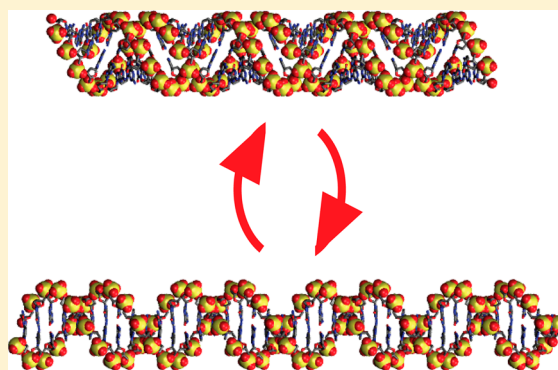


## Transitions of Double-Stranded DNA Between the A- and B-Forms

James T. Waters,<sup>†</sup> Xiang-Jun Lu,<sup>‡</sup> Rodrigo Galindo-Murillo,<sup>§</sup> James C. Gumbart,<sup>†</sup> Harold D. Kim,<sup>†</sup> Thomas E. Cheatham, III,<sup>§</sup> and Stephen C. Harvey<sup>\*,||</sup><sup>†</sup>School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, United States<sup>‡</sup>Department of Biological Sciences, Columbia University, New York, New York 10027, United States<sup>§</sup>Department of Medicinal Chemistry, L.S. Skaggs Pharmacy Institute, University of Utah, Salt Lake City, Utah 84112, United States<sup>||</sup>Department of Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

## Supporting Information

**ABSTRACT:** The structure of double-stranded DNA (dsDNA) is sensitive to solvent conditions. In solution, B-DNA is the favored conformation under physiological conditions, while A-DNA is the form found under low water activity. The A-form is induced locally in some protein–DNA complexes, and repeated transitions between the B- and A-forms have been proposed to generate the forces used to drive dsDNA into viral capsids during genome packaging. Here, we report analyses on previous molecular dynamics (MD) simulations on B-DNA, along with new MD simulations on the transition from A-DNA to B-DNA in solution. We introduce the A-B Index (ABI), a new metric along the A-B continuum, to quantify our results. When A-DNA is placed in an equilibrated solution at physiological ionic strength, there is no energy barrier to the transition to the B-form, which begins within about 1 ns. The transition is essentially complete within 5 ns, although occasionally a stretch of a few base pairs will remain A-like for up to ~10 ns. A comparison of four sequences with a range of predicted A-phobicities shows that more A-phobic sequences make the transition more rapidly than less A-phobic sequences. Simulations on dsDNA with a region of roughly one turn locked in the A-form allow us to characterize the A/B junction, which has an average bend angle of 20–30°. Fluctuations in this angle occur with characteristic times of about 10 ns.

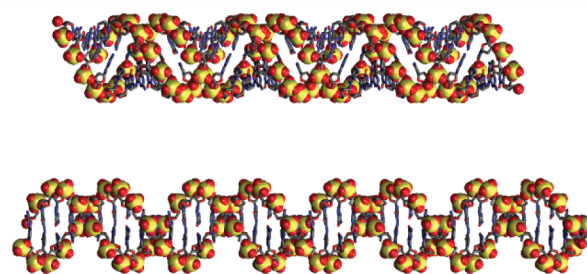


## INTRODUCTION

Rosalind Franklin was the first to remark that the X-ray fiber diffraction pattern of DNA, and therefore its structure, depends on the water activity.<sup>1</sup> She designated the low- and high-humidity forms as “A” and “B”, respectively. The latter was, of course, the structure proposed by Watson and Crick,<sup>2</sup> while the structure of the A-form was reported shortly later by Franklin and Gosling.<sup>3</sup>

Figure 1 shows models of a 40 base pair double-stranded DNA (dsDNA) duplex in idealized A- and B-forms. Among the most visible structural differences is that A-DNA has a substantially smaller rise per residue (2.6 Å) than B-DNA (3.4 Å), so the A-DNA duplex is shorter than the B-DNA molecule. In addition, the base pairs of B-DNA are essentially perpendicular to the molecular axis, while those of A-DNA have an inclination on the order of 20°. (For definitions of the DNA helicoidal parameters such as rise and inclination, see refs 4 and 5.)

The first DNA crystal structure was a surprise, as it was left-handed and had a dinucleotide repeat.<sup>6</sup> Rich called it Z-DNA, both to emphasize how different it was from A-DNA and B-DNA, and because the backbone was not a smooth helix, but a zigzag, due to the dinucleotide repeat.



**Figure 1.** Forty base pairs of B-DNA (bottom) and A-DNA (top). Two distinguishing features are easily seen at this scale. First, A-DNA is much shorter than B-DNA, manifested in the rise per base pair (2.6 Å for A-DNA vs 3.4 Å for B-DNA). Second, the base pairs are essentially perpendicular to the helix axis in B-DNA, but not in A-DNA; the helicoidal parameter that quantifies this is the inclination (~15–20° for A-DNA vs ~0° for B-DNA).

**Special Issue:** J. Andrew McCammon Festschrift

**Received:** February 29, 2016

**Revised:** April 27, 2016

**Published:** May 2, 2016

Table 1. DNA Sequences Used in the Simulations

name	sequence
BigF	AAATTTTGAAAAATTTTTCAAAAAATTTTGAAAAATTTT
LilF	GGGGGTACCCCTGGGGGCCCCCTAGGGGGTACCCCGGGG
MedF	AGATCAGATCTCAGATCGATCGATCTCATGATCTGAGATG
RandF	AGAAGAAGTTAAGCGCCCGAGTATTACCTATTTCGGCTGCT

DNA strongly favors the B-form in vivo, because of the high water activity. There are proteins that bind to Z-DNA,<sup>7,8</sup> and it has been suggested that Z-DNA may play a role in the pathogenesis of vaccinia virus.<sup>9</sup> But for many years, A-DNA was largely regarded as a laboratory artifact, with no biological role.

Regions of DNA with local stretches of A-form do occur in some protein crystal structures.<sup>10,11</sup> Cryo-electron microscopic studies on the rod-shaped SIRV2 virus found that the entire dsDNA genome is in the A-form,<sup>12</sup> because it is completely covered by proteins, giving the solvent very limited access to the double helix. In addition, it has been found that the A form protects DNA in spores from damage during desiccation.<sup>13</sup>

We recently proposed that repeated DNA transitions between the A- and B-forms are responsible for the DNA translocation through the ATP-consuming motors of icosahedral dsDNA bacteriophages.<sup>14</sup> In this “scrunchworm” model, the shortening and lengthening of dsDNA by these cyclic transitions are captured by a coupled grip-release cycle, thus rectifying the motion and advancing the DNA into the capsid. If this is correct, then DNA is not a passive substrate, driven into the capsid by lever-like motions of the motor proteins, as is commonly assumed; instead, it is an active component of the packaging motor. Our model predicts that different DNA sequences will generate different forces during packaging by phage motors,<sup>14</sup> because different base pair steps have differing “A-phobicities”.<sup>15</sup> AT-rich sequences are A-phobic (particularly AA•TT steps), while GC-rich sequences (particularly GG•CC steps) are A-philic.

This study is aimed at examining the “A-ness” and “B-ness” of dsDNA under a variety of conditions and how those characteristics depend on the sequence of the molecule, using molecular dynamics (MD) simulations. We developed a new parameter, the A-B Index (ABI), to quantify the conformations of individual base pair steps along the continuum between the A- and B-forms. We then monitored the transitions of dsDNA molecules with different sequences from the A-form to the B-form when the molecules are free in solution. We extended some of those simulations while holding a region of the DNA locked in the A-form to generate models for the A/B junction whose characteristics we also quantified. Finally, we examined the ABI distributions for different base pair steps in long MD simulations on DNA free in solution under physiological conditions, using the Ascona B-DNA Consortium (ABC) database<sup>16–19</sup> and further runs on one molecule from the ABC set.<sup>20,21</sup>

## METHODS

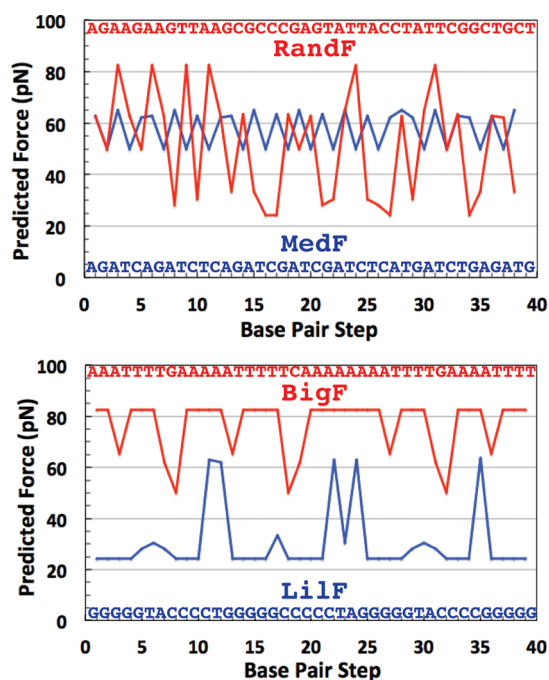
**Molecular Dynamics Simulations.** Models of dsDNA (A-DNA or B-DNA) were solvated with TIP3P water<sup>22</sup> and sufficient ions to neutralize the system in 150 mM NaCl, which favors the B-DNA conformation. Each system was briefly minimized, then brought to  $T = 310$  K, and the solvent was equilibrated while the DNA conformation was restrained. Restraints were then released, and we monitored the progress of the simulation over periods ranging from 10 to 32 ns.

Except as noted, all simulations were carried out with NAMD,<sup>23</sup> using the CHARMM36 force field<sup>24,25</sup> and SHAKE<sup>26</sup> restraints on the bonds, permitting a 2 fs time step. We simulated the isobaric–isothermal ( $NPT = \text{constant}$  number of particles, constant pressure and constant temperature) ensemble with rectangular periodic boundary conditions and the particle mesh Ewald (PME)<sup>27,28</sup> algorithm for treating long-range forces. The initial size of the periodic box was  $50 \times 50 \times 170$  Å. This provided 15 Å for the solvation layer between the DNA and the sides of the box in the lateral direction and 15 Å between the DNA and each end of the box in the direction of the DNA axis after it expands to the B-form.

**Design of DNA Molecules.** We carried out simulations on four molecules, each containing 40 base pairs, but with different sequences (Table 1). Ideal models of A-DNA and B-DNA with these sequences were built using the Web 3DNA server at Rutgers University, [w3dna.rutgers.edu](http://w3dna.rutgers.edu).<sup>29</sup>

Sequence design was based on the predicted forces that would be generated in the  $A \Rightarrow B$  transition in physiological solvent. For base pair step  $X$ , the predicted force is  $F = \Delta G_X / \Delta z$ , where  $\Delta G_X$  is the free energy difference between the B- and A-forms for step  $X$  according to Tolstorukov et al.,<sup>15</sup> and assuming that the free energy difference is expended over a distance of  $\Delta z = 0.8$  Å per base pair, the difference in the rise of the A- and B-DNA double helices.<sup>14</sup> Tolstorukov et al. found the midpoint for the B-A equilibrium for different sequences in mixed water-trifluoroethanol solvents and analyzed the data with a model that assigned different A-phobicities to each of the 10 unique base pair steps. We have used values from their 10-parameter fit (column D-10 in their Table 2). “LilForce” is GC rich, while “BigForce” is AT rich. “MedForce” is composed primarily of base pair steps that are predicted to have intermediate A-phobicities. These three sequences are not only compositionally biased; they are enriched in base pair steps with the desired A-phobicities. “RandForce” was generated with a target composition of equal parts A, C, G, and T. Like MedForce, it is predicted to generate an intermediate force in the viral packaging experiments but the step-to-step values fluctuate more in RandForce than in MedForce. Figure 2 shows the predicted forces for each base pair step in these sequences.

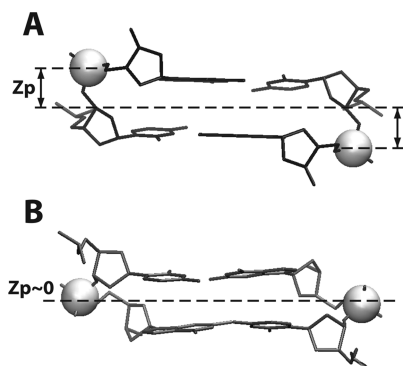
**A Reaction Coordinate for the  $B \rightleftharpoons A$  Transition.** A-DNA and B-DNA differ in many ways, including base pair inclination, X-displacement, rise, twist and sugar pucker. Some structural biologists would argue that “A-DNA” (or “B-DNA”) implies some minimum number of base pairs in the appropriate conformation, though there is no agreement on how many are needed (three? five? more?), or how close they must be to the canonical structure. Few actual DNA structures are purely A-form or B-form. As just one example, individual nucleotides in crystal structures identified as “B-DNA” have a range of sugar puckers, sometimes including puckers that lie in the C3'-endo region of the pseudorotation map<sup>30,31</sup> that are traditionally associated with A-form DNA and RNA double helices. Thus,



**Figure 2.** Forces predicted to be generated by the A $\Rightarrow$ B transition in the scrunchworm model<sup>14</sup> for the four different sequences of Table 1, based on the values of  $\Delta G_{BA}$  from Tolstorukov et al.<sup>15</sup> and the assumption that this energy is expended over a distance of 0.8 Å per base pair, the difference in the rise of the A- and B-DNA double helices.

there is no unique parameter or combination of parameters that can rigorously distinguish between A-like and B-like structures.

Nonetheless, it would be useful to have a single scalar quantity to simplify the characterization of DNA structures along the A–B continuum. We have chosen to use a linear combination of two structural parameters that has previously been shown to be useful for distinguishing A- and B-like characteristics of individual base pair steps.<sup>11</sup> These are the glycosidic torsion angle  $\chi$ , and the position of the phosphate group relative to the mean base pair plane, designated  $Z_p$ .<sup>32</sup> Figure 3 gives a structural view of  $Z_p$ . We normalize  $\chi$  and  $Z_p$ ,

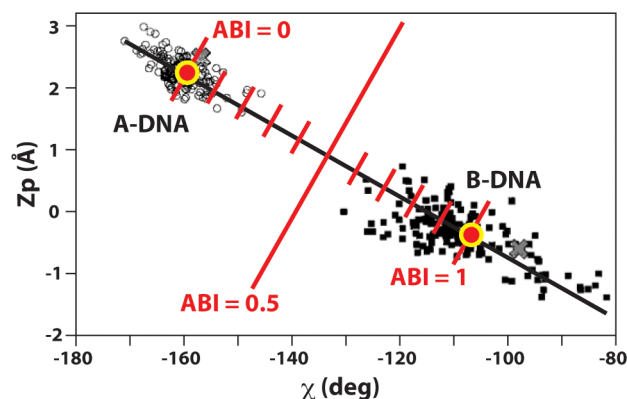


**Figure 3.** Phosphate group has different positions in A-DNA and B-DNA. This difference is made quantitative by  $Z_p$ , the average displacement of the two phosphate groups joining successive nucleotides relative to the mean base pair plane between the corresponding successive base pairs. In the A- and B-forms,  $Z_p \sim 2$  Å and  $Z_p \sim 0$ , respectively.<sup>32</sup>

relative to their differences in A-DNA and B-DNA, then use a linear combination to define the ABI for a given base pair step  $x$

$$ABI = \frac{1}{2} \left[ \frac{Z_p(x) - Z_p(A)}{Z_p(B) - Z_p(A)} + \frac{\chi(x) - \chi(A)}{\chi(B) - \chi(A)} \right] \quad (1)$$

where  $Z_p(A) = 2.2$  Å,  $Z_p(B) = -0.4$  Å,  $\chi(A) = -157^\circ$ , and  $\chi(B) = -108^\circ$ , the values for A- and B-DNA (Figure 4).



**Figure 4.** Definition of the ABI. Black points are representative conformations of individual base pair steps defined in two-dimensional ( $\chi, Z_p$ ) space, taken from high-resolution crystal structures of the A- and B-forms.<sup>11</sup> (Reprinted with permission from ref 11. Copyright 2000 Elsevier.) Ideal configurations of A- and B-forms are shown by the red circles with yellow outlines. The ABI is defined by a linear combination of normalized values of  $\chi$  and  $Z_p$  (eq 1), and the ABI scale is shown here in red. ABI values cluster around 0 and 1 for the A- and B-forms, respectively. Note that values outside the range  $0 \leq ABI \leq 1$  do occur.

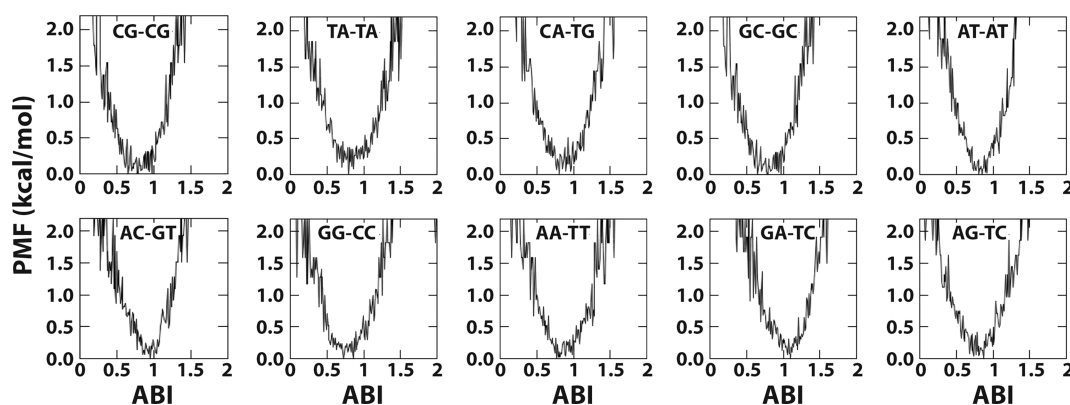
DNA helicoidal parameters and the ABI for conformations from the simulations were determined using 3DNA,<sup>33,34</sup> which can be downloaded from the 3DNA Web site at Columbia University, x3dna.org. The ABI is now incorporated into 3DNA as of version 2.3.

Any single parameter is necessarily an oversimplification of the multidimensional DNA conformational space. We will discuss the utility and limitations of the ABI below.

**Database of B-DNA Structures.** To characterize the distributions of ABI in B-DNA in solution and to search for possible sequence-dependent effects, we determined the ABI probability distributions for the 10 unique dinucleotide base pair steps, using the ABC (Ascona B-DNA Consortium) database of DNA structures, generated in multimicrosecond molecular dynamics simulations,<sup>16–19</sup> and further converged runs of the GAAC repeat 18-mer from that set.<sup>20,21</sup> The simulations that generated that database were carried out with the AMBER package,<sup>35</sup> using the parm99 force field<sup>36,37</sup> with the bsc0 modifications,<sup>38</sup> with the SPC/E water model<sup>39</sup> in a solvent containing 150 mM KCl at  $T = 300$  K.

**Calculation of Bend Angles at A/B Junctions.** We calculated bend angles using the DSSR (detecting secondary structures) functionality in 3DNA. The DNA molecule was divided into three segments, one corresponding to the restrained base pairs in the center of the molecule, and one for each of the free ends. DSSR gives the optimum vector along the axis of each fragment, and we calculated each bend angle from the dot product between the axes of the two fragments on either side of the junction.





**Figure 5.** Potentials of mean force for the A-B Index for each of the 10 unique base pair steps, derived from the ABC database of MD simulations on B-DNA and its extension.<sup>16–21</sup>

## RESULTS AND DISCUSSION

**ABI Distributions in B-DNA in Solution.** As seen in Figure 4, there are variations in the structures of base pair steps, even inside DNA crystals. These are partly because the environment of a given base pair step depends on the sequence in which it is buried, as became evident with the first crystal structure of B-DNA.<sup>40</sup> For example, the structure of the central GC step will be somewhat different in the sequences GGCC and TGCA.

The series of ABC MD simulations on B-DNA examined the sequence-dependence of DNA structure,<sup>16–21</sup> and that database offers an excellent opportunity to examine variations in the A-B Index. We have calculated the ABI probability distributions for each of the 10 unique base pair steps from that database, and these are expressed as potentials of mean force in Figure 5, using the relationship

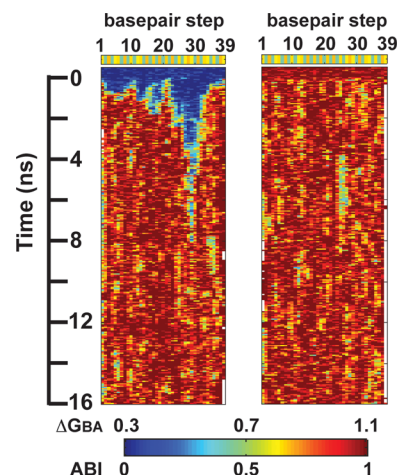
$$PMF = -RT \ln(P)$$

where  $P$  is the probability,  $R$  is the universal gas constant, and  $T = 300$  K is the temperature.

The PMFs have subtle differences. The lowest free energy states have ABIs in the range  $0.7 \leq \text{ABI} \leq 1$ , except for the GA-TC step. This step is evidently the most A-phobic with the AMBER force field under these solvent conditions with the minimum free energy state around  $\text{ABI} \sim 1.1$ . The A-phobicity of this step is also reflected by the fact that excursions of the ABI to values below 0.5 cost about 2 kcal/mol, whereas most other steps reach ABI values as low as  $\sim 0.2$ – $0.3$  before reaching that energy level. Surprisingly, there is little difference in the distributions for the GG-CC step and that of the AA-TT step, although the former is predicted to be the least A-phobic step by the Tolstorukov scale,<sup>15</sup> while the latter is the most A-phobic. This may represent a deficiency in the force field.

**The Transition from A-DNA to B-DNA.** To determine the characteristics of the  $A \Rightarrow B$  transition, we equilibrated a solution of 150 mM NaCl around a dsDNA molecule restrained in its initial conformation, then released the restraints and monitored DNA's behavior. We repeated this process for the four sequences shown in Table 1; for each sequence, we ran four simulations starting with the molecule in the A-form and four others starting from the B-form.

Figure 6 shows the ABI trajectories for a typical pair of simulations, on the MedF sequence, starting from the A-DNA and B-DNA conformations. The main panels indicate the



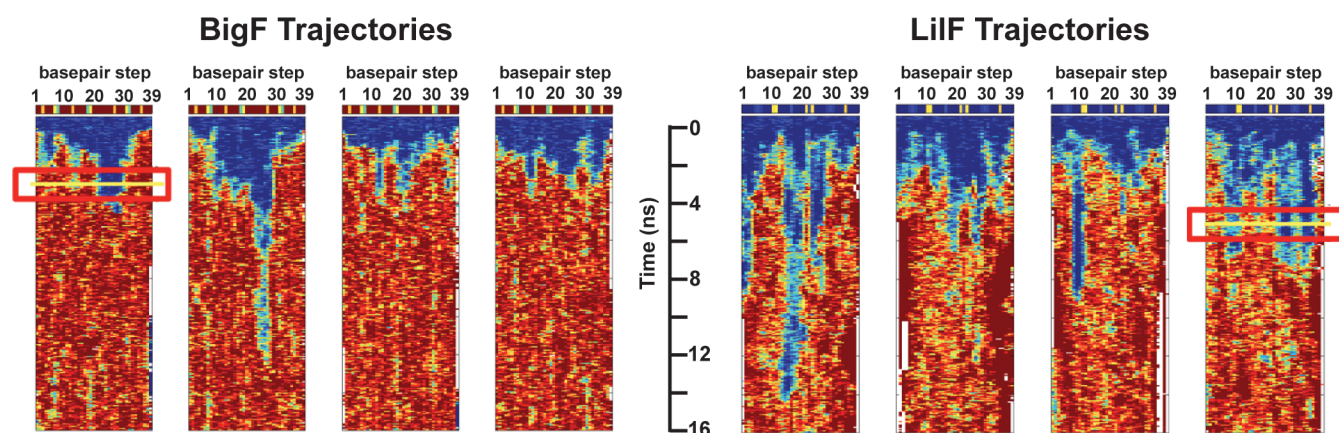
**Figure 6.** ABI trajectories for two DNA molecules, each with the MedF sequence, starting from the A-DNA conformation (left) and the B-DNA conformation (right). The bar at the bottom gives color scales for the ABI and for  $\Delta G_{BA}$  from the Tolstorukov scale.<sup>15</sup> The latter allows the reader to interpret the color bar at the top, which indicates the A-phobicity of each base pair step. As shown in Figure 2, there is little variation in the A-phobicity of the MedF sequence, so there is little variation in the colors in this bar. The DNA was equilibrated with restraints holding it in either the A- or B-form; the last 1 ns of equilibration is included in the figure, and  $t = 0$  corresponds to the point at which restraints were released.

evolution of the ABI for each base pair step, color-coded from blue ( $\text{ABI} \leq 0$ ) to red ( $\text{ABI} \geq 1$ ).

The structure that was equilibrated in the B-form maintains B-like characteristics throughout the trajectory, as seen in the overall redness of that plot. Individual base pair steps have fluctuating values of ABI, and base pair steps 26–27 flicker into a conformation with  $\text{ABI} \sim 0.5$  during the time period from  $t \sim 4$  to  $t \sim 7$  ns. According to the distributions in Figure 5, these are not unusual excursions in the value of the ABI but the fact that two steps are involved is a hint of cooperativity. All simulations on all four sequences starting from the B-form resemble this one.

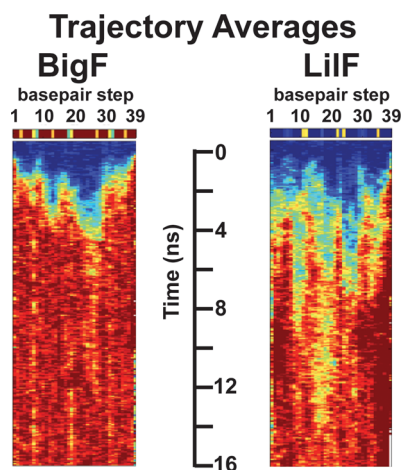
For the structure equilibrated as A-DNA, the transition to B-DNA begins within the first 1–2 ns and is complete by  $t \sim 8$  ns with steps 27–29 persisting in the A-form longer than the rest of the structure.

Do the putative A-phobicities from the Tolstorukov scale make any difference in the trajectories? Figure 7 answers that



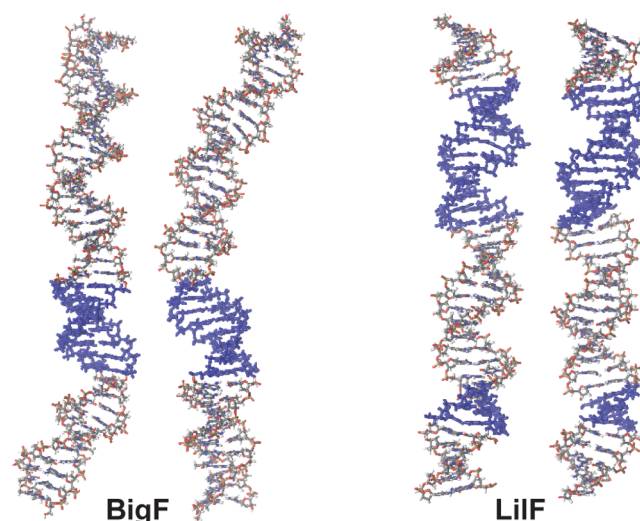
**Figure 7.** ABI trajectories for the BigF and LilF sequences, starting from the A-form. Color scales are the same as in Figure 6. A comparison of the color bars at the top of each trajectory with those for MedF (Figure 6) shows that the BigF sequence is more A-phobic (red), while the LilF sequence is less A-phobic (blue). The yellow bars and red boxes identify regions that have both A-DNA and B-DNA, and A/B junctions, as discussed below. White patches within a trajectory indicate distorted structures where the mean base pair plane is ill-defined, so the values of  $Z_p$  and ABI cannot be calculated by 3DNA.

question by comparing trajectories with the LilF sequence against those with the BigF sequence. Four trajectories are shown for each sequence, to give an idea of the variability. As in Figure 6, we see that there are no significant barriers between the A- and B-forms: transitions from A-DNA to B-DNA begin within 1–2 ns after release of the restraints, and all structures become B-form within a few nanoseconds. Some local regions of A-form DNA persist for several nanoseconds but there is considerable variability as to which regions do so. According to the Tolstorukov scale (cf. Figure 2), the BigF sequence is much more A-phobic than the LilF sequence, and there is a correlation between this prediction and the fact that the LilF sequence takes longer on average to convert to B-DNA than does the BigF sequence. This is also seen in Figure 8, which shows the average values of the ABI from the data in Figure 7, plotted as two comparative trajectories. This result suggests that the CHARMM36 force field captures the relative A-phobicities predicted by Tolstorukov et al.<sup>15</sup>



**Figure 8.** Averaged trajectories for the BigF and LilF sequences, from the data in Figure 7. Color scales are the same as in Figures 6 and 7. The relative A-phobicities of the two sequences are manifested by a more rapid conversion to B-DNA in the BigF molecule versus lingering A-like structures in the LilF molecule.

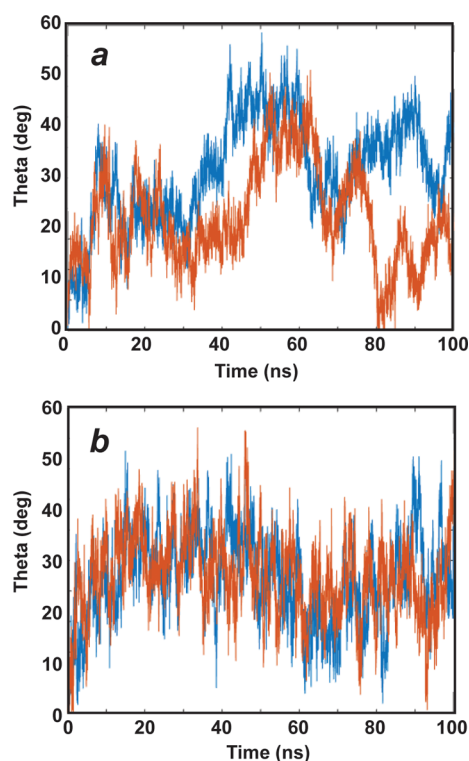
**The A/B Junction.** The persistence of local structures with low values of ABI in Figures 7 and 8 offers opportunities to look at the junction between A-DNA and B-DNA in a single molecule. Figure 9 shows the structures identified with the red boxes and yellow lines in Figure 7 with regions of low ABI values highlighted in blue.



**Figure 9.** Snapshots of junctions in BigF DNA and LilF DNA. The two views of each molecule are orthogonal to one another (rotated by 90° about the vertical axis), so that both the magnitude and direction of the bend angles at the A/B junctions can be seen. There are pronounced bends at both junctions in the BigF molecule, but the LilF molecule is essentially unbent. These structures were captured early in the transitions from A-DNA to B-DNA (red boxes and yellow lines in Figure 7).

To characterize the junction more completely, we captured the BigF structure at  $t = 3$  ns, restrained base pairs 22–29 to the A-like structure they had at that instant (the blue region in the BigF molecule in Figure 9), and extended the MD simulation over an additional 100 ns. Figure 10 shows the resulting trajectories. It takes about 10 ns for the bend angles to reach their equilibrium values. Subsequent fluctuations occur





**Figure 10.** On hundred nanosecond trajectories of the bend angles at the two A/B junctions in the BigF molecule. (a) bend angle between the A-DNA segment and the long arm of the molecule (the upper arm in Figure 9). (b) Bend angle between the A-DNA segment and the short arm of the molecule (the lower arm in Figure 9). Each 100 ns simulation was repeated twice, and the two trajectories are indicated in red and blue.

on time scales of  $\sim 1$ –50 ns with bend angles covering a broad range of roughly  $0^\circ \leq \theta \leq 50^\circ$ .

DNA bending at A/B junctions could account for a puzzling fact about the packaging of dsDNA into bacteriophages. After many years of debate over exactly how many base pairs of DNA are packaged per hydrolyzed ATP molecule, it was shown that DNA packaging proceeds in two phases. First is the “burst phase” in which ATP hydrolysis drives the packaging of 10 base pairs (bp) of DNA. The “dwell phase” follows during which five ADP molecules are replaced by five ATPs. There are four 2.5-bp steps in the 10 bp burst; during each of these, one molecule of ATP is hydrolyzed and one inorganic phosphate molecule is released.<sup>41–43</sup> The fifth ATP hydrolysis event resets the motor at the end of the dwell phase.<sup>42</sup>

Almost all models for the packaging mechanism make the logical assumption that lever-like protein motions are responsible for driving the DNA forward and propose interactions between the motor proteins and the DNA phosphate groups. If such interactions are involved, it is difficult to understand why the DNA advances in 2.5-bp steps, rather than an integral number. Our scrunchworm model proposes that DNA translocation is not due to lever-like protein motions but is instead the product of repeated transitions between B-DNA and A-DNA with these cyclic shortening/lengthening motions captured by a coupled grip-release cycle.<sup>14</sup> One strength of that model is that it offers a possible explanation for the 2.5-bp step size.

The explanation is as follows. B-DNA has a rise of 3.4 Å, while A-DNA has a rise of 2.6 Å. The 2.5-bp step size of the

packaging process thus represents an advance of  $2.5 \times 3.4 \text{ Å} = 8.5 \text{ Å}$ . In the scrunchworm model, an 8.5 Å advance requires the transition of  $8.5 \text{ Å} / (3.4 \text{ Å/bp} - 2.6 \text{ Å/bp}) \sim 10.6 \text{ bp}$  from the A-form to the B-form. This is almost exactly one turn of dsDNA, whether in the canonical B-form (10 bp/turn),<sup>2</sup> the canonical A-form (11 bp/turn),<sup>3</sup> or the 10.4 bp/turn measured in solution.<sup>44</sup> This suggests that the  $B \rightleftharpoons A$  transition is cooperative and raises the question, why is the cooperative unit one turn of the double helix? We argued that if the two A/B junctions at the end of the cooperative unit are each bent by an angle  $\theta$ , there will be minimum steric hindrance in a narrow channel if the two bends are separated by an integral number of helical turns.<sup>14</sup> (In this configuration, the two B-DNA arms are trans to one another, making them parallel, whereas at half-odd integral turns, the two arms will be cis and the angle between them will be  $2\theta$ .) At the time we made this suggestion, there were no data on the magnitude of the angle at A/B junctions, but the data in Figure 10 suggest that with  $\theta \sim 20^\circ$  or more a cooperative unit of one turn of the double helix would be strongly favored under the foregoing argument about steric hindrance.

**Utility and Limitations of the ABI.** Describing a multidimensional conformational space by a single parameter has one great advantage: it simplifies things. But it also has risks: it oversimplifies things. To truly understand any particular dsDNA structure, one must examine it with an analysis tool like 3DNA,<sup>33,34</sup> Curves+,<sup>45</sup> or some other program that determines the helicoidal parameters,<sup>46</sup> using the standard coordinate frame and structural conventions.<sup>5</sup>

The utility of the ABI is demonstrated by the equilibrium distributions for different base pair steps (Figure 5) and by the simplicity of the chromograms monitoring the  $A \Rightarrow B$  transition (Figures 6–8). To allow the reader to critically assess the correlations between the ABI and the other parameters commonly used to distinguish the A- and B-forms, the Supporting Information includes complete 3DNA analyses of the two structures in Figure 9.

## CONCLUSIONS

We have examined the transitions of DNA molecules from the A-DNA conformation to the B-DNA conformation in MD simulations using a B-genic solvent (150 mM NaCl). We investigated four different sequences with a range of predicted A-phobicities, using a new metric, the A-B Index (ABI) to describe the structures. In all cases, there is no energy barrier to the initiation of the A-to-B transition. This is consistent with the positive values of  $\Delta G_{BA}$  reported by Tolstorukov et al.<sup>15</sup> The conversion begins within about a nanosecond and is typically complete within 5 ns, although A-like structures sometimes persist in regions of 3–5 base pairs for up to 10 ns more. Molecules containing sequences that are predicted to be more A-phobic<sup>15</sup> make the transition more rapidly than those containing less A-phobic sequences.

These simulations, along with an analysis of the extended ABC database,<sup>16–21</sup> show that even under physiological conditions there are substantial fluctuations in the ABI about the average values deduced from X-ray crystallography. This is not surprising, considering the modest free energy differences between the A- and B-DNA conformations for individual base pair steps.<sup>15</sup>

When a segment of about one turn of the double helix is locked in the A-form, MD simulations in the same B-genic solvent reveal that the bend angle at the A/B junction has an

average value of  $\sim 20\text{--}30^\circ$ , and that it fluctuates with a characteristic time on the order of 10 ns. The size of these bend angles offers some support for our previous proposal<sup>14</sup> that the cooperative unit for transitions between A-DNA and B-DNA in a narrow channel is one turn of the double helix.

The A-form of the DNA double helix, which was long regarded as little more than a laboratory artifact, does occur locally in DNA–protein complexes.<sup>12,13</sup> It has been found to play a protective role under desiccating conditions, it is the form adopted by the genome of a filamentous virus,<sup>12</sup> and it may play a critical role in the generation of force by viral dsDNA packaging motors.<sup>14</sup> The simulations reported here expand our understanding of this increasingly important DNA conformation.

There are, however, important unanswered questions about the forces driving transitions between the A- and B-forms and how these depend on solvent conditions and DNA sequence. Jayaram et al. gave an interesting breakdown of the differential free energy contributions from phosphate–phosphate repulsions, solvation, and counterion condensation, and how each of these depends on DNA conformation (A vs B) in mixed water–ethanol solvents.<sup>47</sup> They showed that in aqueous NaCl the favorable counterion solvation favors the B-form whereas at low water activity counterion–DNA interactions dominate, favoring the A-form. Our simulations have focused on kinetic issues, so they are not sufficiently long to provide the well-converged equilibrium distributions necessary for decomposing the factors driving the transition. It would be interesting to revisit this question with longer MD simulations on the sequences used in our study.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b02155.

Analyses on the two structures in Figure 9, including the values of the ABI and all other helicoidal parameters for each base pair step, to enable the reader to evaluate the ability of the ABI to identify A-like and B-like conformations. (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: steharv@mail.med.upenn.edu. Telephone: 404-444-3551.

### Author Contributions

S.C.H. proposed the project and oversaw the research. J.T.W. had primary responsibility for implementation and quantitative analysis of the new simulations, whereas J.T.W., H.D.K., J.C.G., and S.C.H. all contributed to their design and interpretation. X.-J.L. incorporated the definition of the A-B Index into 3DNA and contributed to the analysis of the new simulations. R.G.-M. and T.E.C. carried out the analysis of the data from the ABC simulations, provided Figure 5, and contributed to the interpretation of the new simulations. S.C.H. had primary responsibility for writing the manuscript but all authors contributed to it.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Supported by grants from the National Institutes of Health (R01-GM096889 to X.-J. L., T01-GM112882 to H.D.K., and R01-GM081411 to T.E.C.) and by a National Science Foundation CAREER award (MCB-1452464 to J.C.G.). The new MD simulations were carried out on the TACC Stampede system at the University of Texas, Austin, supported by NSF award OCI-1053575. The ABC MD simulations utilized computational resources from the Blue Waters sustained-petascale computing project (NSF OCI 07-25070 and PRAC OCI-1036208).

## ■ ABBREVIATIONS:

ABC, Ascona B-DNA Consortium; ABI, A-B Index; bp, base pair; dsDNA, double-stranded DNA; MD, molecular dynamics; PME, particle mesh Ewald

## ■ REFERENCES

- (1) Franklin, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171*, 740–741.
- (2) Watson, J. D.; Crick, F. H. C. A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738.
- (3) Franklin, R. E.; Gosling, R. G. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* **1953**, *172*, 156–157.
- (4) Dickerson, R. E. Definitions and Nomenclature of Nucleic Acid Structure Components. *Nucleic Acids Res.* **1989**, *17*, 1797–1803.
- (5) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; et al. A Standard Reference Frame for the Description of Nucleic Acid Base-Pair Geometry. *J. Mol. Biol.* **2001**, *313*, 229–237.
- (6) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double-Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, *282*, 680–686.
- (7) Zhang, S. G.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A. Zuoitin, a Putative Z-DNA Binding Protein in *Saccharomyces Cerevisiae*. *EMBO J.* **1992**, *11*, 3787–3796.
- (8) Kim, Y. G.; Lowenhaupt, K.; Oh, D. B.; Kim, K. K.; Rich, A. Evidence That Vaccinia Virulence Factor E3L Binds to Z-DNA in Vivo: Implications for Development of a Therapy for Poxvirus Infection. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1514–1518.
- (9) Kim, Y. G.; Muralinath, M.; Brandt, T.; Percy, M.; Hauns, K.; Lowenhaupt, K.; Jacobs, B. L.; Rich, A. A Role for Z-DNA Binding in Vaccinia Virus Pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 6974–6979.
- (10) Mohr, S. C.; Sokolov, N. V.; He, C. M.; Setlow, P. Binding of Small Acid-Soluble Spore Proteins from *Bacillus Subtilis* Changes the Conformation of DNA from B to A. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 77–81.
- (11) Lu, X. J.; Shakked, Z.; Olson, W. K. A-Form Conformational Motifs in Ligand-Bound DNA Structures. *J. Mol. Biol.* **2000**, *300*, 819–840.
- (12) DiMaio, F.; Yu, X.; Rensen, E.; Krupovic, M.; Prangishvili, D.; Egelman, E. H. Virology. A Virus That Infects a Hyperthermophile Encapsidates a-Form DNA. *Science* **2015**, *348*, 914–917.
- (13) Whelan, D. R.; Hiscox, T. J.; Rood, J. I.; Bambery, K. R.; McNaughton, D.; Wood, B. R. Detection of an En Masse and Reversible B- to a-DNA Conformational Transition in Prokaryotes in Response to Desiccation. *J. R. Soc., Interface* **2014**, *11*, 20140454.
- (14) Harvey, S. C. The Scrunchworm Hypothesis: Transitions between a-DNA and B-DNA Provide the Driving Force for Genome Packaging in Double-Stranded DNA Bacteriophages. *J. Struct. Biol.* **2015**, *189*, 1–8.
- (15) Tolstorukov, M. Y.; Ivanov, V. I.; Malenkov, G. G.; Jernigan, R. L.; Zhurkin, V. B. Sequence-Dependent Ba Transition in DNA

Evaluated with Dimeric and Trimeric Scales. *Biophys. J.* **2001**, *81*, 3409–3421.

(16) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E., 3rd; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; et al. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on D(Cpg) Steps. *Biophys. J.* **2004**, *87*, 3799–3813.

(17) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; et al. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. II. Sequence Contact Effects on the Dynamical Structures of the 10 Unique Dinucleotide Steps. *Biophys. J.* **2005**, *89*, 3721.

(18) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. 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.

(19) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., 3rd; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. Muabc: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272–12283.

(20) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., 3rd On the Absence of Intrahelical DNA Dynamics on the Mus to Ms Timescale. *Nat. Commun.* **2014**, *5*, 5152.

(21) Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., 3rd Convergence and Reproducibility in Molecular Dynamics Simulations of the DNA Duplex D(Gcagcaacgaacgacgc). *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1041–1058.

(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(23) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with Namd. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(24) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(1) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(25) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; Mackerell, A. D., Jr. Optimization of the Charmm Additive Force Field for DNA: Improved Treatment of the Bi/Bii Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348–362.

(26) Ryckaert, J.-P.; Ciccotti, G. 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.

(27) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(28) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193–4194.

(29) Zheng, G.; Lu, X. J.; Olson, W. K. Web 3dna—a Web Server for the Analysis, Reconstruction, and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nucleic Acids Res.* **2009**, *37*, W240–246.

(30) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.

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

(32) El Hassan, M. A.; Calladine, C. R. Conformational Characteristics of DNA: Empirical Classifications and a Hypothesis for the Conformational Behaviour of Dinucleotide Steps. *Philos. Trans. R. Soc., A* **1997**, *355*, 43–100.

(33) 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*, S108–S121.

(34) Lu, X. J.; Olson, W. K. 3dna: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3*, 1213–1227.

(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. R.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. Amber: A Computer Program for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Elucidate the Structures and Energies of Molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(36) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(37) Cheatham, T. E., 3rd; Cieplak, P.; Kollman, P. A. A Modified Version of the Cornell Et Al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862.

(38) Perez, A.; Luque, F. J.; Orozco, M. Dynamics of B-DNA on the Microsecond Time Scale. *J. Am. Chem. Soc.* **2007**, *129*, 14739–14745.

(39) Berendsen, H. J. C.; Grigera, J. R. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(40) Dickerson, R. E.; Drew, H. R. Structure of a B-DNA Dodecamer. II. Influence of Base Sequence on Helix Structure. *J. Mol. Biol.* **1981**, *149*, 761–786.

(41) Moffitt, J. R.; Chemla, Y. R.; Athavan, K.; Grimes, S.; Jardine, P. J.; Anderson, D. L.; Bustamante, C. Intersubunit Coordination in a Homomeric Ring ATPase. *Nature* **2009**, *457*, 446–450.

(42) Chistol, G.; Liu, S.; Hetherington, C. L.; Moffitt, J. R.; Grimes, S.; Jardine, P. J.; Bustamante, C. High Degree of Coordination and Division of Labor among Subunits in a Homomeric Ring ATPase. *Cell* **2012**, *151*, 1017–1028.

(43) Liu, S.; Chistol, G.; Hetherington, C. L.; Tafoya, S.; Athavan, K.; Schnitzbauer, J.; Grimes, S.; Jardine, P. J.; Bustamante, C. A Viral Packaging Motor Varies Its DNA Rotation and Step Size to Preserve Subunit Coordination as the Capsid Fills. *Cell* **2014**, *157*, 702–713.

(44) Wang, J. C. Helical Repeat of DNA in Solution. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 200–203.

(45) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. Curves+ Web Server for Analyzing and Visualizing the Helical, Backbone and Groove Parameters of Nucleic Acid Structures. *Nucleic Acids Res.* **2011**, *39*, W68–73.

(46) Lavery, R.; Sklenar, H. The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.

(47) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. Free Energy Analysis of the Conformational Preferences of A and B Forms of DNA in Solution. *J. Am. Chem. Soc.* **1998**, *120*, 10629–10633.