

Structural Basis of Pathway-Dependent Force Profiles in Stretched DNA

Daniel R. Roe* and Anne M. Chaka

Physics Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8443, Gaithersburg, Maryland 20878-8443

Received: July 16, 2009; Revised Manuscript Received: September 24, 2009

Understanding the mechanical properties that determine the flexibility of DNA is important, as DNA must bend and/or stretch in order to function biologically. Recent single-molecule experiments have shown that above a certain loading rate double-stranded DNA is more stable when stretched from the 3' termini than when stretched from the 5' termini. Unfortunately these experiments cannot provide insight into the structural basis for this behavior. We have used molecular dynamics simulations combined with umbrella sampling to study the stability and structural changes of a 30 bp double-stranded DNA oligomer during stretching from either the 3' termini or the 5' termini. At extensions greater than $1.7\times$ the 3' stretched structure is more stable than the 5' stretched structure due to retention of twice the number (80%) of native hydrogen bonds between base pairs and a higher degree of base stacking. This difference results from greater dissipation of the stretch force via conformational flexibility of the phosphate backbone when pulled from the 3' ends, whereas in the 5' stretch the force is borne more directly by the base pair hydrogen bonds leading to rupture. In addition, stretching from the 5' end produces a greater widening of the major groove that increases solvent exposure and hydrolysis of the base pair hydrogen bonds. These results demonstrate that 3' stretching and 5' stretching in DNA are fundamentally different processes.

Introduction

DNA plays the extremely important role of storing the genetic information required for life as we know it. DNA is by no means a static molecule—many of the processes in which DNA is involved require it to bend and stretch, or even separate into individual strands. For example, DNA undergoing polymerase activity is known to undergo shortening, while DNA coated with the RecA protein is observed to be extended by about 50%.¹ In addition, short DNA segments have been shown to be flexible, stretching up to $1.1\times$ their canonical B-form length absent any external force.² There is also interest in using DNA as a nanoscopic length and force standard for the calibration of experiments on the nanometer/piconewton scale,³ which requires that the response of DNA to applied force be well-characterized for a variety of extensions. In order to understand the complex biological behavior of DNA or to use it as a standard reference material, it is necessary to study the underlying mechanical properties that determine the dynamics of DNA.

Over the past decade and a half many different experimental techniques have been used to probe these properties of DNA at the single-molecule level.^{4–6} In such experiments, DNA is typically tethered on one end to a surface and on the other end to a force measuring probe (which can be an optical trap, atomic force microscope tip, etc.). The distance of the probe to the surface is varied, and the force response of DNA as it is stretched is measured. Initial experiments of this kind showed that the force required to stretch λ -DNA (>100 bp) reached a plateau at ≈ 65 –70 pN, where it could be reversibly stretched up to $\approx 1.7\times$ its contour length with very little force increase.^{5,6} It was theorized that this force plateau indicated a highly cooperative transition from a B-DNA conformation to what was termed an overstretched or “S-DNA” conformation.⁵ However, it was subsequently proposed by Rouzina et al. that instead of

a transition to a new conformation the force plateau was indicative of force-induced melting of DNA.^{7–9} These two different interpretations of experimental results illustrate the importance of creating models to explain such discrepancies. While experiments have been able to characterize the physical properties of the elastic behavior of DNA quite well, there is still much that is unknown about the specific structural changes that occur in DNA during stretching.

Although the majority of DNA stretching experiments have been conducted by stretching from the 5' terminal ends, recent studies have also examined stretching from the 3' terminal ends. There are varying reports as to whether there is any significant difference between these two types of stretch. Hatch et al. used magnetic tweezers to stretch DNA at very low loading rates (near zero) from either the 5' or 3' ends, and observed no difference in rupture force between the two types of stretching.¹⁰ Albrecht et al. used a molecular force balance setup in which a reference DNA duplex was pulled from the 5' ends and a sample DNA duplex was pulled from the 3' ends.¹¹ Their measurements showed that at low loading rates (≈ 90 pN/s) stretching was symmetrical, indicating no difference in rupture profiles between the two types of stretch and consistent with the results from Hatch and co-workers. However, at high loading rates ($>9 \times 10^5$ pN/s) an asymmetry in the force balance developed. This asymmetry was the result of the sample DNA duplexes stretched at the 3' ends having a higher stability than the reference DNA duplexes stretched from the 5' ends. When a subsequent experiment had the reference and sample both stretched from the 5' ends, symmetry was restored. These experiments hint at a complex stretching mechanism for DNA that depends on both the rate of pulling and where DNA is stretched from. However, in both of these experiments only the rupture force can be measured; no data are available for the force vs displacement behavior prior to rupture. As a result, there are no details on

* Corresponding author. Telephone: (301) 975-8741.

what the stretching mechanism might be or what structural changes might give rise to the observed behavior.

Several computational studies have been conducted in which an attempt to model DNA stretching experiments is made, either through direct manipulation and adiabatic minimization of internal helical coordinates^{5,12–14} or through use of molecular dynamics (MD) simulations.^{15–18} Lebrun and Lavery employed the former method with a distance-dependent dielectric to approximate the effects of solvent, and in two separate studies compared several different types of stretching including 5′–5′ and 3′–3′ termini stretching.^{12,13} They found in their simulations that DNA could be stretched up to about 2 times its contour length without disrupting any Watson–Crick (WC) base pairing, but the final structure of the DNA after extension depended on how the DNA was stretched. When DNA was stretched from the 5′ terminal ends, it adopted a narrow, fiberlike structure with highly inclined base pairs. In contrast, when DNA was stretched from the 3′ terminal ends, the DNA adopted a flat, ladderlike ribbon structure. Structures resembling ones from the 5′^{14,16,17} and 3′¹⁵ terminal stretch paths have been observed in several other simulations of DNA stretching as well.

In contrast to the experimental results of Albrecht et al., the simulations of Lebrun and Lavery showed an energetic preference for the 5′ fiber structure over the 3′ ribbon structure, which they attributed to better base-stacking interactions in the fiber. However, they did point out that this could be due to the simplicity of the distance-dependent dielectric solvent model used, which may have led to overdamping of charges. In addition, since what Lebrun and Lavery calculated were just enthalpies, they cannot be used to directly compare stability, as this requires the calculation of free energies. They also noted that the 5′ fiber structure had shorter phosphate–phosphate distances than the 3′ ribbon structure. Albrecht et al. theorized that the unfavorable electrostatic interactions introduced by the shorter distances between the negatively charged phosphates in the 5′ fiber may explain why they observed 5′ stretching to be less stable than 3′ stretching in their experiments,¹¹ but this has not yet been tested.

In an attempt to resolve this discrepancy between recent experimental results and early computational studies, we systemically investigate the factors that may be responsible for the observed difference between 5′–5′ and 3′–3′ stretching. We employ computational methods to model DNA stretching in order to obtain a detailed mechanistic understanding of each type of stretch. Due to limited resources, previous computational studies have had to make use of either short sequences (in which end effects can play a large role), approximate and often inaccurate solvent models, or both. In addition, ongoing development of the force fields used in computational simulations of DNA have revealed issues with certain parameters¹⁹ which may have affected the results of some earlier studies. Whereas previous computational studies have looked at stretching DNA sequences up to 15 bp in length,¹² here we study the stretching of a 30 bp DNA sequence, which is similar to the length of DNA studied in the Albrecht et al. experiments. This is also long enough for three complete turns of DNA, and should lessen the impact of any end effects (terminal fraying, etc.). Our simulations are performed using explicit solvent and making use of the latest nucleic acid force field parameters from Amber.¹⁹ To our knowledge these are currently the largest simulations of this kind.

In order to compare the stability of the 5′–5′ stretching pathway to the 3′–3′ stretching pathway, we used the umbrella sampling (US) method in a serial fashion to estimate the free

energy of DNA along each pathway. Previous studies of DNA stretching using US were run in a parallel fashion, where the starting structures for each US window were taken from stretching simulations at high velocity.^{16,18} This has the potential to introduce any structural artifacts resulting from the high velocity of stretching into the US simulations. In the serial umbrella sampling (SUS) method the initial coordinates of each US window are the final coordinates of the previous window; therefore each successive window benefits from a longer equilibration time. The longer equilibration time generates a better representation of the equilibrium ensemble of stretched structures than molecular dynamics (MD) simulations, and is closer to experimental conditions. For this reason we felt that SUS would be a better choice.

In this work we will show that the distribution of the stretching force throughout the strands of DNA differs for each stretching pathway, resulting in the DNA adopting significantly different final stretched conformations. At large extensions ($>1.7\times$) the 3′ stretched conformation of DNA becomes more stable than the 5′ stretched conformation. A detailed analysis of the specific structural changes (chi angles, sugar puckers, and number of Watson–Crick (WC) hydrogen bonds) that occur during stretching is used to delineate the molecular mechanism responsible for force propagation along the strands of DNA, as well as explain the observed difference in stability. Using observations from previous experimental observations and the current SUS simulation results, we expand upon the rate-dependent mechanism of DNA stretching proposed by Albrecht et al.¹¹

Methods

All molecular dynamics (MD) simulations were performed with PMEMD from Amber 10²⁰ using the FF99 force field²¹ with PARMBSC0 corrections for nucleic acid parameters,¹⁹ which have been shown to be considerably more accurate for simulations of nucleic acid structure than FF99 alone.²² All stretching simulations were performed on a 30 bp sequence used by Morfill et al.²³ (5′-CGTTGGTGCGGATATCTCGGTAGTGG-GATA-3′). This sequence was chosen as it was shown to have a force response similar to larger stretches of DNA (i.e., the existence of a force plateau). The initial coordinates of this sequence were assigned a B-DNA conformation using the NAB program in Amber. This system was then solvated with enough TIP3P²⁴ water molecules so that box dimensions after minimization and equilibration were approximately 4.5 nm \times 4.5 nm \times 20.5 nm. Charge neutralization was achieved with the use of 58 Na⁺ ions, corresponding to a concentration of ≈ 0.23 M. The final system size was 41 290 atoms (1902 DNA, 58 Na⁺, 39 330 solvent). The system was then minimized using a combination of steepest descent and conjugate gradient methods followed by equilibration MD under NVT conditions while restraining the position of all solute heavy atoms; the minimization/MD procedure was then repeated several times while gradually relaxing the restraints.

All MD simulations were conducted under NPT conditions after minimization and equilibration. A constant temperature of 300 K was maintained throughout each simulation by use of a Berendsen²⁵ thermostat with a coupling constant of 1.0 ps. Constant pressure was maintained at 1.0 bar using isotropic position scaling with a relaxation time of 1.0 ps. The PME method was used for long-range electrostatics with a cutoff of 8.0 Å. The SHAKE algorithm²⁶ was used to constrain bonds to hydrogen, allowing a 0.002 ps MD time step. Coordinates were recorded every 500 steps (1 ps).

Serial umbrella sampling (SUS) simulations were performed for both 5' and 3' stretching. The reaction coordinate was either the distance between the C5' atoms of each 5' terminal residue or the distance between the C3' atoms of each 3' terminal residue. For each stretch type the initial umbrella sampling window was started at a relative extension of $1.0\times$ (104.11 Å for 5' and 98.22 Å for 3'). This distance was maintained during the simulation by applying a harmonic distance restraint with a force constant of $2.0 \text{ kcal mol}^{-1} \text{ Å}^{-1}$ between the atoms defining the reaction coordinate. The final structure from the $1.0\times$ simulation was then used as the starting structure for a simulation at $1.025\times$ extension. Simulations were performed in this manner in $0.025\times$ increments up to a relative extension of $1.975\times$ for a total of 40 SUS simulation windows for each stretch type. Each US window was run for 1 ns. While coordinates were written every 500 steps (1 ps), restraint data were written every time step, for a total of 500 000 restraint data points per US window. The first half of these data was considered to be equilibration, and the last half of these data was used to reconstruct the free energy along the reaction coordinate (i.e., the potential of mean force) using the Weighted Histogram Analysis Method (WHAM)²⁷ as implemented by Alan Grossfield.²⁸ All structure pictures were rendered using VMD 1.8.6.²⁹

Results

Serial Umbrella Sampling Simulations. Serial umbrella sampling simulations were used to calculate the potential of mean force (PMF) along both the 5' and the 3' stretching pathways of DNA. Since the C5'–C5' and C3'–C3' reaction coordinates are slightly different due to the fact that the C3' atoms are located on the sugar ring and the C5' atoms are located off the sugar ring, the absolute stretch distance is not an appropriate reaction coordinate for comparing the two types of stretching; a C5'–C5' distance of 104 Å represents no stretching for 30 bp of DNA, while a C3'–C3' distance of 104 Å represents slight stretching. Therefore, relative stretch was chosen as the reaction coordinate in order to directly compare the different types of stretch. Relative stretch is defined as the distance between the C5' or C3' terminal atoms of each strand divided by their initial distance in the equilibrated B-form structure (104.11 Å for 5' stretching and 98.22 Å for 3' stretching).

Simulations were run at relative stretches of $1.0\times$ to $1.975\times$ in $0.025\times$ increments, for a total of 40 umbrella sampling windows. The initial coordinates of each umbrella sampling window were the final coordinates of the previous window, with the exception of the first window at $1.0\times$ which used the equilibrated and minimized B-DNA conformation. See Methods for full details on the SUS simulations.

A good measure of the overall stability and structural integrity of DNA during stretching is the fraction of native Watson–Crick (WC) hydrogen bonds present as a function of relative stretch. Figure 1A,B shows these data for the 5' and 3' stretching pathways. A hydrogen bond was considered present if the heavy atom distance was less than 3.5 Å. It is seen that, as the DNA is stretched, it retains a much higher degree of native base pairing along the 3' stretching pathway compared to the 5' stretching pathway, particularly at extensions greater than $1.7\times$. When stretched from the 5' termini, DNA begins to lose hydrogen bonds right away, in contrast to stretching from the 3' termini, where there is not significant hydrogen bond loss until an extension of $1.3\times$. The final 5' stretched conformation

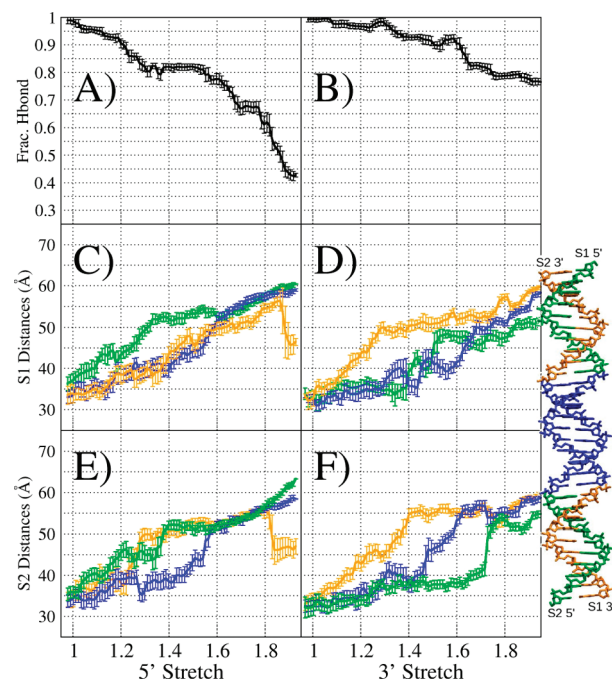


Figure 1. (A, B) Fraction of native WC hydrogen bonds present as a function of relative stretch for 5' and 3' stretching, respectively. (C, D) 5Prime (green, bases 1–10), Central (blue, bases 11–20), and 3Prime (orange, bases 21–30) region distances of strand 1 as a function of relative stretch for 5' and 3' stretching, respectively. (E, F) 5Prime, Central, and 3Prime region distances of strand 2 as a function of relative stretch for 5' and 3' stretching, respectively. The points of each line are the binned averages of results from all US windows, ± 1 standard deviation.

has only $\approx 40\%$ of its hydrogen bonds remaining, while the final 3' stretched conformation has $\approx 80\%$ of its hydrogen bonds remaining.

In order to understand the difference in hydrogen bonding between the two types of stretching, a detailed study of how the stretching force affects the structure of DNA is necessary. Two questions of interest are the following: (1) How is the stretching force distributed throughout the DNA helix during the course of a stretching experiment (i.e., how close to an ideal spring does each individual strand of duplex DNA behave)? (2) Is the distribution of stretching force different for 5' stretching than it is for 3' stretching? In either a 5' or a 3' stretching experiment, each individual strand of DNA has force applied to it at one end only. In order to follow the effect of the stretching force as it travels through a strand, each strand is divided into three regions: 5Prime, defined as the first 10 bases of the strand starting from the 5' terminus (green bases in Figure 1); Central, defined as the 10 bases in the middle of the strand (blue bases in Figure 1); and 3Prime, defined as the last 10 bases in the strand near the 3' terminus (orange bases in Figure 1). Each region has an initial distance of approximately 34 Å in B-DNA. The dynamic behavior of each region was monitored by measuring the C5'–C3' distance of the strand within that region versus the overall relative stretch. This essentially allows direct observation of the effect of the stretching force as it propagates through each strand. If the forces are distributed relatively evenly throughout the helix, one would expect similar extension behavior from the same regions of each strand, whereas if the forces are not evenly distributed their extensions should be different.

Figure 1C,E shows the strand region distances for 5' stretching. The 5Prime region begins to extend first in each strand;

this is not surprising considering these regions are closest to the stretching force. After a very brief lag the 3Prime and Central regions also begin to extend at a relative stretch of $\approx 1.2\times$. The short lag time indicates that the strands are somewhat coupled; the 3Prime region of each strand is likely experiencing stretching forces via base pairing with the corresponding 5Prime region of the opposite strand. In strand 2, the 3Prime region extends before the Central region, but in strand 1 the 3Prime and Central regions extend at roughly the same time. At a relative extension of $1.65\times$, each region in both strands reaches an extension of ≈ 50 Å; after this the extension of all regions occurs at approximately the same rate, indicating at this point in the 5' stretch simulations that the stretching force has become evenly distributed throughout both strands. By the end of the 5' stretch US simulation the extension of the 3Prime region of each strand begins to drop sharply. This corresponds to the strands beginning to dissociate from each other.

Figure 1D,F shows the strand region distances for 3' stretching. As was the case for 5' stretching, during 3' stretching the region closest to the stretching force (3Prime in this case) begins to extend first in each strand. However, the extension of the Central and 5prime regions does not immediately follow, as was the case in 5' stretching. It is only at a relative stretch of $\approx 1.4\times$ that the Central and 5Prime regions of strand 1 begin to extend. The Central region of strand 2 begins to extend here as well, followed by the 5Prime region at an extension of $1.7\times$. Examination of the structures reveals that the extension of the 5Prime region of strand 2 remains low until $1.7\times$ due to the fraying of two base pairs at that end of the DNA duplex. The longer lag time indicates that during 3' stretching the strands are not as coupled to each other through base pairing as they are during 5' stretching. At a relative extension of $\approx 1.75\times$ the distances of strand 2 become approximately equal, reaching ≈ 55 Å, although the 5 prime region is slightly smaller. However, the region distances in strand 1 at the same point range from 45 to 55 Å. This is in contrast to 5' stretching, where after $1.65\times$ each region of both strands appeared to be highly correlated, and indicates that during 3' stretching the stretch force is not distributed as evenly throughout the strands.

The potential of mean force (PMF) and stretching force (calculated from the slope of the PMF curves) as a function of DNA extension are shown for both 5' and 3' stretching in Figure 2B and 2A, respectively. The force and energy curves for each type of stretch are initially indistinguishable. The force curves (Figure 2A) for 3' and 5' stretching show an initial gradual increase in stretching force until around a relative stretch of $1.1\times$, after which the force plateaus; correspondingly, the PMF curves (Figure 2B) increase with constant slope. The average force over the plateau region (defined as a relative stretch of $1.1\times$ to $1.6\times$) is 107 ± 37 pN for 5' stretching and 115 ± 33 pN for 3' stretching. These results are in good agreement with experimental values of the force plateau, which have been reported in the range of ≈ 65 pN²³ to ≈ 133 pN³⁰ for short oligonucleotides.

The forces and stabilities of 5' and 3' stretching begin to differ greatly after a relative stretch of $\approx 1.65\times$. For the 5' stretch the PMF and the force both begin to increase rapidly, and by a relative stretch of $1.9\times$ the force has increased past 600 pN. It should be noted that $1.65\times$ is also the point in 5' stretching where the strand lengths become approximately equal (Figure 1C,E). This is also approximately where the force is observed to rise from the plateau region in experiments ($1.7\times$). During 3' stretching, however, the PMF stays much lower and the force remains relatively flat until an extension of $\approx 1.75\times$, after which

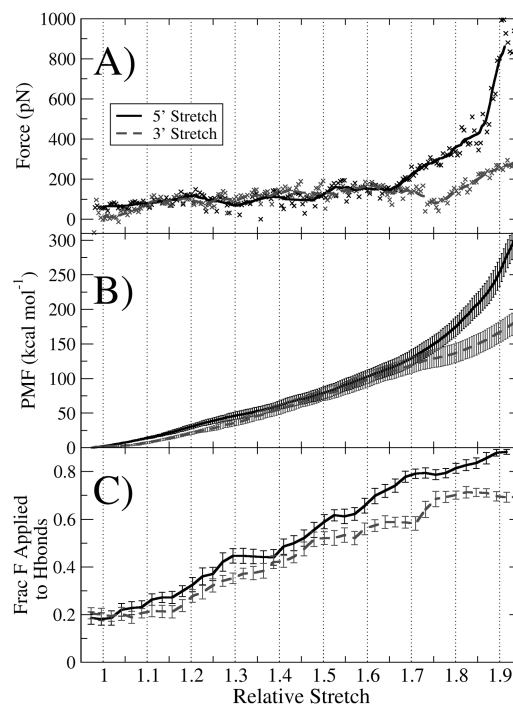


Figure 2. Stretching force, PMF (free energy), and average fraction of stretching force applied across the WC hydrogen bonds as a function of relative stretch for SUS simulations of 5' (solid lines) and 3' (dashed lines) stretching. (A) Stretching force calculated from the slope of the PMF curves. (B) Final PMF curves for each run were obtained from averaging the PMF curves from the first and last halves of the SUS data; error bars are half the absolute difference. (C) The average fraction of hydrogen bonding was calculated for each SUS window; error bars are ± 1 standard deviation.

the force begins to increase. However, by $1.9\times$ the force and PMF are still far below those seen for 5' stretching. These results indicate that at extensions larger than $\approx 1.7\times$ DNA is more stable when stretched from the 3' ends than when stretched from the 5' ends.

Figure 2C shows the fraction of restraint force applied across the WC hydrogen bonds during DNA stretching. The points of each line are the binned averages of results from all US windows, ± 1 standard deviation. The more aligned the hydrogen bond is to the axis of the stretching force, the greater the force that hydrogen bond will feel, increasing the chance it will be broken. More force is applied across the WC hydrogen bonds in 5' stretching than in 3' stretching. At a relative extension of $1.7\times$, the fraction of stretching force applied to the WC hydrogen bonds in 5' stretching is 0.76, whereas in 3' stretching it is 0.59. At the final extension the fraction of stretching force applied has become more similar for the 5' and 3' stretching (0.84 versus 0.71). It is interesting to note that although the stretching force values have become more similar, almost twice as many hydrogen bonds are lost during 5' stretching compared to 3' stretching (Figure 1A,B). The reason for this will be discussed in the next section, where the specific structural details of each stretching pathway are examined.

The results from the SUS simulations show that at extensions greater than $1.7\times$ the 3' stretched structure has a lower free energy and is consequently more stable than the 5' stretched structure. It may be that the higher degree of hydrogen bonding in the 3' stretched structure is the reason behind the higher stability, but in order to fully understand how these conformations are attained, a more detailed analysis is required. In the next section we examine the structural basis for the enhanced

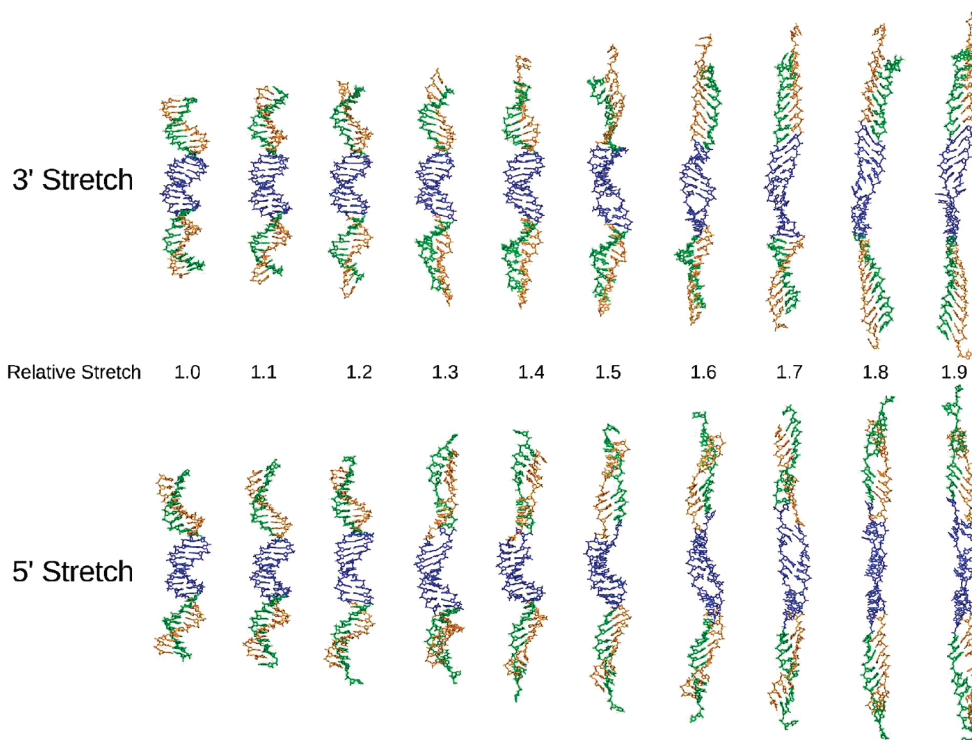


Figure 3. Final structures of SUS windows from 3' and 5' stretching at relative stretch values ranging from 1.0 \times to 1.9 \times . In each strand green denotes the 5Prime region (bases 1–10), blue denotes the Central region (bases 11–20), and orange denotes the 3Prime region (bases 21–30).

stability of the 3' stretched conformation and propose a theoretical mechanism for 5' vs 3' stretching.

It is worth mentioning that we did attempt to run the umbrella sampling simulations in parallel fashion, using structures taken from MD simulations of DNA stretching at constant velocity. The resulting PMF (data not shown) diverges from the SUS PMF at a relative stretch of 1.15 \times . By 1.4 \times it is ≈ 10 kcal/mol larger, and by 1.7 \times it is ≈ 20 kcal/mol larger. We do note that although these SUS simulations are better converged, the free energies calculated from them represent only an upper bound to the free energy.

Structural Details of 5' and 3' Stretched Conformations.

The final structures from umbrella sampling windows ranging from a relative stretch of 1.0 \times to 1.9 \times in 0.1 \times intervals are shown for both 5' and 3' stretching in Figure 3. Here it is seen that the structures adopted during the course of the 5' stretch are quite different than the structures adopted during the course of the 3' stretch. Over the course of 5' stretching the DNA becomes narrower (i.e., the distance between the strands decreases) with the bases in each strand highly inclined toward the 3' termini, somewhat similar to 5'-stretched structures proposed by Lebrun and Lavery.^{12,13} However, in the Lebrun and Lavery structures the WC base pairing remained completely intact, and about every other stacking interaction was interrupted. The structures in Figure 3 clearly show loss of WC base pairing, and base stacking interactions are interrupted about every four bases. Other simulations of 5' stretching have observed structures with similar losses of base pairing and stacking.^{16–18}

At a relative stretch of 1.3 \times there are already some gaps in the base stacking interactions throughout the 5' stretched structure. By 1.7 \times the 5' stretched structure has multiple gaps in base stacking and has lost several base pairs due to WC hydrogen bond disruption. In addition, the major groove width along the 5' stretched DNA increases to become as large as 30 Å, while the minor groove width remains constant at around 8 Å (groove width analysis performed with CURVES³¹). The

effect of this is to push base pairs into the solvent through the major groove. This is reflected in the large change in solvent exposed surface area (calculated using the linear combination of pairwise overlaps method³² as implemented in the SANDER module of Amber), which rises from its initial value of $\approx 10\,500$ Å² to $\approx 12\,000$ Å² by a relative stretch of 1.7 \times , and ends up at $\approx 12\,600$ Å².

In contrast, over the course of 3' stretching the DNA becomes wider (i.e., the distance between the strands increases) and becomes mostly unwound. The bases tend to incline toward the 5' termini, although the degree of inclination appears to be less than that seen during 5' stretching, particularly toward the Central region of the DNA. The 3' stretched conformations here resemble those seen by Lebrun and Lavery,^{12,13} except again that their structures showed no base pairing loss; this is probably because the simple distance-dependent dielectric used to model solvent in those simulations would tend to disfavor loss of WC hydrogen bonds. These 3' stretched conformations also resemble those seen by Konrad and Bolonick in their simulations of a dodecamer.¹⁵

The 3' stretched structures show mostly intact base pairing, consistent with the results shown in Figure 1B, as well as intact base stacking. In fact, the 3' stretched structures remain largely undisturbed until a relative stretch of 1.5 \times , where some pairing/stacking interactions are lost in the Central region and at both termini. Unlike 5' stretching, during 3' stretching both the major and minor groove widths increase, and by 1.9 \times they are almost the same (18 and 15 Å, respectively). As a consequence, the bases are pushed in toward the minor groove and away from solvent. The solvent accessible surface area increases from $\approx 10\,500$ to $\approx 11\,000$ Å² by a relative stretch of 1.7 \times , and ends up at $\approx 11\,500$ Å². The bases in the 3' stretched structures are overall less solvent exposed, which may explain why base stacking interactions appear to be disrupted far less in the 3' stretched structures compared to the 5' stretched structures. This

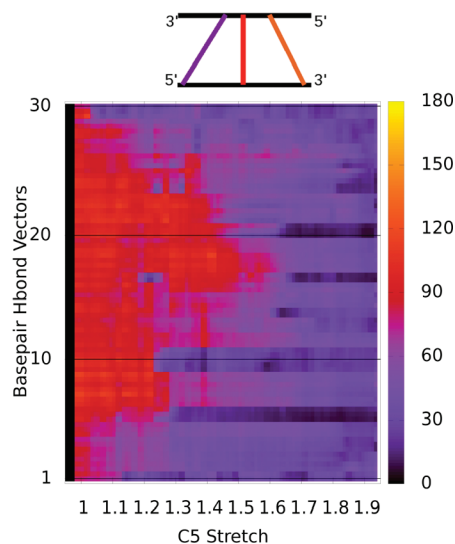


Figure 4. Inclination of WC hydrogen bonds with respect to the axis of stretching between base pairs as a function of relative stretch from SUS simulations of 5' stretching. An inclination of 90° is considered perpendicular to the axis of stretching, 0° indicates inclination toward the 3' terminus, and 180° indicates inclination toward the 5' terminus. The diagram depicts these three types of inclination.

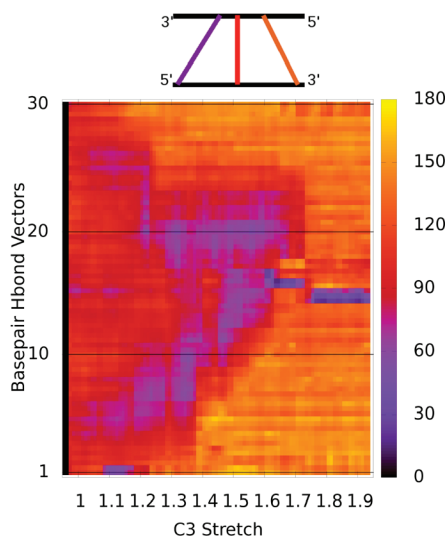


Figure 5. Inclination of WC hydrogen bonds with respect to the axis of stretching between base pairs as a function of relative stretch from SUS simulations of 3' stretching.

movement of the bases away from solvent toward the minor groove is also what leads to the overall widening of the DNA.

The inclination of the base pairs with respect to the helical axis has a direct bearing on the structural stability of DNA. The native WC hydrogen bonds between a given base pair lie along the plane formed by the base pair, so as the base pair becomes more inclined, so do the hydrogen bonds. As the bases become more inclined, a greater portion of the stretching force acts upon the hydrogen bonds, increasing the probability they will rupture (as pointed out by Harris et al.¹⁷).

A plot of hydrogen bond angles with respect to the axis of stretching is shown for 5' stretching in Figure 4 and 3' stretching in Figure 5. For the purposes of this study an inclination of 90° indicates the hydrogen bond is perpendicular to the axis of stretching, an inclination of 0° indicates alignment with the axis of stretching toward the 3' terminus, and 180° indicates alignment with the axis of stretching toward the 5' terminus.

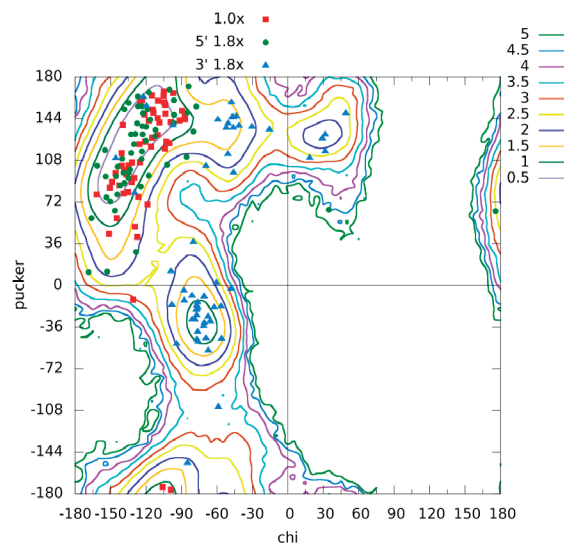


Figure 6. Surface showing the distribution of chi vs sugar pucker values for all DNA bases calculated from two-dimensional histogram bin populations from all SUS windows (both 5' and 3' stretching, 80 windows total). Points representing the chi and sugar pucker values of bases in the final structures at relative stretch values of 1.0 \times , 5' 1.8 \times , and 3' 1.8 \times are also shown using red squares, green circles, and blue triangles, respectively.

The hydrogen bond angles in the B-DNA conformation are all at $\approx 90^\circ$, perpendicular to the axis of stretching. During 5' stretching, the angles of the hydrogen bonds at the termini immediately begin moving toward 0°, inclining toward the 3' terminus of each strand. This change in inclination happens in a stepwise fashion, moving from the termini toward the center of the DNA. By an overall relative extension of 1.65 \times , all of the hydrogen bonds in the 5' stretched structure have inclinations of less than 45°. It is interesting to note that it is around this relative extension value that the individual strand lengths become approximately equal (Figure 1C–F) and the stretching force begins to rise from the plateau region (Figure 2). The apparent stepwise nature of the transition is indicative of a cooperative process.

The inclination of hydrogen bond angles is quite different during 3' stretching, however. The hydrogen bonds at the termini begin to incline toward 180°, inclining toward the 5' terminus of each strand. However, the hydrogen bonds adjacent to the termini initially incline in the opposite direction, toward 0°, before they eventually begin to incline toward 180° as well. This effect appears somewhat cooperative at one end of the DNA (near base pair 1, T–A), but not as much at the other, where it appears to stall at base pair 23, C–G. In direct contrast to 5' stretching, by the end of 3' stretching a variety of hydrogen bond inclinations are observed, with hydrogen bonds near the termini having inclinations greater than 135° and hydrogen bonds near the Central region having inclinations less than 135°.

In addition to inclination, we evaluate the structural changes in DNA in terms of the sugar pucker and the chi dihedral angle.³³ The sugar pucker defines the direction in which the five atoms of the sugar ring deviate from a plane, and the chi angle measures rotation of a base around the glycosidic bond, and by definition can be either anti (less than -90°) or syn (greater than -90°).

Figure 6 shows a surface representing the distribution of chi vs pucker values of each base from all serial umbrella sampling windows for 5' and 3' stretching. The value for each bin was calculated using the formula $-k_B T \ln(N_{\text{Bin}}/N_0)$, where k_B is

Boltzmann's constant, T is the temperature (300 K), N_{Bin} is the bin population, and N_0 is the most populated bin. In addition, the final chi vs pucker values for DNA at extensions of $1.0\times$, $5' 1.8\times$, and $3' 1.8\times$ are shown as red squares, green circles, and blue triangles, respectively. The sugar pucker values were calculated using a pseudodihedral angle as described by Altona and Sundarlingam.³⁴

At a relative stretch of $1.0\times$ (corresponding to the B-DNA form), the chi vs pucker values for all bases cluster mostly in the broad minimum in the upper-left quadrant of the landscape; the bases adopt all anti conformations, and have mostly C2'-endo and C1'-exo pucker values. There is almost no penalty for shifting between these pucker values, and the shift from C2'-endo to C1'-exo correlates with a chi angle shift toward -180° . When stretched in the $5'$ direction, the chi/pucker values of the bases are still clustered in the same broad minimum, and the bases are still in all anti conformations. The sugar pucker values are also similar but are slightly more restricted than $1.0\times$, shifting to mostly C1'-exo and O3'-endo.

When stretched in the $3'$ direction, however, the distribution of chi vs pucker values looks quite different. For the $3'$ stretch at $1.8\times$ almost none of the bases remain in the original minimum, instead clustering in three new minima. The majority of the chi vs pucker values now fall into the minimum in the lower-left quadrant, in which the bases adopt mostly syn conformations (chi $\approx 75^\circ$) and a C2'-exo pucker. These bases are located at the termini of the DNA (in the 5Prime and 3Prime regions of each strand) and have hydrogen bond inclination values close to 180° (Figure 5). The bases in the second most populated minimum (back in the upper-left quadrant) are also syn, with chi values around -50° , and adopt a C1'-exo pucker. These bases are located around the Central regions of each strand, with hydrogen bond inclination values closer to 120° . The final minimum (upper-right quadrant) has even larger syn chi angle values, and also adopts a C1'-exo pucker—these bases have lost base pairing at the termini of the DNA. The adoption of syn chi values by even some pyrimidine bases (which is highly unfavorable) reflects the extreme nature of the forces imposed on the DNA due to stretching. It appears that in $3'$ stretching the energy of the stretching force initially goes toward adopting this metastable conformation, characterized by increased backbone flexibility. This is in contrast to $5'$ stretching where the backbone is not as flexible, and so the energy of the stretching force goes immediately toward breaking the WC hydrogen bonds.

Conclusions

Through computational methods we have determined a model for DNA stretching that explains the observed difference between $5'-5'$ and $3'-3'$ stretching. It is seen that $3'$ and $5'$ stretching are initially similar in terms of stability, and that the "overstretching transition" is primarily the result of reorganization of the sugar-phosphate backbone. At extensions greater than $1.7\times$ the $3'$ stretched structure becomes more stable than the $5'$ stretched structure, consistent with experimental results.¹¹ The reason for this is in the structural details of the different conformations adopted by DNA over the course of the $5'$ and $3'$ stretching pathways. Our results indicate that there is a fundamental difference in how the stretching force propagates along the phosphate-sugar backbone due to the availability of elongated metastable conformations during $3'$ stretching that are not accessible during $5'$ stretching. This conformational flexibility of the backbone during $3'$ stretching allows the force to be dissipated during elongation without significant disruption

of WC hydrogen bonded base pairs. In contrast, the relative lack of backbone flexibility during $5'$ stretching results in the near-immediate distribution of the stretching force across the WC base pairs, resulting in tighter coupling of strand motion throughout the length of the DNA and more rapid and extensive rupture of the WC hydrogen bonds.

During $3'$ stretching, DNA adopts a metastable conformation characterized by significant changes in sugar pucker and chi dihedral angle values. As the DNA extends, the bases in a given strand undergo a series of conformational changes which involve the chi dihedral angle adopting a syn conformation and the sugar pucker adopting a C2'-exo or C1'-exo conformation. This additional flexibility in each strand results in the stretching force propagating from the ends of the individual strands of the DNA toward the center. Since the stretching force initially goes into the adoption of this new conformation, less of the stretching force is applied directly across the WC hydrogen bonds, allowing the DNA to maintain its base pairing interactions longer. After this conformation is adopted the force begins to rise from the plateau region, indicating at that point that the hydrogen bonds are now being disrupted. Due to the changes in the backbone, the bases turn in toward the minor groove and away from solvent as DNA is stretched from the $3'$ termini, resulting in the major and minor groove widths both increasing and eventually having similar values. This widening is accompanied by a flattening and unwinding of the double helical structure. The bases in this $3'$ stretched structure are less solvent exposed, which allows the DNA to better maintain base stacking interactions.

In contrast, during $5'$ stretching the backbone does not have similar flexibility, as indicated by how little the chi and sugar pucker values change throughout $5'$ stretching. As a result, the bases in a given strand immediately incline toward the $3'$ terminus and the stretching force is applied across the WC hydrogen bonds, resulting in their being disrupted much earlier than is the case in $3'$ stretching. Indeed, the stretching force propagates faster throughout each strand of DNA in $5'$ stretching. In contrast to $3'$ stretching, during $5'$ stretching the major groove widens while the minor groove width remains the same, turning the bases into the solvent which allows easier disruption of base pairing and base stacking interactions. The relative instability of $5'$ stretched DNA does not appear to result from unfavorable electrostatics due to decreased phosphate distances, as had been previously suggested.¹¹ Instead, our results show $5'$ stretching to be less stable due to the increased solvent exposure of the $5'$ stretched conformation combined with more of the stretching force applied directly across the WC hydrogen bonds. By $1.7\times$ the stretching force has made the strands taut; this is where the force begins to rise from the plateau region as there is nothing more to do but break WC hydrogen bonds, which is made easier by the increased solvent exposure.

While previous studies have shown that different stretched conformations of DNA can arise depending on whether DNA is pulled from the $3'$ or $5'$ termini, they were unable to predict which might be more stable. This study is the first comparison of the free energies along these pathways, which are essential for any meaningful comparison of stability. In addition, these DNA stretching simulations are currently the largest of their kind, in part due to the use of an explicit solvent model; implicit solvent models may not provide the necessary accuracy needed, particularly in a study such as this where multiple base pairs can be disrupted and exposed to solvent over the course of stretching. Although prior MD simulations were able to show some structural differences between the pathways, it was not

possible to determine the relative stabilities, necessitating the use of an enhanced sampling method like umbrella sampling. SUS simulations allowed the computation of free energies, which showed that at extensions greater than $1.7\times$ DNA is more stable when stretched from the 3' termini than when stretched from the 5' termini.

It should be emphasized that these results are currently more relevant for interpretation of experiments in the high loading rate regime since the free energies calculated from the SUS simulations are not completely converged. At low loading rates (below 9×10^5 pN/s) Albrecht et al. did not observe any difference between 5' and 3' stretching using a molecular force balance,¹¹ and Hatch et al. did not observe any difference in rupture forces for 5' and 3' stretching at loading rates close to zero.¹⁰ Albrecht et al. suggested that at low loading rates the DNA duplex unbinds from force induced melting,¹¹ as has also been suggested by Rouzina et al.^{7–9} It may be that at slower stretching velocities much of the DNA structure has already been melted by the time the extensions necessary to observe the effect predicted in this study are reached. This is partially supported by the fact that there is a certain degree of base pair loss in the SUS simulations; it is possible if they could be extended to the microsecond time scale base pairing could be lost before extensions greater than $1.7\times$ could be reached. It is known that a high concentration of salt (≈ 1 M) can significantly stabilize DNA during stretching,³⁵ so it is possible that stretching experiments using higher salt concentrations might be better able to detect the difference between 5' and 3' DNA stretching at lower loading rates.

Acknowledgment. We thank Gordon Shaw and John Pratt for useful insights into DNA stretching experiments. D.R.R. thanks Christina Bergonzo and Tom Cheatham for helpful discussions about simulating DNA. Computational resources used for this study include the Raritan cluster at NIST, and the Biowulf cluster at the National Institutes of Health. This work was funded by an NRC Postdoctoral Fellowship.

References and Notes

- (1) Allemand, J.; Bensimon, D.; Croquette, V. *Curr. Opin. Struct. Biol.* **2003**, *13*, 266–274.
- (2) Mathew-Fenn, R. S.; Das, R.; Harbury, P. A. B. *Science* **2008**, *322*, 446–449.
- (3) Rickgauer, J. P.; Fuller, D. N.; Smith, D. E. *Biophys. J.* **2006**, *91*, 4253–4257.
- (4) Neuman, K. C.; Nagy, A. *Nat. Methods* **2008**, *5*, 491–505.
- (5) Cluzel, P.; Lebrun, A.; Heller, C.; Lavery, R.; Viovy, J. L.; Chatenay, D.; Caron, F. *Science* **1996**, *271*, 792–794.
- (6) Smith, S. B.; Cui, Y.; Bustamante, C. *Science* **1996**, *271*, 795–799.
- (7) Williams, M. C.; Wenner, J. R.; Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **2001**, *80*, 874–881.
- (8) Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **2001**, *80*, 882–893.
- (9) Rouzina, I.; Bloomfield, V. A. *Biophys. J.* **2001**, *80*, 894–900.
- (10) Hatch, K.; Danilowicz, C.; Coljee, V.; Prentiss, M. *Phys. Rev. E: Stat. Nonlinear Soft Matter Phys.* **2008**, *78*, 011920.
- (11) Albrecht, C. H.; Neuert, G.; Lugmaier, R. A.; Gaub, H. E. *Biophys. J.* **2008**, *94*, 4766–4774.
- (12) Lebrun, A.; Lavery, R. *Nucleic Acids Res.* **1996**, *24*, 2260–2267.
- (13) Lavery, R.; Lebrun, A. *Genetica* **1999**, *106*, 75–84.
- (14) Kosikov, K. M.; Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1999**, *289*, 1301–1326.
- (15) Konrad, M. W.; Bolonick, J. I. *J. Am. Chem. Soc.* **1996**, *118*, 10994.
- (16) MacKerell, A. D. J.; Lee, G. U. *Eur. Biophys. J.* **1999**, *28*, 415–426.
- (17) Harris, S. A.; Sands, Z. A.; Laughton, C. A. *Biophys. J.* **2005**, *88*, 1684–1691.
- (18) Piana, S. *Nucleic Acids Res.* **2005**, *33*, 7029–7038.
- (19) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (20) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (22) Svozil, D.; Sponer, J. E.; Marchan, I.; Perez, A.; Cheatham, T. E., III; Forti, F.; Luque, F. J.; Orozco, M.; Sponer, J. *J. Phys. Chem. B* **2008**, *112*, 8188–8197.
- (23) Morfill, J.; Kuhner, F.; Blank, K.; Lugmaier, R. A.; Sedlmair, J.; Gaub, H. E. *Biophys. J.* **2007**, *93*, 2400–2409.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 935.
- (25) Berendsen, H.; Postma, J.; VanGunsteren, W.; Dinola, A.; Haak, J. J. *J. Chem. Phys.* **1984**, *81*, 3690.
- (26) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 341.
- (27) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1021.
- (28) Grossfield, A. Wham: weighted histogram analysis method; <http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html>.
- (29) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (30) Noy, A.; Vezhenov, D. V.; Kayyem, J. F.; Meade, T. J.; Lieber, C. M. *Chem. Biol.* **1997**, *4*, 519–527.
- (31) Lavery, R.; Sklenar, H. Curves 5.1: helical analysis of irregular nucleic acids; <http://www.ibpc.fr/UPR9080/Curindex.html>.
- (32) Weiser, J.; Shenkin, P. S.; Still, W. C. *Biopolymers* **1999**, *50*, 373–380.
- (33) vanHolde, K. E.; Johnson, W. C.; Ho, P. S. *Principles Of Physical Biochemistry*, 1st ed.; Prentice Hall Inc.: New York, 1998.
- (34) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
- (35) Clausen-Schaumann, H.; Rief, M.; Tolkendorf, C.; Gaub, H. E. *Biophys. J.* **2000**, *78*, 1997–2007.

JP906749J