} \times \mathbf{L}_{i,i+1})$. This definition is illustrated in Figure 2. A major advantage of the pseudoroll

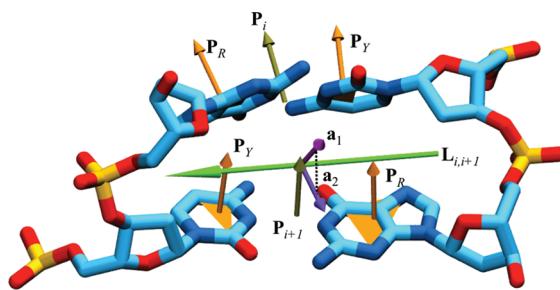


Figure 2. Definition of the pseudoroll angle. \mathbf{P}_R and \mathbf{P}_Y are the normal vectors to individual bases, and \mathbf{P}_i and \mathbf{P}_{i+1} are the normal vectors to the base pairs defined in eq 2. $\mathbf{L}_{i,i+1}$ is the vector between the centers of mass of the atoms on the two strands that are part of base pairs i and $i+1$. \mathbf{a}_1 and \mathbf{a}_2 are the cross products between \mathbf{P}_i and $\mathbf{L}_{i,i+1}$ and between \mathbf{P}_{i+1} and $\mathbf{L}_{i,i+1}$, respectively (eq 1). The roll angle $\rho_{i,i+1}$ is defined as the signed angle between \mathbf{a}_1 and \mathbf{a}_2 , which represents the angle between base pair normal vectors in the plane perpendicular to $\mathbf{L}_{i,i+1}$.

angle is that no RMSD fit with respect to an idealized base pair is needed, speeding up the calculation significantly. The pseudoroll angle was tested for correlation with the actual roll angle as defined by 3DNA³⁹ using a survey of protein–DNA crystal structures as described in the Supporting Information; a list of the PDB codes of the crystal structures is shown in Table S1.

Adaptive Umbrella Sampling Simulations. Adaptive umbrella sampling³¹ is a popular technique for determining a free energy surface in terms of a specified reaction coordinate (here, the central roll angle of a DNA dodecamer). This method determines the free energy iteratively from a sequence of simulations. During each simulation, histograms are collected to count the number of times configurations along the reaction coordinate are visited. The first simulation is unbiased, and the system is propagated using the normal force field. In subsequent simulations, a biasing potential is added to the force field. This biasing potential equals the negative of the estimated free energy surface for the normal, unbiased system. This free energy is calculated from the histograms of all previous simulations, using reweighting to account for the

presence of the biasing potential. Initially, the calculated free energy surface is inaccurate, since only a small part of the space near the initial configuration will have been sampled. The biasing potential artificially increases the energy of visited configurations however, which forces the system to explore new areas. Eventually, the biasing potential will remove all free energy barriers, and sampling will become diffusive, visiting all parts of the reaction coordinate space with equal probability. When this happens, the calculated free energy surface will not change upon new iterations, and the simulation has converged. Illustrative examples of the adaptive umbrella sampling method include studies of the *trans*-to-*cis* isomerization of a peptide bond in a bacterial collagenase⁴¹ and the rotamer dynamics of a threonine in a designed peptide.⁴²

Adaptive umbrella sampling simulations³¹ were performed to determine the free energy surface of DNA dodecamers as a function of the central roll angle. Ten DNA sequences of the form 5'-CGCGANNTCGCG-3' were studied, where NN is one of the 10 possible base dimers (GG, GA, GC, GT, AA, AG, AT, TA, TG, CG). In each simulation, sampling of the central pseudoroll angle was enhanced. Each system was simulated using both the CHARMM27³² and AMBER parambsc0²⁷ force fields; thus, a total of 20 simulations were done. The simulations were performed using an in-house modified version of CHARMM⁴³ in which the ADUMB module (which performs adaptive umbrella sampling) was interfaced with the RXNCOR⁴⁴ module (which defines the reaction coordinates). All simulations were performed in a ~150 mM KCl solution of TIP3P explicit water;⁴⁵ the water layer between the DNA heavy atoms and the edge of the rhombic dodecahedral box was at least 12 Å. Ions were initially positioned using the SOLVATE program.⁴⁶ Periodic boundary conditions were used with the particle mesh Ewald method⁴⁷ to treat long-range electrostatic interactions, and a shifting function was used to treat the van der Waals interactions. SHAKE constraints⁴⁸ were employed for all bonds containing hydrogen atoms, and a time step of 2 fs was used. Construction of the initial coordinates, heating, and equilibration of each system are described in the Supporting Information.

In the production runs, adaptive umbrella sampling³¹ was applied to the pseudoroll angle for the central base pair step as described above. During each simulation, histograms of the pseudoroll angle with bin size 1° were collected. Every 500 ps, the potential of mean force in terms of the pseudoroll angle was calculated using all previous histograms from all runs using the weighted histogram analysis method (WHAM),⁴⁹ and the new biasing potential U_{bias} was calculated from the potential of mean force using

$$U_{\text{bias}} = \beta^{-1} \ln(e^{-\beta F(\rho)} + e^{-\beta U_{\text{max}}}) \quad (3)$$

where $F(\rho)$ is the free energy surface calculated using WHAM,⁴⁹ β is the inverse temperature, and U_{max} is a cap on the biasing potential. This cap was imposed in order to prevent numerical instabilities in the use of this coordinate. For example, if one of the bases in a base pair flipped out of the DNA during the simulation, the vectors \mathbf{P}_R and \mathbf{P}_Y for this base pair could point in nearly opposite directions. If this happened, $|\mathbf{P}_R + \mathbf{P}_Y|$ became a very small number, causing \mathbf{P}_i to become numerically unstable. This instability typically resulted in large biasing forces and ultimately in a failure to conserve energy, as well as the flipping of bases or the formation of Hoogsteen base pairs. These are very high energy structures; the experimental

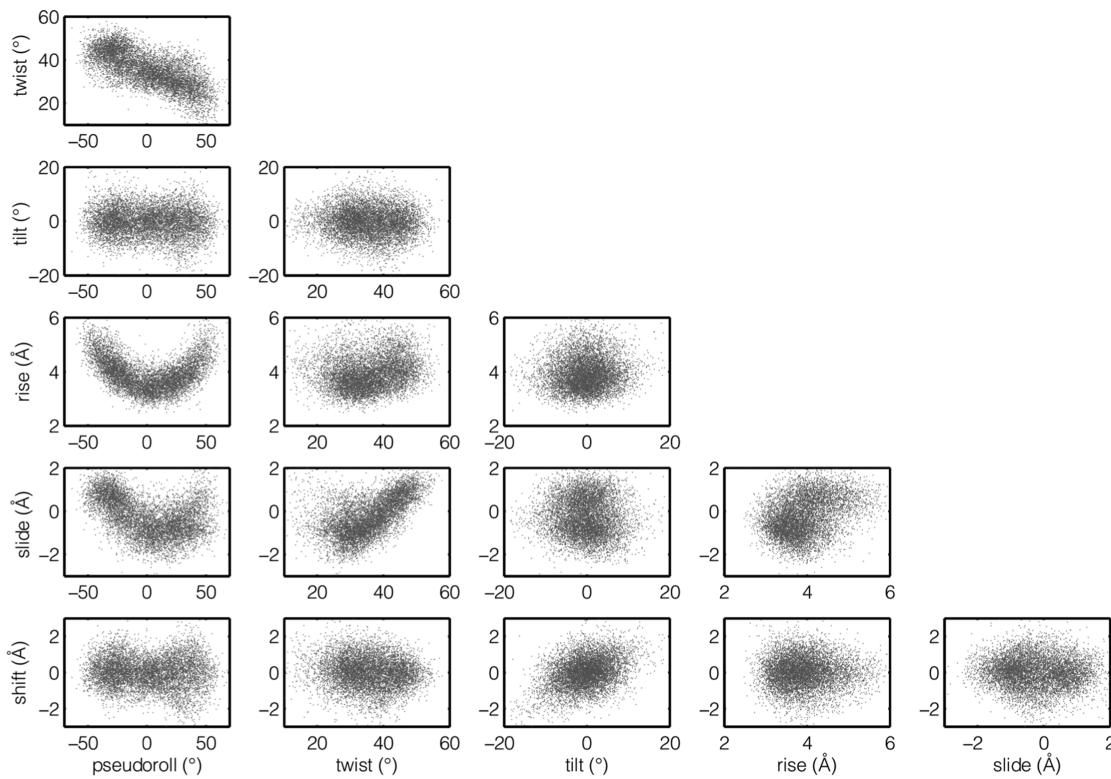


Figure 3. Scatterplots of pairs of base pair step parameters.

activation energies for these processes have been estimated to be at least 12–15 kcal/mol in otherwise straight DNA.^{18,50} By biasing the simulation toward extreme values of the pseudoroll angle, the adaptive umbrella sampling method encouraged these numerical instabilities when a cap was not used.

The value of the cap was chosen by running test simulations with caps starting at 15 kcal/mol and decreasing the cap whenever the above-mentioned numerical instabilities occurred, until a reasonably stable simulation was obtained. The resulting values for the cap (10 kcal/mol for AMBER and 8 kcal/mol for the CHARMM simulations) are below estimates of the activation energy for base flipping from previous experimental and computational studies.^{50,51} The biasing potentials were expanded in a 12th order trigonometric potential. In the AMBER simulations, 80 readjustments of the umbrella potential were performed, for a total production time of 40 ns per sequence, while the CHARMM simulations used 60 readjustments, or 30 ns of total production time per sequence. These values reflect the fact that the biasing cap was 2 kcal/mol higher for AMBER; hence, the AMBER simulations took longer to converge.

Analysis. Base pair step parameters were calculated using 3DNA,³⁹ and the helical axes and overall bending angles were obtained with Curves+.³⁵ The free energy surfaces below 3 kcal/mol were fitted to parabolas to determine the mean and standard deviations of the roll angle. Free energy surfaces as a function of the bending angle and the distribution of the charge density due to surrounding ions in the solution were determined as well. PLs were obtained in three ways: A_r , based on the dependence of end-to-end distance along the helical axis on contour length; A_w , based on the autocorrelation function of the tangent vector to the helical axis; and A_n , based on the autocorrelation function of the normal vector to the base pair planes. In addition, a PL A_r^* , which included a

contribution for the longitudinal stiffness of DNA, was obtained as well. These analyses were similar to Mazur's treatment,⁵² and the detailed procedures are provided in the Supporting Information. A simulation of a bent conformation of the DNA strand with central base pair CG under harmonic restraints on heavy atoms in a very large cubic water box ($120 \times 120 \times 120 \text{ \AA}$, with a 40 \AA water margin) at 150 mM KCl was conducted to test for size effects. Procedures for this simulation are also provided in the Supporting Information.

Nearest Neighbor Plate Models. The former four metrics of the PL (A_r , A_r^* , A_w , and A_n) are all based on direct extrapolations of the behavior of the simulated DNA strands. As a result, the metrics average over the behavior of the well-sampled central dimers as well as the peripheral DNA, which can affect the values of all four PL metrics if the sampling of the peripheral DNA is affected by end effects or has not fully converged. Additionally, two of the metrics (A_r and A_n) are particularly sensitive to static bending in the DNA due in part to large roll angles.⁵³ Last, extrapolation from the short length scales of the simulated DNA leads to large uncertainties in the predicted PLs. To circumvent these potential problems in estimating the average PL directly from short-strand simulations, the well-sampled local statistics of the central two base pairs were used to construct two sequence-independent coarse grained models. Henceforth called nearest neighbor plate (NNP) models, they are similar to the models developed by Mergell et al.⁵⁴ and Lu et al.⁵³ They are not intended as a replacement for other coarse-grained models of DNA^{54,55} but as an alternative way of examining the long length scale behavior from our short length simulation data. Although sequence effects are important in DNA flexibility both at the base pair step level and at the nearest neighbor level,^{24a–c} the simulations presented here did not provide enough statistics to make these models sequence dependent.

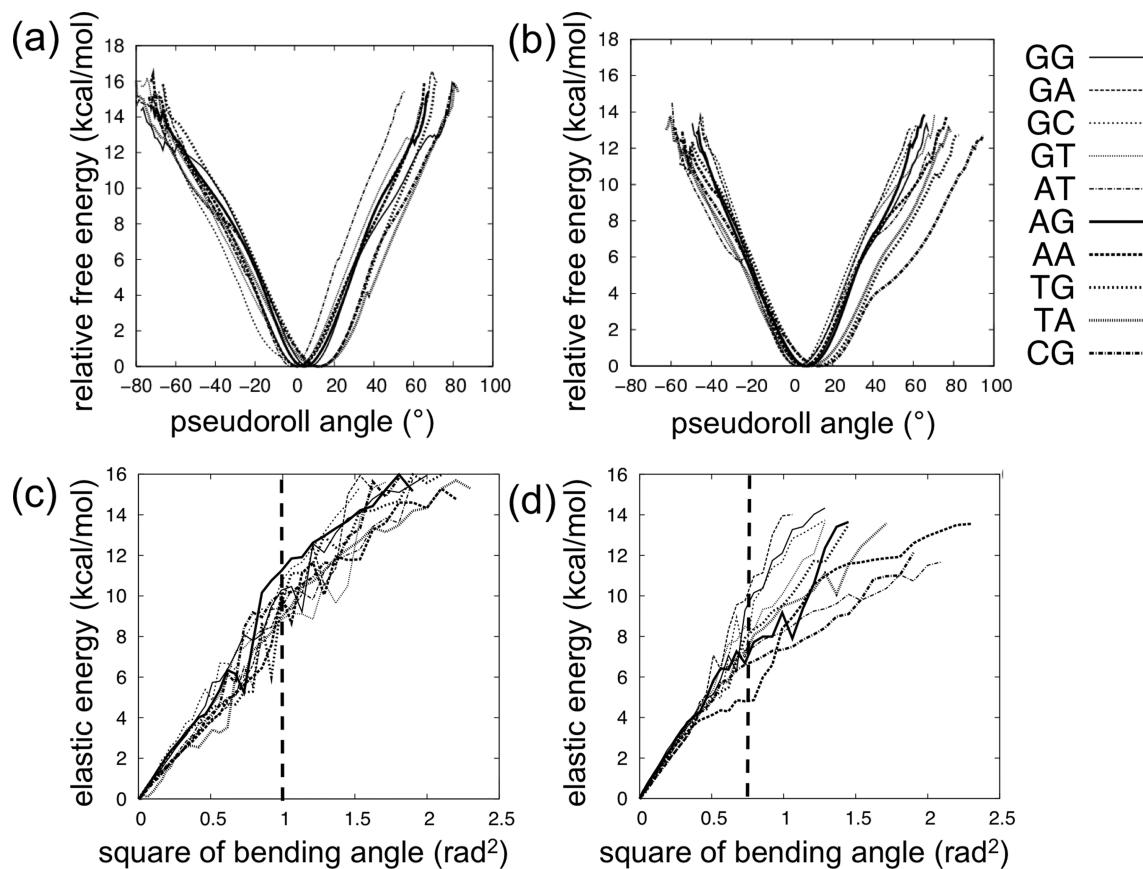


Figure 4. (a) Converged free energy surfaces as a function of the pseudoroll angle for the 10 different central basepair steps using the AMBER force field. (b) The CHARMM force field. (c) Elastic energy as a function of the square of the bending angle for AMBER simulations. (d) For CHARMM simulations. No convergence was obtained right of the dotted vertical lines in c and d (see text).

Therefore, we have constructed these models by averaging over all possible base pair steps as described below; consequently, results from the models will be approximate, since sequence dependence is not explicitly accounted for.

Both models assume that values of the roll (ρ), twist (ω), tilt (τ), shift (Dx), slide (Dy), and rise (Dz) between adjacent base pairs or “plates” ($i, i + 1$) are independent of second nearest neighbors ($i - 1, i + 2$). The models differ in the way these parameters are constructed. The first method, equivalent to linear response theory,⁵⁶ uses the fluctuations of the six dimer step parameters about their equilibrium values to determine a dimer-averaged covariance matrix Σ from the simulations (Table S2). As the mean values of the six dimer step parameters are dimer-dependent, it was necessary to subtract the means separately for each dimer prior to computing Σ . In constructing Σ , the data were reweighted to account for the biasing potential and dimer degeneracy (data from nonpalindromic sequences were counted twice). Σ along with the set of weighted averages for the individual motions (Table S3) fully describes a multivariate normal distribution, from which sets of correlated random numbers Z were generated according to

$$Z = \mu + Ax \quad (4)$$

Here, x is a vector of independent random numbers pulled from a standard normal distribution with mean μ and A is generated from Σ through the spectral decomposition:

$$A = U\Lambda^{1/2} \text{ where } \Sigma = U\Lambda U^T \quad (5)$$

Sets of vectors Z were used to assemble long strands of DNA one base pair at a time following a set of operations described in Lu et al.⁵³

The second NNP model was introduced to capture the quadratic relationships between roll and rise and roll and slide found in the simulations (Figure 3) and to include the observed anharmonicity in the free energy as a function of roll angle (Figures 4a and b). In this method, the rigid body motions were parametrized as functions of the roll angle. Similar to the first model, the distributions of roll angles from the simulations were averaged as a function of deviation from the mean for each dimer, and the data were reweighted to account for the biasing potential and dimer degeneracy. The quadratic dependencies of rise and slide on roll and the linear dependencies of the remaining motions were determined for each dimer using weighted least-squares fits (eq 6 and Figure S1). Upon subtracting out the fits to individual dimers, the residuals were used to determine a 5×5 covariance matrix Σ' (Table S4). As in the first method, the spectral decomposition of Σ' yielded the matrix A' used to generate the multivariate normal distribution z with the same covariance as the residuals. Adding these fluctuations to the fits resulted in base pair step parameters:

$$\begin{pmatrix} \omega \\ \tau \\ Dx \\ Dy \\ Dz \end{pmatrix} = \begin{pmatrix} -0.1405\rho + 35.31 \\ -0.0032\rho + 0.08 \\ -0.0024\rho - 0.03 \\ 0.000943\rho^2 - 0.0191\rho - 0.77 \\ 0.000467\rho^2 - 0.0001\rho + 3.48 \end{pmatrix} + A'z \quad (6)$$

Parameterization of both models was performed using AMBER data, since the AMBER simulations used a higher free energy cap than CHARMM and thus sampled more conformational space. While both models could be extended to model sequence-dependent properties, the statistics needed to construct such models would require more extensive MD simulations than are presented here. Each NNP model was used to construct 1000 48510-bp strands, corresponding to the length of λ bacteriophage DNA. An average PL for both models was obtained through the A_a metric by fitting a line to the exponential decay of the tangent–tangent correlation function.

RESULTS

We performed adaptive umbrella simulations of 10 different double-stranded DNA molecules using a modified definition of the roll angle of the central base pair step as a reaction coordinate. The sequences of the DNA molecules differ at the central base pair step. In these simulations, the DNA bends significantly in both directions at this base pair step, exploring roll angles between -70° and $+70^\circ$. This range covered all roll angles in our survey of protein–DNA crystal structures (Table S1 of the Supporting Information). For comparison, the most extreme negative and positive roll angles are -27° and $+24^\circ$ in the 147 base pair nucleosome,^{4b} and -54° and $+38^\circ$ in a 145 base pair nucleosome with stretched DNA.⁵⁷

Convergence. Convergence of the simulations was judged according to the absolute value of the maximum difference between the free energy surfaces on successive runs. Since the use of a cap on the biasing potential restricts the ability of the system to sample the region of the free energy surface above the cap, convergence cannot be expected for this part of the free energy surface and was excluded from the convergence analysis. Table S5 of the Supporting Information shows this difference for the last three runs in each of the simulations. For each system, the free energy converged to within 0.1 kcal/mol.

Tests of the Pseudoroll Angle Reaction Coordinate. An essential test for the pseudoroll angle is how well it reproduced the actual roll angle as calculated by 3DNA. The pseudoroll angle was found to be strongly correlated with the actual roll angle for all base pair steps, as shown in Figure S2 of the Supporting Information. The correlation coefficient between the pseudoroll angle and actual roll angle was found to be 0.997 for the crystal structures in the survey, 0.998 for simulations conducted using the AMBER force field, and 0.997 for simulations conducted using the CHARMM force field. The mean difference between the two was found to be less than 1° , and RMSDs were found to be $1\text{--}2^\circ$. Detailed statistics further quantifying the excellent agreement between the pseudoroll and true roll angle are shown in Table S6 in the Supporting Information.

The DNA strands maintained stable base pairing throughout these simulations, although in some of the simulations the bases flipped out partially for approximately 50–100 ps before returning to a Watson–Crick base paired configuration. Toward the end of the CHARMM simulation for central

base pair step CG, the central base pairs buckled outward as shown in Figure S3 of the Supporting Information. This caused a lower free energy for large positive roll angles in the corresponding free energy surface. For all strands, there is a V-shaped correlation between the pseudoroll angle of the central base pair step and the overall bending angle, as shown by the scatterplots in Figure S4 of the Supporting Information. The correlation is V-shaped because both positive and negative roll angles reflect bending of the DNA that results in a positive bending angle. There is considerable scatter in this correlation due to conformational fluctuations in the two ends of the DNA, as well as rotation of the two pairs of the central step relative to the DNA helical axis. Despite the scatter, the correlation confirms the observation that most of the DNA bending flexibility is due to changes in the roll angle.³⁰

Short Length Behavior Shows Increased Flexibility at Large Roll Angles. A comparison of the mean and standard deviation of the roll angles for each of the 10 base pair steps in our simulations with those from previous simulations^{24c} and crystal structure surveys⁵⁶ is shown in Table S7 in the Supporting Information. The pyrimidine–purine (RY) base pair steps have a positive equilibrium roll angle of about 10° and an increased flexibility due to decreased base pair stacking, which is consistent with the molecular dynamics simulations of the ABC consortium^{24c} and surveys of crystallographically determined DNA structures.⁵⁶ The free energy surfaces as a function of the pseudoroll angle are shown in Figure 4a and b; a close-up view for near zero values of the roll angle is shown in Figure S5 of the Supporting Information. The AMBER and CHARMM force fields gave similar results for values of the roll angle near zero but began to differ for large positive or negative roll angles. The AMBER simulations were able to sample larger deviations of the roll angle because a higher cap on the biasing potential was used. Whereas the free energy surface is quadratic for small and intermediate roll angles, for large roll angles, the free energy surface is linear in roll angle, indicating a relative reduction in the free energy cost of DNA deformation at large roll angles. We observed this linear dependence at high roll angles for all strands and both force fields.

The elastic energy was calculated from the free energy in terms of a global bending angle as described in the Supporting Information and is plotted against the square of the bending angle in Figure 4c and d. Although the elastic energy is usually thought of as a potential energy in the context of coarse-grained models such as the WLC model, it contains entropic contributions from the motions of atoms within each base pair and therefore is a free energy in the context of fully atomistic models of DNA. Consistent with the WLC model, all simulations showed a quadratic dependence of the elastic energy on θ for $\theta < 50^\circ$. Despite sampling a wide range of roll angles and obtaining good convergence of the free energy surface as a function of the roll angle, we were unable to converge the free energy surface as a function of the DNA bending angle for high bending angles ($>60^\circ$ for AMBER and $>50^\circ$ for CHARMM). The reason for this is two-fold: we did not directly bias the bending angle, and the use of the biasing cap prevented convergence at very large bending angles (because these occur at large free energies). Due to the biasing cap, the convergence cannot be easily remedied by longer sampling. For small and intermediate bending angles, the free energy surfaces in terms of bending angle appear to be similar in shape to those determined by Zacharias using screw axis bending as a reaction coordinate, and the minimum of the free

Table 1. Persistence Length of the DNA Strands, Calculated by Four Different Methods^a

force field	AMBER				CHARMM			
	base pair step	A_r (Å)	A_r^* (Å)	A_a (Å)	A_n (Å)	A_r (Å)	A_r^* (Å)	A_a (Å)
GG	568.2	489.0	568.2	422.3	639.8	599.8	675.0	377.6
GA	479.9	419.6	431.9	471.3	652.6	616.4	646.6	546.9
GC	680.7	654.8	667.7	448.6	620.0	610.9	704.5	385.9
GT	611.6	538.6	586.2	465.3	622.9	593.8	674.7	451.4
AT	508.0	422.8	417.5	730.5	620.8	593.0	632.4	547.4
AG	576.7	520.6	565.4	582.6	579.6	548.8	607.2	425.0
AA	494.0	429.9	453.3	650.1	531.0	523.4	589.9	540.1
TG	642.8	564.6	599.9	370.2	576.3	567.8	599.7	340.1
TA	433.6	405.4	370.0	498.2	601.7	572.6	620.4	475.4
CG	487.5	464.3	465.8	299.6	522.3	545.9	606.8	333.8

^a A_r is determined by fitting the dependence of tangent vector autocorrelation function on contour length as specified in eq S4. A_a is determined by fitting the autocorrelation function of normal vectors to the bases as specified in eq S5. A_n is determined by fitting the dependence of end-to-end distances along the Curves+ helical axis on contour length using eq S6, while A_r^* includes a contribution for the longitudinal stiffness in this fit (eq S7).

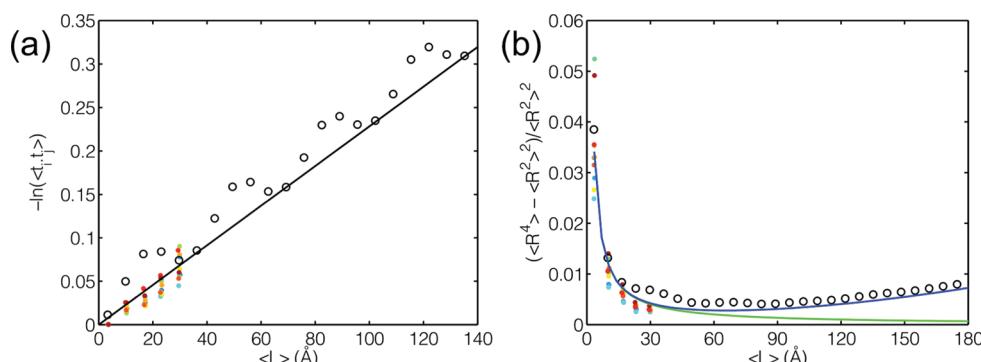


Figure 5. (a) Behavior of the alignment of tangent vectors ($t_i \cdot t_j$) as a function of DNA contour length. Results from the NNP model (open circles) are compared with results directly from MD (dots, colored from blue to red according to the order of the base pairs in Table 1). The black curve is the best fit line to the NNP model, used to extract a PL of 438 Å via eq S4. (b) Relative variance in the square of the end-to-end separation, R^2 , as a function of DNA contour length. Results from the NNP model (open circles) are compared with results directly from MD (dots, colored as in a). The green and blue curves are the relative variance expected from longitudinal and the sum of longitudinal and bending fluctuations, respectively.

energy surface is in a similar position, at about 10°.²⁸ Despite this, it is clear from the data that our free energy cost of bending for large angles is higher than obtained by Zacharias (e.g., Zacharias found a value of 5.5 kcal/mol for bending through a 90° angle), which confirms that the presence of type II kinks lowers the free energy cost of bending.

Long Length Behavior Is in Agreement with WLC Model. The PL of each DNA fragment can be determined by examining the dependence of structural parameters on contour length, as described in the Supporting Information and shown in Figure S6. This represents an extrapolation of the short length scales of the simulations to much larger distances, which can magnify even small errors. As a consequence, numerical values of the PLs depended on the method of analysis. The equations yielded close fits, however, which increased our confidence that the behavior of DNA on larger length scales is similar to what was observed in our simulations. The fitted PLs are shown in Table 1 and Figure S7 of the Supporting Information and are in qualitative agreement with the accepted value of approximately 500 Å.^{7a} The calculated PLs are generally slightly longer for the CHARMM simulations. The AMBER values vary over a significant range, with the PLs for central base pair steps including A–T being particularly low. This may be because these base pair steps combine with the neighboring A–T base pairs to form an A–T rich region, and

consistent with the experimental observations,²² these regions are more flexible than the average DNA sequence. PLs based on the contour length (eqs S6 and S7, Figure S6 of the Supporting Information) were improved when accounting for DNA stretching, with values of A_r^* shorter and closer to 500 Å than values of A_r . The DNA stretching moduli obtained from the A_r^* fits correspond to stretching moduli in the range of 400–2800 pN, which is the same order of magnitude as the experimentally measured values (approximately 1000 pN).⁵⁸

In addition, NNP models were used to determine a persistence length from the statistics of the well-sampled central base pairs. The two models differ in the way the interaction of neighboring base pairs was parametrized. The first NNP model assumed linear and harmonic relationships between all base pair step parameters, while the second incorporated roll angle anharmonicity as well as the non-linearity observed between rise-roll and slide-roll step parameters. The latter method describes all base pair motions as functions of the roll angle, with the resulting fits shown in Figure S1 of the Supporting Information. The U-shaped relationships of rise and slide to the roll angle are likely due to geometric requirements, as the two base pairs must separate if they are not to collide with each other when the roll angle becomes large in either direction. The negative correlation between roll and twist is in qualitative agreement with trends

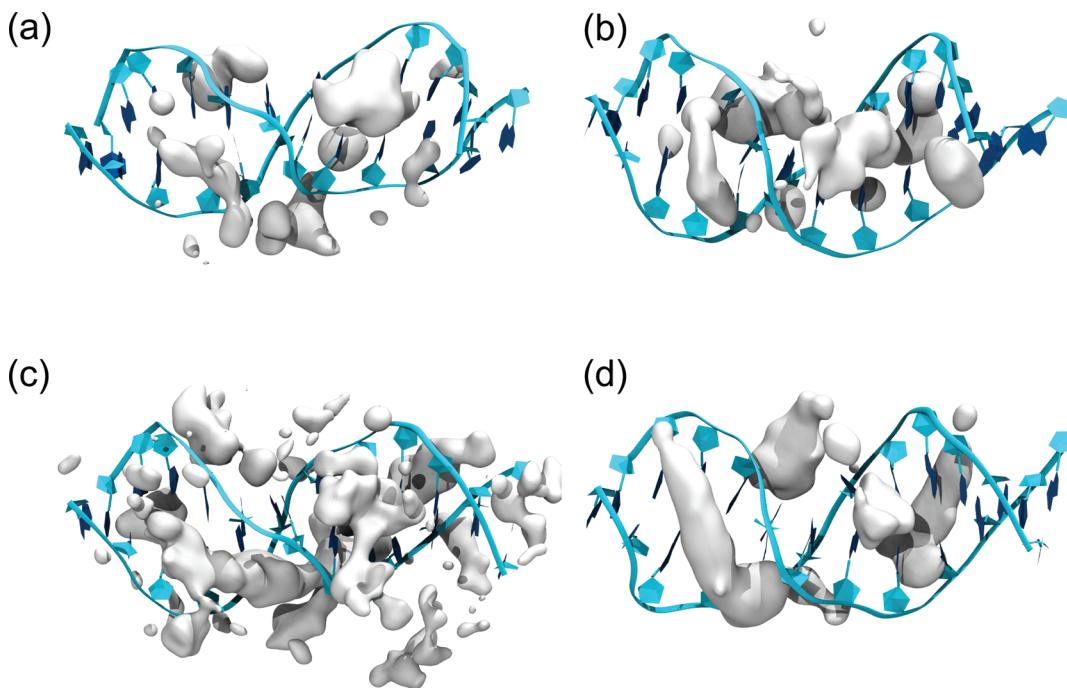


Figure 6. Isosurface of charge density at $0.0015e/\text{\AA}^3$ from ions around DNA from CG AMBER simulation (a) for frames with a pseudoroll angle less than -30° ; (b) for frames with a pseudoroll angle greater than 30° ; (c and d) similar isosurfaces from CG CHARMM simulation for frames with pseudoroll angles less than -30° or greater than 30° , respectively.

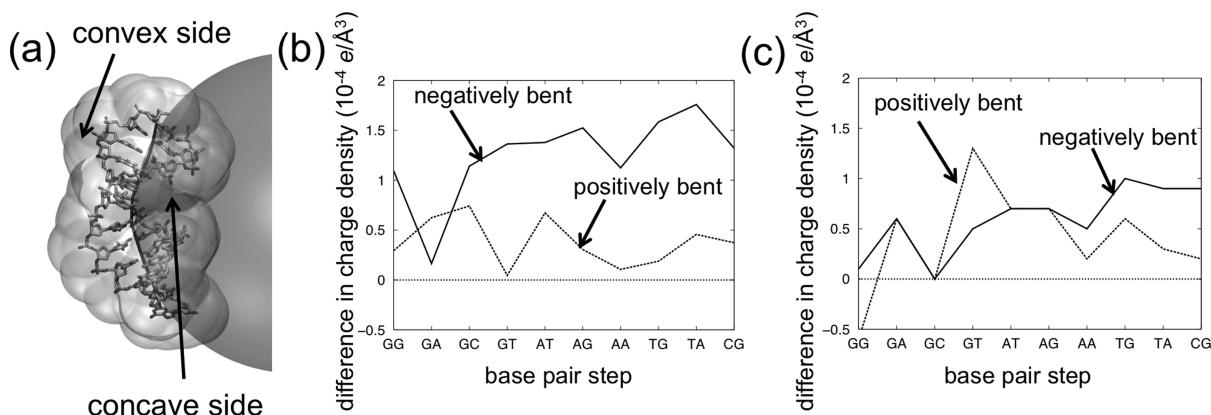


Figure 7. (a) Volumes within which average ion densities were calculated. The volume considered was within 5\AA of the DNA (white surface) and was divided into concave and convex sides of the DNA based on a sphere fitted to the helical axis generated by Curves+ (gray sphere, black curve). (b) Difference between charge density due to ions in the concave and convex sides as a function of base pair step and DNA bending for simulations using the AMBER force field. (c) Difference between charge density due to ions in the concave and convex sides as a function of base pair step and DNA bending for simulations using the CHARMM force field.

observed in DNA crystal structures.⁵⁶ The average PL was obtained from the slope of the logarithm of the tangent–tangent correlation function $C(L) = \langle \mathbf{t}_i \cdot \mathbf{t}_{i+1} \rangle$ computed from 1000 strands of 48 510-bp DNA (the length of λ bacteriophage DNA), as shown in Figure 5a. Fitting was performed to the bottoms of the oscillations so as not to be affected by periodic deviations of the tangent vector about the helical axis due to the helicity of DNA.⁵² Using data from the AMBER simulations, the linear response⁵⁶ and anharmonic methods yielded PL values of $437 \pm 1\text{\AA}$ and $438 \pm 1\text{\AA}$, respectively, identical within error. Their similarity is the result of the PL being largely determined by the variance in the roll and tilt angles common to the two models, as anharmonic terms become irrelevant on length scales many times larger than the base pair separation.⁵⁹ Both PL values are shorter than the experimentally measured

values of around 500\AA ^{12b} but close enough to suggest that our simulations are consistent with the experimental long length scale behavior of DNA.

Because of the relative shortness of the simulated strands, the MD data lack the oscillations seen in longer DNA strands⁵² (Figure 5a). The oscillations are also suppressed by the smoothing effect caused by fitting a tangent curve (possessing inherent stiffness) to the DNA segment by Curves+. Further complications of inferring PLs from short strands arise from the significant role of longitudinal fluctuations on these short length scales, as shown by the difference in the PL values obtained by the A_r and A_r^* metrics. Their importance can also be seen by decomposing the fluctuations in the end-to-end distances into stretching and bending components (Figure 5b). Only for separations above 100\AA , or 30 base pairs, do bending

contributions to the fluctuations take over. The two NNP models display the correct limiting behavior at both short and long length scales, while fluctuations at intermediate separations between 20 and 80 Å cannot be accounted for solely by the sum of the limiting stretching and bending behaviors.

Ionic Screening Aids Bending Flexibility at High Roll Angles.

The distributions of ionic charge around the DNA show that when the pseudoroll angle is positive, potassium ions preferentially accumulate in the major groove (Figure 6b and d). When the pseudoroll angle is negative, potassium ions accumulate near the minor groove (Figure 6a and c). In either case, positive charge accumulates on the concave side of the DNA, reaching a very large charge density at extreme roll angles. The isosurface shown represents a charge density of $0.0015e/\text{\AA}^3$, which is equivalent to a difference between potassium and chloride ion concentrations of about 2.5 M, about 16 times the bulk salt concentration of $\sim 150\text{ mM}$ used in this study. The same tendency is observed in the CHARMM and AMBER simulations, although in the CHARMM simulations the potassium ions occupy the minor groove regardless of the conformation of the DNA. While significant patterns were obtained for the distribution of ions, the number of water molecules in the first solvation shell showed no changes with pseudoroll angle, suggesting that the distribution of water molecules is little changed by the bending of DNA.

To further quantify this difference, the ion densities within 5 Å of the DNA and on either the concave or convex side were calculated. The results are shown in Table S8 and Figure 7. Figure 7a illustrates the relevant volumes within which the ion densities were estimated. Table S8 shows the calculated average charge densities due to the ions for DNA bent either negatively (pseudoroll angle less than -30°) or positively (pseudoroll angle greater than 30°). Figure 7b and c show the difference between the charge density on concave and convex sides as a function of base pair step and force field. The overall trends are consistent across all base pairs and force fields and show a congregation of positive charges on the concave side of DNA when DNA is strongly bent. The positive charge density is higher on the concave side of the DNA in all 20 cases when the DNA is bent negatively and in 19 out of 20 cases when the DNA is bent positively.

The magnitude of the difference in charge density varied between AMBER and CHARMM, however. This may be because it depends on the details of interactions between potassium ions and the various base atoms of DNA in the different force fields. In particular, the positive charge density on the concave side of the DNA is higher in CHARMM than in AMBER. This likely reflects a tendency of potassium ions to associate with the minor groove of DNA in the CHARMM force field. This may in turn be because of differences in the way that partial charges are parametrized in the two force fields.

Tests showed that the observed charge congregation was not due to convergence or system size issues. Analyses showed that the ion distributions were equilibrated, and the difference in charge density between concave and convex sides of bent DNA was converged after 20 ns. The 12 Å water margin used in the simulations is only slightly longer than the Debye–Hückel length for a KCl solution of 150 mM concentration ($\sim 7.9\text{ \AA}$), raising the possibility that the observed congregation of positive charges may be an artifact from the low number of ions simulated or the small system size. To test for this possibility, an additional simulation was conducted with a positively bent DNA structure under restraints in an extremely large solvent

box ($120 \times 120 \times 120\text{ \AA}$) with 334 ions as described in the Supporting Information. This simulation gave a very similar accumulation of positive charge density in the minor groove as observed in the smaller simulations.

The preference of cations for the concave side of DNA would be expected to increase its flexibility by screening the repulsion of the backbones. These observations suggest an explanation for the observed decrease in DNA stiffness upon bending through high roll angles. As DNA bends, cations accumulate on the concave side of the DNA, screening the interaction between phosphate backbones and making it easier to further bend the DNA.

DISCUSSION

By performing adaptive umbrella sampling on the roll angles of DNA strands, we have simulated the bending of 12 bp fragments of DNA in atomistic detail. The results are consistent with the known long-range behavior and PL of DNA, but the free energy surfaces as a function of roll angle show evidence that short DNA strands become less stiff when strongly bent. This was observed for all sequences and both force fields (AMBER and CHARMM). Such increased flexibility on the short length scale does not imperil agreement with the WLC model for long length scale behavior, as demonstrated by our calculations using coarse-grained models. The free energy cost of large roll angles is high enough, and the Boltzmann probability of their occurrence small enough, that the behavior of long DNAs is not affected by a possible increase of flexibility at high roll angles. A similar observation is made in the recently proposed “subelastic chain” models⁵⁹ in which the bending free energy is linear in the bending angles. Despite an increased flexibility on short length scales, these models give the same behavior as the WLC model on long length scales, in analogy to the central limit theorem.

Both the short-range PLs obtained directly from the simulations and those from the coarse-grained sampling are in qualitative agreement with the accepted PL of DNA.^{7a} They do not agree perfectly with experimentally obtained values, however, such as the recent measurements of sequence-dependent PLs for short DNA strands from cyclization experiments.²² There are several possible reasons for this. The sequences simulated do not exactly match those for which PLs have been determined experimentally. In particular, the inherent flexibility of base pair steps is known to be affected by the flanking base pairs,^{24a–c} so that the inherent flexibility of each base pair dimer in experimental sequences may be different from that simulated here. It is also difficult to represent the conformation of a complex polymer such as DNA in terms of a simple curve. This makes it difficult to describe DNA flexibility on short length scales in terms of a single PL. It is not trivial to determine a local helical axis for short fragments of DNA; for example, in the Mazur analysis, an adjustable parameter was used to control the bending elasticity of the axis in order to remove artifacts due to the helicity of DNA.⁵² In addition, the calculated PL of DNA differs depending on what structural features of the DNA are used, because of the ability of the base pairs to rotate relative to the surrounding helical axis. Nevertheless, while no absolute quantitative agreement is reached, the calculated PLs from the simulations are in good overall agreement with the experimental values, and the simulations reflect that on the long length scale, DNA is well-described by the WLC model.

In this work, only type I kinks were studied. Localized melting of the DNA duplex (involving type II kinks) and formation of noncanonical structures have been proposed to explain an increase in the inherent flexibility of DNA on short length scales,^{15a} and recent evidence for this has been obtained by NMR relaxation dispersion measurements.¹⁸ Since this localized melting produces numerical instabilities in the simulation approach presented here, our simulations could not evaluate the contribution of type II kinks to DNA flexibility. Consequently, our simulations provide a conservative estimate of the flexibility of DNA, and our estimates of the free energy cost of bending DNA are higher than those produced using the screw-axis coordinate of Zacharias.^{26,28} Nevertheless, we confirm the linear dependence of free energy on roll angle for large roll angles. This increased DNA flexibility would not require type II kinks or significant disruptions in base pairing and would come exclusively from type I kinks.

Due to the use of adaptive umbrella sampling on the roll angle, the simulations presented here produced DNA conformations with a much wider variety of relative base pair positions and orientations compared to those observed in previous simulations. Consequently, it is necessary to consider whether the force fields used here accurately represent the interactions between base pairs for these positions and orientations. In the force fields, this interaction is represented through carefully fitted van der Waals and electrostatic parameters. Detailed quantum chemical studies suggest that the mathematical forms of these terms are adequate, and no additional terms such as special π charges are needed.⁶⁰ More recent comparisons of interaction energies for DNA base stacking complexes from the AMBER force field to high-level quantum calculations indicate strong agreement both at equilibrium distances and at longer distances.⁶¹ In addition, a previous simulation performed in our group correctly predicted experimental trends in the activation energy of unstacking of the fluorescent dye Cy3 from DNA, suggesting that stacking forces are also correctly reproduced by the CHARMM force fields.⁶²

Our study of the distribution of ions around the DNA has also led us to propose a mechanism for the increase in DNA flexibility on short length scales with strong bending. When DNA bends strongly, cations congregate on the concave side of the DNA, reaching a concentration about an order of magnitude greater than their concentration in the bulk solution. This reduces the electrostatic repulsion between phosphate backbones, facilitating the bending. To our knowledge, this is the first molecular mechanism proposed for increased DNA flexibility on short length scales that does not involve the partial melting of the DNA. The finding that positive charges aid the flexibility of DNA is in agreement with several experimental observations. The flexibility of DNA was found to increase when it is in contact with positively charged surfaces,^{21a} when chemically modified by tethering a cationic group to the major groove,^{21b} and when one backbone was artificially neutralized.^{21d} The latter effect had also been reproduced in simulations.^{21c} In addition, bending of DNA in the nucleosome core particle is aided by the presence of arginine side chains extending into the minor groove.^{4b} Increased bending caused by asymmetric neutralization of phosphates on the concave side has also been found to provide 1–2 kcal/mol to the binding of EcoRV.⁶³

The ability of cations to congregate in the minor groove has been observed previously in simulations and is a result of

coordination of the ions by carbonyl and amine groups and ring nitrogens of the bases. For example, a small region of high sodium ion density was observed in microsecond-length simulations of DNA,⁶⁴ and a 3D-RISM study of ion distributions around DNA also showed the substantial presence of potassium ions in the minor groove.⁶⁵ A coarse grained simulation of DNA with explicit ions showed that the PL dropped by about 25% when the salt concentration was increased from 0.1 to 1 M.⁶⁶ Although these studies mainly considered straight DNA, similar effects are to be expected in bent DNA, especially since the bending of DNA brings coordinating groups on the bases closer together on the concave side.

While the preference of cations for the concave side of the DNA presents a possible mechanism for the increase in DNA flexibility upon strong bending, the simulations do not rule out other possible contributions. For example, the decrease in stacking energies upon bending might also play a role. An attempt was made to assess the possible role of stacking energies from our simulations, but the data were too noisy to draw any useful conclusions. We will examine the possible role of stacking energies in future work.

ASSOCIATED CONTENT

S Supporting Information

Details on the crystal structure survey, simulations, and analyses are provided. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 813-974-8762. Fax: 813-974-3203. E-mail: avandervaart@usf.edu.

Author Contributions

[#]These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Daniel Barr, Andrew Chizmeysha, Dmitry Matyushov, and Petra Fromme for helpful discussions. We thank Jeff Klauda for providing parameter files for the AMBER parambsc0 force field in CHARMM format. We thank the ASU Fulton HPCI, USF Research Computing, TeraGrid, and XSEDE for computer time. This work was supported by NSF CAREER award no. CHE-1007816 to A.v.d.V., and NSF DMS-0714953 to M.F.T.

REFERENCES

- (1) Sarai, A.; Kono, H. Protein-DNA recognition patterns and predictions. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 379–398.
- (2) Allemann, R. K.; Egli, M. DNA recognition and bending. *Chem. Biol.* **1997**, *4*, 643–650.
- (3) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **2000**, *8*, 1015–1023.
- (4) (a) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *J. Mol. Biol.* **2002**, *319*, 1097–1113. (b) Richmond, T. J.; Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **2003**, *423*, 145–150.
- (5) Strahl, B. D.; Allis, C. D. The language of covalent histone modifications. *Nature* **2000**, *403*, 41–45.

- (6) Kratky, O.; Porod, G. Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 1106–1123.
- (7) (a) Taylor, W. H.; Hagerman, P. J. Application of the method of phage-T4 DNA ligase-catalyzed ring closure to the study of DNA structure. 2. NaCl dependence of DNA flexibility and helical repeat. *J. Mol. Biol.* **1990**, *212*, 363–376. (b) Bustamante, C.; Smith, S.; Liphardt, J.; Smith, D. Single molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279–285. (c) Bustamante, C.; Bryant, Z.; Smith, S. Ten years of tension: single-molecule DNA mechanics. *Nature* **2003**, *421*, 423–427.
- (8) Cloutier, T. E.; Widom, J. Spontaneous sharp bending of double-stranded DNA. *Mol. Cell* **2004**, *14*, 355–352.
- (9) Wiggins, P. A.; van der Heijden, T.; Moreno-Herrero, F.; Spakowitz, A.; Phillips, R.; Widom, J.; Dekker, C.; Nelson, P. C. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnol.* **2006**, *1*, 137–141.
- (10) Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. B. Remeasuring the Double Helix. *Science* **2008**, *322*, 446–449.
- (11) Yuan, C. L.; Chen, H. M.; Lou, X. W.; Archer, L. A. DNA bending stiffness on small length scales. *Phys. Rev. Lett.* **2008**, *100*, 018102.
- (12) (a) Du, Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5397–5402. (b) Vologodskaia, M.; Vologodskii, A. Contribution of the intrinsic curvature to measured DNA persistence length. *J. Mol. Biol.* **2002**, *317*, 205–213. (c) Mastrianni, A. J.; Sivak, D. A.; Geisser, P. L.; Alivisatos, A. P. Probing the conformational distributions of subersistence length DNA. *Biophys. J.* **2009**, *97*, 1408–1417. (d) Witz, G.; Rechendorff, K.; Adamcik, J.; Dietler, G. Conformation of circular DNA in two dimensions. *Phys. Rev. Lett.* **2008**, *101*, 148103.
- (13) Crick, F. H. C.; Klug, A. Kinky helix. *Nature* **1975**, *255*, 530–533.
- (14) Sobell, H. M.; Tsai, C. C.; Gilbert, S. G.; Jain, S. C.; Sakore, T. D. Organization of DNA in chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 3068–3072.
- (15) (a) Yan, J.; Marko, J. F. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.* **2004**, *93*, 108108. (b) Travers, A. DNA dynamics: Bubble 'n' flip for DNA cyclisation? *Curr. Biol.* **2005**, *15*, R377–R379.
- (16) Yuan, C.; Rhoades, E.; Lou, X. W.; Archer, L. A. Spontaneous sharp bending of DNA: role of melting bubbles. *Nucleic Acids Res.* **2006**, *34*, 4554–4560.
- (17) Forties, R. A.; Bundschuh, R.; Poirier, M. G. The flexibility of locally melted DNA. *Nucleic Acids Res.* **2009**, *37*, 4580–4586.
- (18) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **2011**, *470*, 498–502.
- (19) Peters, J. P.; Maher, L. J. DNA curvature and flexibility *in vitro* and *in vivo*. *Q. Rev. Biophys.* **2010**, *44*, 23–63.
- (20) Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic solutions of polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179–246.
- (21) (a) Podesta, A.; Indrieri, M.; Brogioli, D.; Manning, G. S.; Milani, P.; Guerra, R.; Finzi, L.; Dunlap, D. Positively charged surfaces increase the flexibility of DNA. *Biophys. J.* **2005**, *89*, 2558–2563. (b) Moulaei, T.; Maehigashi, T.; Lountos, G. T.; Komeda, S.; Watkins, D.; Stone, M. P.; Marky, L. A.; Li, J. S.; Gold, B.; Williams, L. D. Structure of B-DNA with cations tethered in the major groove. *Biochemistry* **2005**, *44*, 7458–7468. (c) Kosikov, K. M.; Gorin, A. A.; Lu, X. J.; Olson, W. K.; Manning, G. S. Bending of DNA by asymmetric charge neutralization: All-atom energy simulations. *J. Am. Chem. Soc.* **2002**, *124*, 4838–4847. (d) Okonogi, T. M.; Alley, S. C.; Harwood, E. A.; Hopkins, P. B.; Robinson, B. H. Phosphate Backbone Neutralization Increases Duplex DNA Flexibility: A Model for Protein Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4156–4160.
- (22) Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15421–15426.
- (23) Mills, J. B.; Hagerman, P. J. Origin of the intrinsic rigidity of DNA. *Nucleic Acids Res.* **2004**, *32*, 4055–4059.
- (24) (a) 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. Molecular dynamics simulations for the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(C_pG) steps. *Biophys. J.* **2004**, *87*, 3799–3813. (b) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., III; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **2005**, *89*, 3721–3740. (c) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jarayam, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299–313. (d) Lankas, F.; Spackova, N.; Moakher, M.; Enkhbayar, P.; Sponer, J. A measure of bending in nucleic acids structures applied to A-tract DNA. *Nucleic Acids Res.* **2010**, *38*, 3414–3422.
- (25) Lankas, F.; Lavery, R.; Maddocks, J. H. Kinking occurs during molecular dynamics simulation of small DNA minicircles. *Structure* **2006**, *14*, 1527–1534.
- (26) Curuksu, J.; Zakrzewska, K.; Zacharias, M. Magnitude and direction of DNA bending induced by screw-axis orientation: influence of sequence mismatches and abasic sites. *Nucleic Acids Res.* **2008**, *36*, 2268–2283.
- (27) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817–3829.
- (28) Curuksu, J.; Zacharias, M.; Lavery, R.; Zakrzewska, K. Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res.* **2009**, *37*, 3766–3773.
- (29) Savalyev, A.; Materese, C. K.; Papoian, G. A. Is DNA's Rigidity Dominated by Electrostatic or Nonelectrostatic Interactions? *J. Am. Chem. Soc.* **2011**, *133*, 19290–19293.
- (30) Dickerson, R. E. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **1998**, *26*, 1906–1926.
- (31) (a) Hooft, R. W. W.; van Eijck, B. P.; Kroon, J. An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J. Chem. Phys.* **1992**, *97*, 6690–6694. (b) Bartels, C.; Karplus, M. Multidimensional Adaptive Umbrella Sampling: Applications to Main Chain and Side Chain Peptide Conformations. *J. Comput. Chem.* **1997**, *18*, 1450–1462.
- (32) (a) Foloppe, N.; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104. (b) MacKerell, A. D., Jr.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105–120.
- (33) El Hassan, M. A.; Calladine, C. R. The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *J. Mol. Biol.* **1995**, *251*, 648–664.
- (34) Lavery, R.; Sklenar, H. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* **1989**, *6*, 655–667.
- (35) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkovicu, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.
- (36) Bhattacharya, D.; Bansal, M. A self-consistent formulation for analysis and generation of non-uniform DNA structures. *J. Biomol. Struct. Dyn.* **1989**, *6*, 635–653.

- (37) Babcock, M. S.; Pendault, E. P. D.; Olson, W. K. Nucleic acid structure analysis: Mathematics for local cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.* **1994**, *237*, 125–156.
- (38) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. B-DNA twisting correlates with base pair morphology. *J. Mol. Biol.* **1995**, *247*, 34–48.
- (39) Lu, X.-J.; Olson, W. K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (40) Lu, X.-J.; Olson, W. K. Resolving the discrepancies among nucleic acid conformational analyses. *J. Mol. Biol.* **1999**, *285*, 1563–1575.
- (41) Spiriti, J.; van der Vaart, A. Mechanism of the Calcium-Induced trans-cis Isomerization of a Non-Prolyl Peptide Bond in Clostridium histolyticum Collagenase. *Biochem.* **2010**, *49*, 5314–5320.
- (42) Spiriti, J.; Bogani, F.; Ghirlanda, G.; van der Vaart, A. Modulation of protein stability by O-glycosylation in a designed Gc-MAF analog. *Biophys. Chem.* **2008**, *134*, 157–167.
- (43) 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. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (44) Kottalam, J.; Case, D. A. Dynamics of Ligand Escape from the Heme Pocket of Myoglobin. *J. Am. Chem. Soc.* **1988**, *110*, 7690–7697.
- (45) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (46) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science* **1996**, *271*, 997–999.
- (47) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (48) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (49) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte-Carlo data-analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
- (50) Coman, D.; Russu, I. M. A nuclear magnetic resonance investigation of the energetics of basepair opening pathways in DNA. *Biophys. J.* **2005**, *89*, 3285–3292.
- (51) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *J. Chem. Theory Comput.* **2009**, *5*, 3105–3113.
- (52) Mazur, A. K. Evaluation of Elastic Properties of Atomistic DNA Models. *Biophys. J.* **2006**, *91*, 4507–4518.
- (53) Lu, X.-J.; El Hassan, M. A.; Hunter, C. A. Structure and conformation of helical nucleic acids: analysis program (SCHNAAp). *J. Mol. Biol.* **1997**, *273*, 668–680.
- (54) Mergell, B.; Ejtehadi, M.; Everaers, R. Modeling DNA structure, elasticity, and deformations at the base-pair level. *Phys. Rev. E* **2003**, *68*, 021911.
- (55) (a) Dans, P. D.; Zeida, A.; Pantano, S.; Coarse Grained, A Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725. (b) Savelyev, A.; Papoian, G. A. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20340–20345. (c) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. DNA Nanotweezers Studied with a Coarse-Grained Model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101. (d) Olson, W. K.; Colasanti, A. V.; Czapla, L.; Zheng, G., Insights into the Sequence-Dependent Macromolecular Properties of DNA from Base-Pair Level Modeling. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; pp 205–224. (e) Tan, R. K.-Z.; Petrov, A. S.; Devkota, B.; Harvey, S. C., Coarse-Grained Models for Nucleic Acids and Large Nucleoprotein Assemblies. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009; pp 225–236. (f) Knotts, T. A.; Schwartz, D. C.; de Pablo, J. J. A coarse grain model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901. (g) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: A coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy. *Proteins* **2010**, *78*, 1266–1281. (h) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.* **2010**, *132*, 035105.
- (56) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.
- (57) Ong, M. S.; Richmond, T. J.; Davey, C. A. DNA stretching and extreme kinking in the nucleosome core. *J. Mol. Biol.* **2007**, *368*, 1067–1074.
- (58) Wang, M. D.; Yin, H.; Landick, R.; Gelles, J.; Block, S. M. Stretching DNA with optical tweezers. *Biophys. J.* **1997**, *72*, 1335–1346.
- (59) Wiggins, P. A.; Nelson, P. C. Generalized theory of semiflexible polymers. *Phys. Rev. E* **2006**, *73*, 31906.
- (60) (a) Sponer, J.; Leszczynski, J.; Hobza, P. Nature of nucleic acid-base stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs. *J. Phys. Chem.* **1996**, *100*, 5590–5596. (b) Sponer, J.; Leszczynski, J.; Hobza, P. Base stacking in cytosine dimer. A comparison of correlated ab initio calculations with three empirical potential models and density functional theory calculations. *J. Comput. Chem.* **1996**, *17*, 841–850.
- (61) (a) Sponer, J.; Jurecka, P.; Marchan, I.; Luque, J.; Orozco, M.; Hobza, P. Nature of base stacking: Reference quantum-chemical stacking energies in ten unique B-DNA base-pair steps. *Chem.—Eur. J.* **2006**, *12*, 2854–2865. (b) Kolář, M.; Berka, K.; Jurečka, P.; Hobza, P. On the Reliability of the AMBER Force Field and its Empirical Dispersion Contribution for the Description of Noncovalent Complexes. *ChemPhysChem* **2010**, *11*, 2399–2408.
- (62) Spiriti, J.; Binder, J. K.; Levitus, M.; van der Vaart, A. Cy3-DNA stacking interactions strongly depend on the identity of the terminal base pair. *Biophys. J.* **2011**, *100*, 1049–1057.
- (63) Hancock, S. P.; Hiller, D. A.; Perona, J. J.; Jen-Jacobson, L. The Energetic Contribution of Induced Electrostatic Asymmetry to DNA Bending by a Site-Specific Protein. *J. Mol. Biol.* **2011**, *406*, 285–312.
- (64) Perez, A.; Luque, F. J.; Orozco, M. Dynamics of B-DNA on the Microsecond Time Scale. *J. Am. Chem. Soc.* **2007**, *129*, 14739–14745.
- (65) Howard, J. J.; Lynch, G. C.; Pettitt, B. M. Ion and Solvent Density Distributions around Canonical DNA from Integral Equations. *J. Phys. Chem. B* **2011**, *115*, 547–556.
- (66) Savelyev, A. Do monovalent mobile ions affect DNA's flexibility at high salt content? *Phys. Chem. Chem. Phys.* **2012**, *14*, 2250–2254.
- (67) Parkinson, G.; Wilson, C.; Gunasekera, A.; Ebright, Y. W.; Ebright, R. E.; Berman, H. M. Structure of the CAP-DNA complex at 2.5 angstrom resolution: A complete picture of the protein-DNA interface. *J. Mol. Biol.* **1996**, *260*, 395–408.
- (68) Bruner, S. D.; Norman, D. P. G.; Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **2000**, *403*, 859–866..