THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY



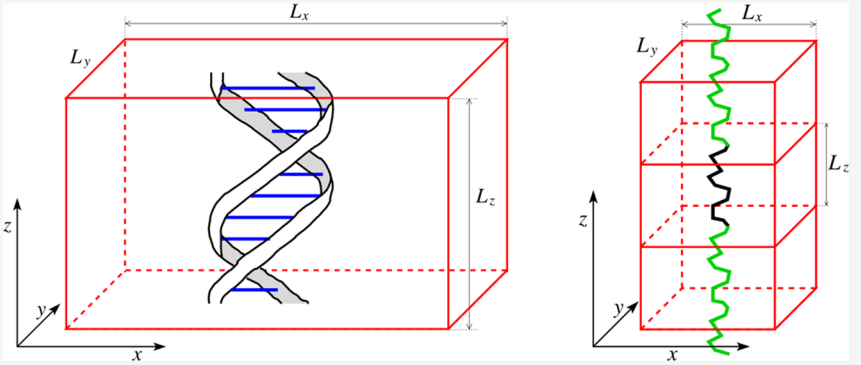
pubs.acs.org/JPCB
Article

# Asymmetric Periodic Boundary Conditions for All-Atom Molecular Dynamics and Coarse-Grained Simulations of Nucleic Acids

Radek Erban* and Yuichi Togashi*

Cite This: J. Phys. Chem. B 2023, 127, 8257−8267






**ABSTRACT:** Periodic boundary conditions are commonly applied in molecular dynamics simulations in the microcanonical (NVE), canonical (NVT), and isothermal−isobaric (NpT) ensembles. In their simplest application, a biological system of interest is placed in the middle of a solvation box, which is chosen 'sufficiently large' to minimize any numerical artifacts associated with the periodic boundary conditions. This practical approach brings limitations to the size of biological systems that can be simulated. Here, we study simulations of effectively infinitely long nucleic acids, which are solvated in the directions perpendicular to the polymer chain, while periodic boundary conditions are also applied along the polymer chain. We study the effects of these asymmetric periodic boundary conditions (APBC) on the simulated results, including the mechanical properties of biopolymers and the properties of the surrounding solvent. To get some further insights into the advantages of using the APBC, a coarse-grained worm-like chain model is first studied, illustrating how the persistence length can be extracted from the local properties of the polymer chain, which are less affected by the APBC than some global averages. This is followed by all-atom molecular dynamics simulations of DNA in ionic solutions, where we use the APBC to investigate sequence-dependent properties of DNA molecules and properties of the surrounding solvent.

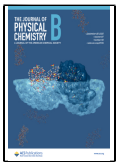


## 1. INTRODUCTION

The structure and function of DNA depend on its nucleotide sequence and on the properties of the surrounding solvent.[1] Since DNA is negatively charged, concentrations of ions are perturbed from their bulk values in the region close to DNA. The resulting 'ion atmosphere' has been studied using ion counting experiments.[2] From the theoretical point of view, all-atom molecular dynamics (MD) simulations can be applied to provide detailed insights into DNA, ions, and water interactions.[3] For example, the effect of mobile counterions, $Na^+$ and $K^+$, on a DNA oligomer was studied by Várnai and Zakrzewska,[4] who used periodic boundary conditions for MD simulations of the solvated DNA oligomer at constant temperature and pressure and studied the counterion distribution around the DNA structure.

However, the applicability of all-atom MD studies is limited to relatively small systems. To simulate larger systems, several coarse-grained approaches have been developed in the literature. In adaptive resolution studies,[5,6] DNA and its immediate neighborhood are simulated using the full atomistic resolution, while a coarse-grained description is used to describe the solvent molecules that are far away from DNA. Solvent can also be treated implicitly in faraway regions.[6] To model even larger systems, the DNA molecule itself can be described by coarse-grained models.[7−10] Examples vary from models using several coarse-grained sites per nucleotide[11] to Brownian dynamics simulations.[12,13] Using a systematic 'bottom-up approach', the interaction potential between coarse-grained sites can be

Received: June 8, 2023
Revised: July 31, 2023
Published: September 15, 2023







https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

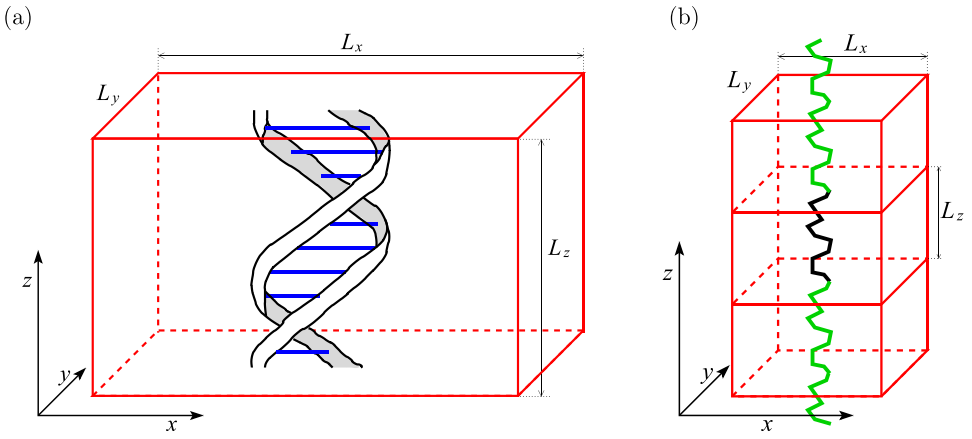



**Figure 1.** (a) Schematic of the simulation domain given by eq 1 for the case with 10 base pairs of DNA, i.e., for $n = 1$. (b) Discrete worm-like chain segment in an APBC simulation is denoted by the black line. Its periodic image copies are visualized as the green lines.

derived from the underlying atomistic force field, with results dependent on the microscopic force field used.[14]

The properties of ions in bulk water can be studied using all-atom MD simulations,[15,16] which provide 'bottom-up' estimates of the values of diffusion constants of ions. Some biological processes include transport of ions across relatively large distances which are out of reach to all-atom MD simulations. Brownian dynamics descriptions of ions are instead used for modeling such systems.[17,18] While the transport of ions in bulk water can be described on a sufficiently long time scale as a standard Brownian motion, more detailed coarse-grained stochastic models of ions in bulk water have to be used at time scales studied by MD simulations.[16] Coarse-grained stochastic models of ions can be written as systems of stochastic differential equations or by the generalized Langevin equation.[16,19] To parametrize such models, detailed all-atom MD simulations of 'long' chains of nucleic acids can be used, but this can be computationally intensive. In this paper, we investigate simulations of effectively infinitely long DNA by applying asymmetric periodic boundary conditions (APBC) in the cuboid computational domain

$$\Omega = [0, L_x] \times [0, L_y] \times [0, L_z] \tag{1}$$

The main idea behind the APBC is that DNA is periodic with the period of 10 base pairs, i.e., the APBC will allow us to use $10n$ base pairs of DNA in domain $\Omega$, where $n \in \mathbb{N}$ is an integer denoting the number of helical pitches. A schematic of our simulation domain is presented in Figure 1a for the case of the simulation with 10 base pairs, i.e., for $n = 1$. The DNA molecule is positioned parallel to the $z$-axis, and we use periodic boundary conditions in the $z$-direction. Such a periodic boundary condition in the $z$-direction is less common in all-atom MD simulation studies, where the biomolecule of interest is often placed in the middle of the computational domain and it is solvated on all its sides by a layer of water molecules separating the biomolecule from the domain boundary.[4,20]

Considering the projection into the $xy$-plane, the DNA molecule is positioned in the middle of the simulated domain. In particular, the DNA molecule is separated by the layer of water molecules from the boundaries of the simulated domain in both the $x$-direction and $y$-direction. While we use periodic boundary conditions in all three directions, there is an asymmetry (highlighted in our terminology APBC): a modeler has a relative freedom to choose the values of $L_x$ and $L_y$ in the computational domain defined by eq 1, while the value of $L_z$ is dictated by the properties of the simulated biomolecule. The imposed DNA periodicity fixes the helical twist of the DNA molecule with the simulation box size $L_z$ chosen such that it exactly corresponds to $n$ helical pitches. However, considering simulations at isothermal−isobaric (NpT) ensemble, standard isotropic barostats introduce fluctuations in the domain size leading to changes in $L_z$ as well. To fix $L_z$, an asymmetric barostat is used in Section 3.

The APBC has been used in previous studies[5,21,22] to simulate effectively infinitely long DNA molecules. Except for the asymmetry between the $z$-direction and $x$-direction (resp. $y$-direction), the APBC can lead to a relatively standard all-atom MD setup, with the domain periodic in all three directions, which was previously used to explore the ion atmosphere around the DNA.[21,22] However, it is more challenging to use the APBC to study the mechanical properties of biopolymers, as we will first illustrate in Section 2 by considering a discrete worm-like chain model. This is followed by all-atom MD simulations of DNA in Section 3, where we present the use of APBC to investigate the mechanical properties of the DNA and the properties of the surrounding solvent.

## 2. WORM-LIKE CHAIN MODEL

Let us consider the discrete worm-like chain (WLC) model where DNA consists of $N$ segments $\mathbf{l}_i$, $i = 1, 2, ..., N$, each having the same length, $\delta$. Denoting the angle between the adjacent $i$th and $(i + 1)$th segments by $\theta_i$, for $i = 1, 2, ..., (N - 1)$, the chain bending energy of the discrete WLC model is

$$\frac{E}{k_B T} = \alpha \sum_{i=1}^{N-1} \theta_i^2 \tag{2}$$

where $\alpha$ is a dimensionless constant. The sum on the right-hand side of eq 2 corresponds to an integral term in the continuous WLC model,[23,24] and eq 2 can be also derived by discretizing this integral term.[1] Then, one can also establish that our dimensionless stiffness parameter $\alpha$ is proportional to the persistence length of the continuous WLC model.[24] This is also true for the discrete WLC model in the limit $\alpha \to \infty$, where we get the persistence length equal to $2\delta\alpha$ to the leading order (see eq 8). To show this, we define the persistence length of the first $j$th segments of the discrete WLC model, for $j \leq N$, by


https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

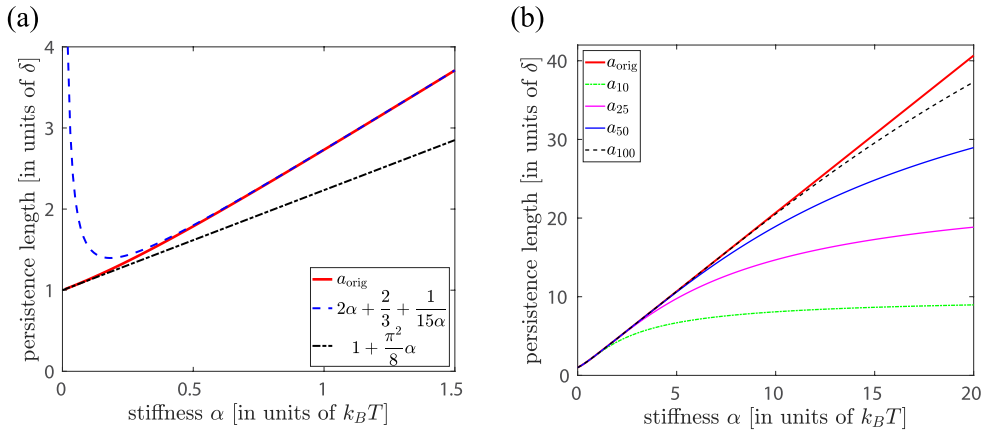



**Figure 2.** (a) Plot of persistence length $a_{orig}$, given by eq 4, as a function of the stiffness parameter $\alpha$, together with asymptotic results given by eqs 8 and 9. The dimensionless parameter $\alpha$ can be viewed to express energy in units $k_BT$, while all persistence lengths are plotted in units of the segment length, $\delta$. (b) Plot of persistence length $a_{orig}$, given by eq 4, and persistence lengths $a_j$, given by eq 5, for $j$ = 10, 25, 50, 100, as a function of parameter $\alpha$.

$$a_j = \left\langle \frac{\mathbf{l}_1}{\delta} \cdot \sum_{i=1}^{j} \mathbf{l}_i \right\rangle \tag{3}$$

That is, $a_j$ is the average value of the projection of the vector connecting the end points of the first and the $j$th segment on the direction of the first segment. Then, the persistence length of the discrete WLC model can be defined as the limit

$$a_{orig} = \lim_{N\to\infty} a_N \tag{4}$$

which effectively is the average value of the projection of the end-to-end vector of a long chain in the direction of the first segment. In this paper, we will only study the discrete version of the WLC model because it will provide more insight into the use of the APBC than the continuous model. In particular, in the following, the WLC model means the discrete WLC model with the energy defined by eq 2.

**2.1. Dependence of Persistence Length on Stiffness Parameter $\alpha$.** The average in eq 3 can be evaluated as

$$a_j = \sum_{i=1}^{j} \frac{\langle \mathbf{l}_1 \cdot \mathbf{l}_i \rangle}{\delta} = \delta \sum_{i=1}^{j} \langle \cos(\theta) \rangle^{i-1} = \delta \frac{1 - \langle \cos(\theta) \rangle^j}{1 - \langle \cos(\theta) \rangle} \tag{5}$$

where the average $\langle \cos(\theta) \rangle = \langle \cos(\theta_i) \rangle$ is independent of the index $i$ because the chain bending energy in eq 2 is symmetric with respect to any permutation of indices. Using spherical polar coordinates,[25] we have

$$\langle \cos(\theta) \rangle = \frac{\int_0^\pi \cos(\theta)\sin(\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta}{\int_0^\pi \sin(\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta} \tag{6}$$

To get formula 6, we note that the distribution of angles between adjacent segments is proportional to $\sin(\theta)\exp[-\alpha\theta^2]$. Using eqs 4, 5, and 6, we deduce (see Appendix A.1)

$$a_{orig} = \frac{\delta}{1 - \langle \cos(\theta) \rangle} \tag{7}$$

$$= \delta\left(2\alpha + \frac{2}{3} + \frac{1}{15\alpha} + O\left(\frac{1}{\alpha^2}\right)\right), \text{ as } \alpha \to \infty \tag{8}$$

$$= \delta\left(1 + \frac{\pi^2}{8}\alpha + O(\alpha^2)\right), \text{ as } \alpha \to 0 \tag{9}$$

In Figure 2a, we present how the persistence length $a_{orig}$ depends on the stiffness parameter $\alpha$ in interval $[0, 3/2]$, illustrating the accuracy of both expansions in eqs 8 and 9. While eq 8 is derived in the limit $\alpha \to \infty$, it approximates the exact result well for persistence lengths satisfying $a_{orig} > 2\delta$ or equivalently for $\alpha > 0.62$. In Figure 2b, we plot the dependence of the persistence length $a_{orig}$ on the stiffness parameter $\alpha$ in a larger interval $[0, 20]$ together with the values of $a_j$ given by eq 5. Using the exact result for $a_{orig}$ given in eq 7, we can rewrite eq 5 as follows

$$a_j = a_{orig}(1 - \langle \cos(\theta) \rangle^j) \tag{10}$$

Considering the limit $\alpha \to \infty$ in eq 6, we have

$$\langle \cos(\theta) \rangle = 1 - \frac{1}{2\alpha} + \frac{1}{6\alpha^2} + O\left(\frac{1}{\alpha^3}\right), \text{ as } \alpha \to \infty$$

where the first three terms of the expansion on the right-hand side provide an approximation of $\langle \cos(\theta) \rangle$ with about 5% relative error for $\alpha > 1$, and the relative error decreases as we increase the value of $\alpha$, for example, the relative error is smaller than 1% for $\alpha > 2$. Substituting this expansion for $\langle \cos(\theta) \rangle$ into eq 10, we obtain that for sufficiently large values of $\alpha$, say for $\alpha > 1$, we can calculate the persistence length $a_{orig}$ from $a_j$ by using the following formula

$$a_{orig} = \frac{a_j}{1 - \left(1 - \frac{1}{2\alpha} + \frac{1}{6\alpha^2}\right)^j} \tag{11}$$

**2.2. Dependence of Persistence Length on APBC.** Considering that the polymer chain is simulated in the domain given by eq 1 with APBC, we have an extra constraint

$$\sum_{i=1}^{N} \mathbf{l}_i = [0,\ 0,\ L_z] \tag{12}$$

where $N$ denotes the number of simulated segments along the $z$-direction. As illustrated in Figure 1b, such a model can be viewed as a model of an (infinitely) long polymer chain by using the periodicity

$$\mathbf{l}_i = \mathbf{l}_{i \bmod N}, \quad \text{for } i \in \mathbb{N} \tag{13}$$

However, substituting eqs 12 and 13 into the definition of persistence length in eq 4, we would obtain that $a_{orig} = \infty$ because the periodic boundary means that the infinitely long



https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

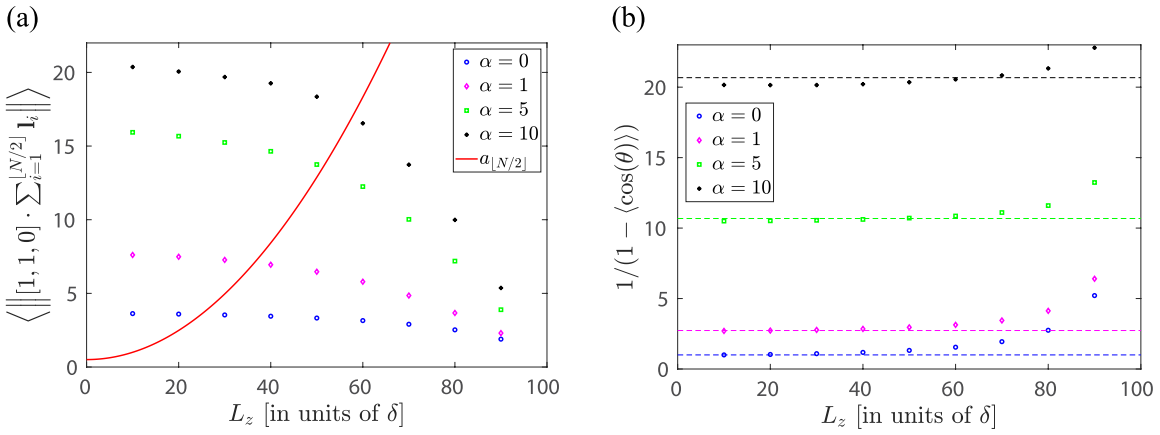



**Figure 3.** (a) Average distance of the polymer middle point from the axis of the polymer, defined by eq 14, estimated from Monte Carlo simulations of the WLC model with $N = 100$ segments for $L_z \in \{10, 20, ..., 90\}$ and $\alpha \in \{0, 1, 5, 10\}$. The red line shows $a_{\lfloor N/2 \rfloor} = a_{50}$ as a function of the domain length $L_z$ (theoretical result given by eq 15 confirmed by simulations for all considered values of $\alpha$). Further details of Monte Carlo simulations are in Appendix A.3. (b) The estimate of $a_{\text{orig}}$ given by eq 16, where $\langle \cos(\theta) \rangle$ is estimated from Monte Carlo simulations of the WLC model with $N = 100$ segments for $L_z \in \{10, 20, ..., 90\}$ and $\alpha \in \{0, 1, 5, 10\}$. The theoretical result (without APBC, independent of $L_z$), given by eq 6, is plotted by dashed lines for each value of $\alpha$.

filament is effectively straight. Since eq 12 postulates that the vector connecting ends of $N$ segments is fixed, we obtain the most variability in this model by looking at the behavior of the $\lfloor N/2 \rfloor$ consecutive segments. Due to the symmetry of the problem and eq 12, the average of the vector $\sum_{i=1}^{\lfloor N/2 \rfloor} \mathbf{l}_i$ is equal to $[0, 0, L_z/2]$ for any value of $\alpha$, but the deviations from this average will be larger when we increase the stiffness parameter $\alpha$. To illustrate this, we define the average distance of the polymer middle point from the axis of the polymer by

$$\left\langle \left\| [1, 1, 0] \cdot \sum_{i=1}^{\lfloor N/2 \rfloor} \mathbf{l}_i \right\| \right\rangle \tag{14}$$

that is, we calculate the (Euclidean) norm of the projection of the vector $\sum_{i=1}^{\lfloor N/2 \rfloor} \mathbf{l}_i$ on the $x$−$y$ plane. The average in eq 14 is plotted in Figure 3a for different values of parameter $\alpha$ and domain length $L_z$. We observe that, for a fixed value of $L_z$, the average in eq 14 increases with the value of the stiffness parameter $\alpha$. Moreover, Figure 3a also shows that the value of the average in eq 14 approaches zero as $L_z$ approaches its maximum possible value, $N\delta$. Indeed, if $L_z = N\delta$, the polymer is straight and the value of the average in eq 14 is exactly equal to zero. On the other hand, if $L_z$ is smaller, then we obtain a larger value of the average in eq 14, especially for polymers with larger persistence lengths (i.e., for large values of $\alpha$).

On the face of it, one possible way to estimate $a_{\text{orig}}$ could be to estimate $a_{\lfloor N/2 \rfloor}$ from our Monte Carlo simulations[26] and then use formula 11 for $j = \lfloor N/2 \rfloor$. However, formula 11 has been derived for the case of the WLC model in the three-dimensional physical space $\mathbb{R}^3$. Considering the APBC, we obtain that $a_{\lfloor N/2 \rfloor}$ is independent of $\alpha$ (see Appendix A.2). We have

$$a_{\lfloor N/2 \rfloor} = \delta \left( \frac{N - \lfloor N/2 \rfloor}{N - 1} + \frac{L_z^2 (\lfloor N/2 \rfloor - 1)}{N(N-1)\delta^2} \right) \tag{15}$$

which simplifies to $a_{\lfloor N/2 \rfloor} \approx (\delta/2) + L_z^2/(2\delta N)$ for large values of $N$. This result is also visualized in Figure 3a. In particular, a better strategy to obtain the real persistence length $a_{\text{orig}}$ from the APBC simulations is to estimate

$$\langle \mathbf{l}_i \cdot \mathbf{l}_{i+1} \rangle = \langle \cos(\theta) \rangle$$

and then use the exact result for $a_{\text{orig}}$ given in eq 7, namely

$$a_{\text{orig}} = \frac{\delta}{1 - \langle \cos(\theta) \rangle} \tag{16}$$

The results are presented in Figure 3b. We observe that the results given by eq 16, where $\langle \cos(\theta) \rangle$ is estimated from Monte Carlo simulations (with APBC), agree well (especially for smaller values of $L_z$) with the theoretical result, which has been obtained without the use of APBC in eq 6. As we increase $L_z$, the error of this estimation increases. In fact, if $L_z = N\delta = 100\ \delta$, then the WLC polymer becomes a straight line, giving $\langle \cos(\theta) \rangle = 1$ and formula 16 diverges, as we can also observe in Figure 3b for larger values of $L_z$.

## 3. APBC IN ALL-ATOM MD SIMULATIONS

In this section, we investigate the use of APBC in all-atom MD models of DNA. Our simulations are performed with 10−100 base pairs (bp) of double-stranded DNA (dsDNA). Since we use APBC, all simulations are effectively simulating (infinitely) long DNA chains. In particular, MD results with the longest simulated chain (100 bp) can be used as the 'ground truth' for the presented APBC simulations with shorter 10−50 bp long DNA chains. We note that the MD simulations of relatively short 50 bp DNA segments without APBC have been previously used in the literature to estimate the DNA persistence length by using a middle section of the simulated DNA segment.[20,27]

We consider 6 types of (infinitely) long DNA sequences, with repeated nucleotides, namely, poly(A), poly(C), poly(AT), poly(CG), poly(AC), and poly(AG), where poly(X) means that the corresponding nucleotide sequence is periodically repeated. We note that these 6 cases correspond to all possible cases of pairs of nucleotides which are repeated infinitely many times. For example, repetitions of dinucleotides AC, CA, TG, and GT all correspond to the poly(AC) case because AC and CA are equivalent due to the periodic boundary conditions along the chain length, and TG is on the complementary strand, with GT being equivalent to TG because of the periodic boundary conditions.



https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

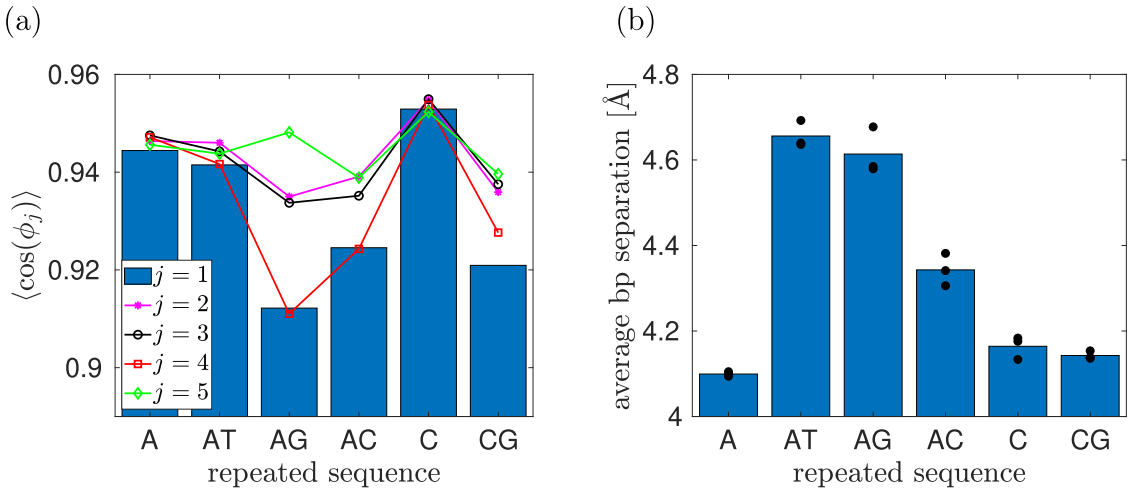



**Figure 4.** Results of all-atom MD simulations of DNA chains with APBC and $N = 10$. (a) The average given by eq 18 for each of the 6 considered sequences of repeated nucleotides is calculated using three independent MD simulations. (b) The average separation between the base pairs calculated using three independent MD simulations (blue bars). The results for each individual realization are plotted as a black dot.

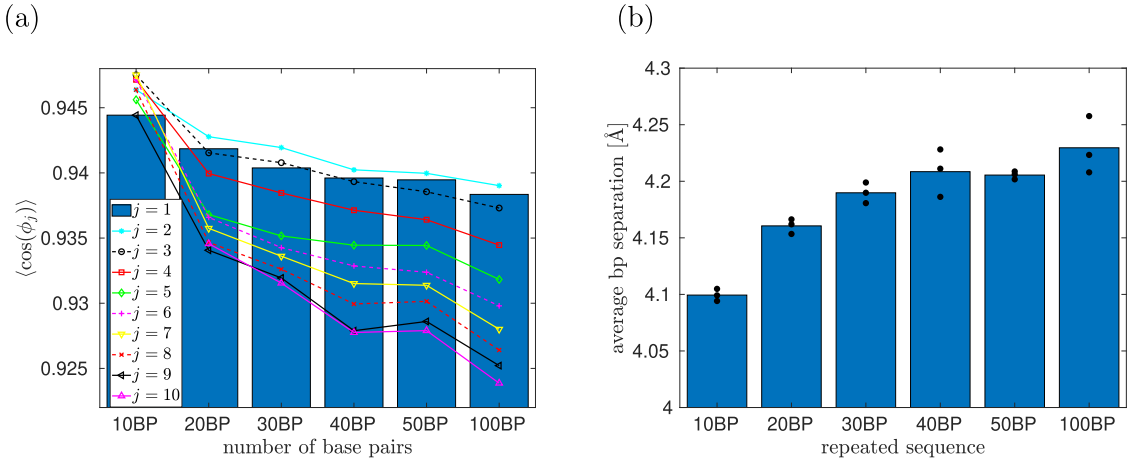



**Figure 5.** Results of all-atom MD simulations with APBC using the poly(A) DNA chain with $N$ in the range 10−100 bp. (a) The average given by eq 18 for each of the 6 values of $N$ considered is calculated using three independent MD simulations. The results are presented for $j = 1, 2, ..., 10$ and $n = 1, 2, 3, 4, 5, 10$. (b) The average separation between the base pairs calculated using three independent MD simulations (blue bars). The results for each individual realization are plotted as a black dot.

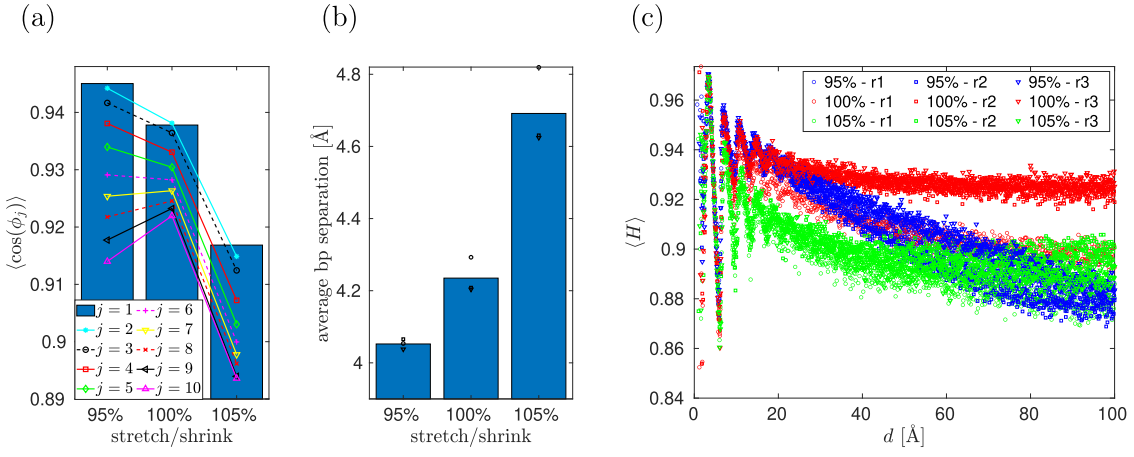



**Figure 6.** Results of MD simulations of 100 bp poly(A) dsDNA with APBC that use the values of $L_z$ given by eq 19. We average over three independent MD time series for each of the presented case. (a) The average in eq 18 for $j = 1, 2, ..., 10$. (b) The average separation of base pairs (blue bars). Dots include the results for individual MD realizations (i.e., we have averaged over the dots to calculate blue bars). (c) The average $\langle H\rangle$ as a function of distance $d$. We present results for the 95% (blue), 100% (red), and 105% (green) cases using different colors. Different symbols (circle, square, triangle of the same color) denote data points calculated by different MD realizations.



https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

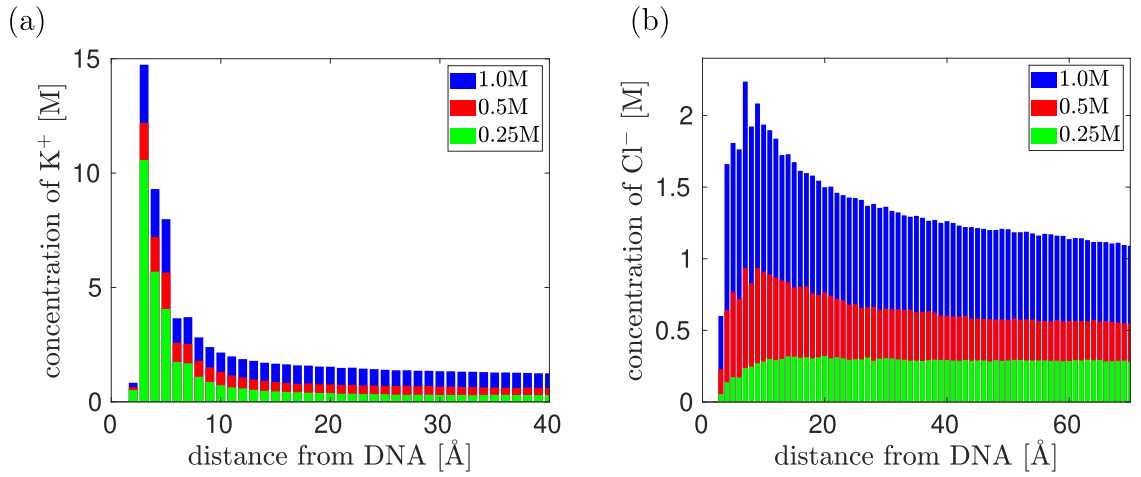



**Figure 7.** Results of all-atom MD simulations with APBC using the poly(A) DNA chain with $N = 10$ bp and three different concentrations of KCl in the bulk. (a) The concentration of $K^+$ ions given by eq 21 as a function of the distance from DNA. (b) The concentration of $Cl^-$ ions given by eq 21 as a function of the distance from DNA.

Each infinitely long sequence is modeled in our computational domain given by eq 1 with APBC using $N = 10n$ base pairs of DNA, where $n$ ranges from 1 to 10. The APBC is implemented along the $z$-direction as it is detailed in Appendix B.1. First, an $(N + 1)$ bp long dsDNA configuration is constructed in such a way that the $(N+1)$th base pair is equivalent to the first base pair translated to the $z$-direction. Then, a nucleotide at the 3'-end of each strand is removed and the bond to the 3'-end (removed) nucleotide is substituted with that to the first base at the 5'-end. The corresponding angles and dihedrals are added to MD structural files as detailed in Table 1 in Appendix B.1. In all MD simulations, we consider domain in eq 1 with $L_x = L_y = 200$ Å and we vary $L_z$. In Figures 4, 5, and 7, we choose $L_z$ as a multiple of $n$ (resp. $N$) with

$$L_z = 3.375N \text{ Å} = 33.75n \text{ Å} \tag{17}$$

while we study the effect of stretching and shrinking of DNA in Figure 6 by using $L_z$ obtained as the 95, 100, and 105% of the value given by eq 17. All MD simulations are done in KCl solutions, with $K^+$ ions neutralizing the negatively charged DNA segments. We use the concentration 150 mM KCl in Figures 4−6, while we vary the concentration of KCl in Figure 7.

When using APBC with polymer models, there are (locally) two important directions: parallel to the polymer chain and perpendicular to the polymer chain. We consider both of them, in Sections 3.1 and 3.2, respectively. In Section 3.1, we study the effects of APBC on the properties of the DNA chain, where we can make direct analogues to the results obtained for the persistence length of the WLC model in Section 2. This is followed by studying the characteristics of the surrounding solvent in Section 3.2, where we investigate the ion atmosphere around DNA for different concentrations of KCl.

**3.1. Dependence of the Polymer Chain Properties on APBC.** The persistence length for our (infinitely) long sequences of dinucleotides can be determined by various experimental and theoretical studies.[28] In Figure 4, we present the results of all-atom MD simulations with APBC of $N = 10$ bp segments using the six cases of repeated dinucleotides. Technical details of these MD simulations are given in Appendix B.

To analyze our MD results, we associate a unit orientation vector $\mathbf{h}_i$ with each base pair, i.e., $i = 1, 2, ..., N$, where $N = 10n$ is the total number of simulated base pairs. Denoting the angle between the $i$th and $(i + j)$th base pair as $\phi_j$, we have $\cos(\phi_j) = \mathbf{h}_i \cdot \mathbf{h}_{i+j}$, which we calculate for all $i = 1, 2, ..., N$. Averaging the calculated results over all possible values of $i$, we have

$$\langle \cos(\phi_j) \rangle = \langle \mathbf{h}_i \cdot \mathbf{h}_{i+j} \rangle \tag{18}$$

where the accuracy of this average is further improved by calculating it as a time average over relatively long MD time series. More precisely, we calculate three independent time series of length 10 ns and sample our results every 10 ps, disregarding the beginning of each simulation as the time required to equilibrate the system, see Appendix B for more details. Considering $N = 10$ (i.e., $n = 1$), we plot the averages in eq 18 in Figure 4a for values $j = 1, 2, 3, 4, 5$. We note that

$$\langle \cos(\phi_j) \rangle = \langle \cos(\phi_{N-j}) \rangle = \langle \cos(\phi_{10n-j}) \rangle$$

because we use APBC. In particular, the values of the averages in eq 18 for $j = 6, 7, ...$ are already represented in Figure 4a by the corresponding values for $j = 1, 2, 3, 4, 5$. In Figure 4, we observe that the results are clearly sequence-dependent for $j = 1$, with the $j = 4$ case providing the best match to the $j = 1$ case. On the other hand, the results are less sequence-dependent for $j = 2$ or $j = 5$. Given the APBC, there is no variation for $j = N = 10$ as we have already observed for the WLC model because of the constraint given by eq 12. In Figure 4b, we present the average separation between the subsequent base pairs for each of the studied case.

Our MD simulations in Figure 4 use the smallest possible value of $N$ (corresponding to $n = 1$), while one can expect that the results of all-atom MD simulations should be less influenced by the APBC for larger values of $n$ (in theory, the APBC-induced errors should decrease to zero in the limit $n \to \infty$). To investigate this further, we study the dependence of our results on $n$ for the poly(A) case in Figure 5a. We use three independent MD simulations for $n = 1, 2, 3, 4, 5, 10$ corresponding to simulations with $N$ ranging from 10 to 100 bp. In each case, we plot the averages given by eq 18 for $j = 1, 2, ..., 10$. We note that this average is trivially equal to 1 in the case $j = 10$ for $N = 10$ bp (because the 1st and the 11th base pairs are identical for $N = 10$ bp), so we omit this artificial value from our plot for 10 bp in Figure 5a. We observe that the results for $n = 1, 2, 3, 4, 5$ are matching some trends of the results for 100 bp. In particular, we can make similar conclusions as in Section 2 that the local properties (smaller values of $j$) are less influenced by using


https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

APBC than the averages estimated over the whole simulated polymer length (for $j$ comparable to $N$). In Figure 5b, we present the average separation between the base pairs. Comparing with Figure 4b, we observe that the separation between the base pairs is more influenced by the sequence of nucleotides (in Figure 4b) than by using APBC, with only 2−3% variations as we go from 10 bp to 100 bp APBC simulations in Figure 5b.

In Figure 3, we have considered the WLC model with $N = 100$ segments while varying the domain length $L_z$. In Figure 6, we present the results of a similar study using all-atom MD simulations with $N = 100$ bp. The middle bars in Figure 6a,b correspond to the results of the poly(A) case with 100 bp, which has already been included in Figure 5. Using eq 17, this corresponds to $L_z = 337.5$ Å. The other simulations correspond to the same setup where we either extend or shrink the value of $L_z$ by 5%, i.e., we use the values of $L_z$ given as

$$L_z = 320.625 \text{ Å}, \quad L_z = 337.5 \text{ Å}, \quad \text{and} \quad L_z = 354.375 \text{ Å} \tag{19}$$

In Figure 6b, we observe that the average separation between base pairs increases as we increase $L_z$. On the other hand, the behavior of averages in eq 18 is less monotonic as we stretch or shrink the DNA chain, see Figure 6a. Another way to visualize the results of all-atom MD simulations is to consider the average in eq 18 as a function of the distance between the base pairs,[20] which is visualized as function $\langle H \rangle$ in Figure 6c. To calculate $\langle H \rangle$, we average $\langle \mathbf{h}_i \cdot \mathbf{h}_j \rangle$ over all pairs $i$ and $j$ such that the corresponding base pairs are the distance $d$ apart. We present this average, $\langle H \rangle$, as a function of the distance $d$ in Figure 6c. The rate of decay of function $\langle H \rangle$ with distance $d$ can be used as an alternative way to define and estimate the persistence length from MD simulations.[20,27] For example, Kameda et al.[20] fit the subset of calculated data points $\langle H \rangle$ corresponding to $d$ in subinterval $3 \le d \le 100$ [Å] by function $f(d) = \exp[-d/a_p]$, where $a_p$ is the parameter defining the persistence length in their study (see Appendix B.3 for more details). Since $L_z$ given by eq 19 are all larger than 100 Å, we can use our data points $\langle H \rangle$ given in Figure 6c to repeat the analysis of Kameda et al.[20] We obtain that the parameter $a_p$ is equal to 60.9, 78.0, and 57.6 nm for the 95, 100, and 105% cases, respectively.

The DNA persistence length has been experimentally determined in the force extension measurements[23,29] to be around 50 nm at the physiological ionic concentrations. This has also been confirmed by different computational studies.[30,31] The dependence of the DNA elastic properties on the sequence of nucleotides has been studied both computationally[30] and experimentally.[28] The experimentally determined persistence length for our sequences of dinucleotides used in Figure 4 have been obtained in the literature[28] as 50.4 nm (poly(A)), 41.7 nm (poly(C)), 42.7 nm (poly(AT)), 49.6 nm (poly(CG)), 50.7 nm (poly(AC)) and 52.6 nm (poly(AG)). In Figure 4, we have used relatively short DNA segments (10 bp) and calculated $\langle \cos(\phi_j) \rangle$ defined in eq 18. This value cannot be directly used, for $j = 1$, in eq 16 to estimate the DNA persistence length because nucleic acids are mechanically softer at the scale of a few base pairs[32] and the WLC model is not applicable at this scale. On the other hand, if we use the value of $\langle \cos(\phi_{10}) \rangle$, which is the average cosine of the angle between the base pairs separated by one helical pitch in formula 16, together with the length of the helical pitch as the length $\delta$ of the segment, then formula 16 gives the values of the persistence length around 50 nm as reported in the experimental studies.[28,29] To get a better fit with the experimental data, the WLC model can be extended to the nonlocal twistable WLC model.[32] The standard WLC model is the best applicable to larger segments of micron-sized molecules.[24]

**3.2. Ion Atmosphere.** The APBC are useful for investigating solvent properties in the direction perpendicular to the polymer chain. In Figure 7, we present the results of such a study, calculating the radial distribution of $K^+$ and $Cl^-$ ions. We use three different concentrations of KCl, namely, 0.25, 0.5, and 1 M. In each case, we use $n = 1$, i.e., we use the APBC with 10 bp of poly(A) dsDNA. The results are calculated by averaging over four independent MD time series, each calculated for 10 ns. After the initial transient (of 1 ns) and at equidistant time intervals of 10 ps, we calculate the distance of each ion from the nearest atom of DNA, so our raw data are given in terms of the histograms

$N_{K^+}(r, \Delta r) =$
[number of ions with the distance from DNA in interval $(r, r + \Delta r)$]

$N_{Cl^-}(r, \Delta r) =$
[number of ions with the distance from DNA in interval $(r, r + \Delta r)$]

To get the radial distribution function, these numbers have to be divided by the volume, $V(r, \Delta r)$, giving the volume of all points that have their distance from the DNA in the interval $(r, r + \Delta r)$. Then, the radial distribution of $K^+$ ions and $Cl^-$ ions is theoretically defined by

$$\varrho_{K^+}(r) = \lim_{\Delta r \to 0} \frac{N_{K^+}(r, \Delta r)}{V(r, \Delta r)}, \text{ and}$$
$$\varrho_{Cl^-}(r) = \lim_{\Delta r \to 0} \frac{N_{Cl^-}(r, \Delta r)}{V(r, \Delta r)} \tag{20}$$

where $r$ is the distance from the DNA. To calculate Figure 7, we approximate the limit in eq 20 by choosing (relatively small) value $\Delta r = 1$ Å and we approximate the DNA as a straight line (or equivalently as a straight cylinder) in the $z$-direction, that is, $V(r, \Delta r) = 2\pi r \Delta r L_z$, giving

$$\varrho_{K^+}(r) \approx \frac{N_{K^+}(r, \Delta r)}{2\pi r \Delta r L_z}, \qquad \varrho_{Cl^-}(r) \approx \frac{N_{Cl^-}(r, \Delta r)}{2\pi r \Delta r L_z} \tag{21}$$

Formulas 21 are visualized in Figure 7 as histograms. We observe that the concentration of $K^+$ ions is higher in the vicinity of DNA because positive ions preferentially visit electronegative sites around DNA, although direct binding of ions to DNA bases is a relatively rare event, as previously reported in the literature.[4]

## 4. DISCUSSION AND CONCLUSIONS

Using MD simulations at constant pressure and temperature, we can solvate the DNA with water and ions, fixing the concentration of ions in the bulk. In Section 3.2, we have presented illustrative results of such all-atom MD investigations with APBC. Such simulations can also be used to estimate other solvent properties, for example, the moments of force distributions on ions, which can be used for parametrizing coarse-grained stochastic models of ions used in multiscale and multiresolution simulations.[16,19] The APBC simulations can also be coupled with coarse-grained models of water to design adaptive resolution simulation techniques.[5,33,34] In Section 3.2, we have presented the results calculated with APBC using $n = 1$ helical pitch. In particular, the simulated domain length is around 3.4 nm long and considerably smaller than the DNA's


https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

persistence length, which is about 50 nm. To study the mechanical properties of DNA, we need to increase the number of helical pitches as we have shown in Section 3.1 with our MD simulation results considering up to $n = 10$ helical pitches along the $z$-direction of the APBC simulation domain given by eq 1.

To get further insight into the correct use of the APBC, we have started our investigation using a discrete worm-like chain (WLC) model in Section 2, where we have observed in Figure 3 that the APBC affect less some local properties of the polymer chains than some global averages. In particular, the persistence length of the polymer chain can be estimated from the local properties of relatively short polymer chains, simulated with the help of APBC. The APBC are also applicable to simulations of biopolymers with larger persistence lengths (for example, actin filaments[35,36]) when a modeler is interested to understand the properties of the surrounding solvent. In Section 2, we have introduced the WLC model in the discretized form, where the polymer chain energy, given by eq 2, only includes the bending energy. Equation 2 is applicable to modeling a WLC polymer in the three-dimensional physical space $\mathbb{R}^3$ without any applied force. Such a model is analyzed in Section 2.1. In Section 2.2, we consider the WLC model with the APBC, which can be either viewed as introducing the additional holonomic constraint given by eq 12 or as stretching the polymer chain using the corresponding constraint force. In particular, the latter interpretation of the APBC also means that the polymer chain energy includes not only the bending energy given by eq 2 but also the corresponding stretching energy.[23,24] This viewpoint can be used for the analysis of the WLC model with constraints, although many analytical studies use a simplified eq 12, which is restricted to the $z$-direction, where the studied problem is effectively one-dimensional.[37] In the case of the APBC, we do have the constraints in the $x$-direction and $y$-direction as well, given by eq 12; see Appendices A.2 and A.3 for more details on the analysis and simulations, respectively, in this case.

In Appendix B, we provide the technical details of all-atom MD simulations, including the treatment of constant pressure simulations. The barostat used is again asymmetric with no fluctuations of $L_z$. In the APBC simulations, we have different treatments of the $z$-direction and all perpendicular directions in the $x$−$y$ plane. Simulations with 2D periodicity have also been used to study the behavior of a slab of water between two metallic walls,[38] which can be treated using three-dimensional Ewald techniques by including the image charges. One advantage of the APBC simulations is that they can be implemented with relatively minor modifications of standard all-atom MD tools,[39−43] as detailed in Appendix B. We note that the number of simulated base pairs, $N$, and the number of helical turns, $n$, in the DNA model with APBC has been fixed to satisfy $N = 10n$ in all presented MD simulations, and thus the model does not allow for over-winding or under-winding of DNA.[44] This can be studied by relaxing the assumption $N = 10n$ in all-atom MD simulations, and by extending our basic WLC model to twistable WLC models.[32,45]

## A. WLC MODEL: ADDITIONAL DERIVATIONS AND SIMULATIONS

Here, we provide additional details needed for the derivation of eqs 8, 9, and 15, and further details behind the Monte Carlo simulations in Figure 3.

### A.1. Derivation of Equations 8 and 9

Considering small values of $\alpha \ll 1$, we have $\exp[-\alpha\,\theta^2] = 1 - \alpha\,\theta^2 + O(\alpha^2)$. Substituting into eq 6 and evaluating the resulting integrals, we obtain

$$\langle \cos(\theta) \rangle = -\frac{\pi^2}{8}\,\alpha + O(\alpha^2), \quad \text{as} \quad \alpha \to 0$$

Substituting this expansion into eq 16, we obtain eq 9 in the limit $\alpha \to 0$. On the other hand, if $\alpha \gg 1$, we can use the double angle formula $\cos(\theta)\sin(\theta) = \sin(2\theta)/2$ and substitution $y = \theta\sqrt{\alpha}$ to deduce

$$\begin{aligned}
&\int_0^{\pi} \cos(\theta)\sin(\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta \\
&\quad = \frac{1}{2}\int_0^{\pi} \sin(2\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta \\
&\quad \approx \frac{1}{2\sqrt{\alpha}}\int_0^{\infty} \sin\left(\frac{2y}{\sqrt{\alpha}}\right)\exp[-y^2]\mathrm{d}y
\end{aligned}$$

where replacing the upper integration limit by $\infty$ only results in an exponentially small (negligible) error. Next, considering the limit $\alpha \to \infty$, we apply the Taylor expansion of the sine function, $\sin(x) \approx x - x^3/3! + x^5/5! - x^7/7!$, and evaluate the resulting integrals to obtain

$$\begin{aligned}
&\int_0^{\pi} \cos(\theta)\sin(\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta \\
&\quad = \frac{1}{2\alpha}\left(1 - \frac{2}{3\alpha} + \frac{4}{15\alpha^2} - \frac{8}{105\alpha^3} + O\left(\frac{1}{\alpha^4}\right)\right)
\end{aligned} \tag{22}$$

Similarly, using the substitution $y = \theta\sqrt{\alpha}$, we get

$$\begin{aligned}
&\int_0^{\pi} \sin(\theta)\exp[-\alpha\,\theta^2]\mathrm{d}\theta \\
&\quad = \frac{1}{2\alpha}\left(1 - \frac{1}{6\,\alpha} + \frac{1}{60\alpha^2} - \frac{1}{840\alpha^3} + O\left(\frac{1}{\alpha^4}\right)\right)
\end{aligned} \tag{23}$$

Substituting eqs 22 and 23 into 6 and dividing the power series, we obtain

$$1 - \langle \cos(\theta) \rangle = \frac{1}{2\alpha}\left(1 - \frac{1}{3\alpha} + \frac{7}{90\,\alpha^2} + O\left(\frac{1}{\alpha^3}\right)\right)$$

as $\alpha \to \infty$. Substituting this expansion into eq 16, we obtain eq 8 in the limit $\alpha \to \infty$.

### A.2. Derivation of Equation 15

Since $a_{\lfloor N/2 \rfloor} = a_1 = \delta$ for $N = 2$ and $N = 3$, eq 15 is satisfied for $N = 2$ and $N = 3$. Considering a general value of $N$, the constraint in eq 12 can be rewritten as

$$\sum_{i=2}^{N} \mathbf{l}_i = [0,\ 0,\ L_z] - \mathbf{l}_1$$

Taking the scalar product of this equation with $\mathbf{l}_1$, we get

$$\sum_{i=2}^{N} \langle \mathbf{l}_1 \cdot \mathbf{l}_i \rangle = [0,\ 0,\ L_z]\langle \mathbf{l}_1 \rangle - \delta^2 = \frac{L_z^2}{N} - \delta^2 \tag{24}$$

where we used $\langle \mathbf{l}_1 \rangle = [0, 0, L_z/N]$. Assuming that $\langle \mathbf{l}_1 \cdot \mathbf{l}_i \rangle$ are equal to each other on the left-hand side of eq 24, we get $\langle \mathbf{l}_1 \cdot \mathbf{l}_i \rangle = (L_z^2/N - \delta^2)/(N - 1)$ for $i = 2, 3, \ldots, N$. Substituting into eq 3, we obtain



https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

$$a_{\lfloor N/2 \rfloor} = \delta + \frac{1}{\delta}\sum_{i=2}^{\lfloor N/2 \rfloor} \langle \mathbf{l}_1 \cdot \mathbf{l}_i \rangle = \frac{\delta^2 N(N - \lfloor N/2 \rfloor) + (\lfloor N/2 \rfloor - 1)L_z^2}{\delta N(N-1)}$$

which is eq 15. Considering large values of $N$, eq 15 simplifies to

$$a_{\lfloor N/2 \rfloor} \approx \frac{\delta}{2} + \frac{L_z^2}{2\delta N}$$

which can also be deduced from eq 24 by considering that the left-hand side of eq 24 is twice the sum needed for the calculation of $a_{\lfloor N/2 \rfloor}$.

### A.3. Monte Carlo Simulations of the WLC Model

In Figure 3, we present some illustrative results estimated using Monte Carlo simulations of the WLC model with $N = 100$ segments for different values of the domain length, $L_z$, and the stiffness parameter, $\alpha$. We use $L_z \in \{10, 20, ..., 90\}$ and $\alpha \in \{0, 1, 5, 10\}$. Given $L_z$ and $\alpha$, our results are sampled over a long Monte Carlo trajectory consisting of $10^{10}$ steps. To implement APBC, we start with an initial configuration of the polymer satisfying eq 12. In our implementation, the initial $z$-coordinate of the $i$th polymer segment is equal to $iL_z/N$, for $i = 1, 2, ..., N$, while its $x$- and $y$-coordinates are chosen to be equal to $(-1)^i\sqrt{(\delta^2 - L_z^2/N^2)/8}$ . Then, the length of each segment is equal to $\delta$ and the eq 12 is satisfied. However, our results are independent of our specific choice of the initial configuration because we use the first $10^6$ steps to equilibrate our polymer configuration, i.e., we evolve the system over $10^6$ steps without saving any data. Then, we use $10^{10}$ steps to sample our results. In each Monte Carlo step, we select (a possible) new configuration of the polymer by choosing two nonadjacent beads at random (say the $i$th and $j$th bead) and we rotate the part of the polymer between the $i$th and $j$th bead along the line connecting these beads. We rotate it by a random angle. In this way, we make sure that the new configuration of the polymer satisfies the eq 12. The updated configuration is then either accepted or rejected using the standard Metropolis–Hastings criterion[26] based on the chain bending energy given by eq 2.

## B. ALL-ATOM MD SIMULATION DETAILS

Throughout the all-atom modeling and simulation, we used CHARMM36 force field, VMD 1.9.3 (for modeling),[39] and NAMD 2.14 (for simulation).[40]

### B.1. APBC Implementation

The periodic DNA models ($N$ bp) used for the APBC are generated as follows. First, an $(N + 1)$ bp long dsDNA configuration (in PDB format) is constructed with 3DNA (Web 3DNA 2.0[41,42]) in such a way that the $(N + 1)$th base pair is equivalent to the first base pair translated to the $z$-direction. Here, we use the base pair step parameter set for B-DNA (calf thymus; generic sequence) with Twist 36.0 degrees and Rise 3.375 Å for all nucleotide sequences. Note that the resulting structure shows a slight (sub-ångström) shift between the first and the $(N + 1)$th (i.e., deleted) bases depending on the length and nucleotide sequence, which is then relaxed in the initial equilibration simulation run. Then, a nucleotide at the 3'-end of each strand is removed, and the corresponding structure (in CHARMM PSF format) is generated by AutoPSF. Finally, by editing the PSF file, the bond to the 3'-end (removed) nucleotide is substituted with that to the first nucleotide (i.e., 5'-end), and the angles and dihedrals (4 and 7 terms for each strand, respectively) are reconnected accordingly. These additional terms are listed in Table 1. Note that, for the CHARMM36 force field, the type of the 5'-end phosphate (usually the first atom of each strand) should be changed from P to P2, as it is no longer at the terminal with the APBC.

**Table 1. List of Additional Terms Connecting the Ends of Each DNA Strand**

| | | | |
|---|---|---|---|
| bonds | | | |
| *N*th O3' | 1st P | | |
| angles | | | |
| *N*th C3' | *N*th O3' | 1st P | |
| 1st O1P | 1st P | *N*th O3' | |
| 1st O2P | 1st P | *N*th O3' | |
| 1st O5' | 1st P | *N*th O3' | |
| dihedrals | | | |
| *N*th C4' | *N*th C3' | *N*th O3' | 1st P |
| *N*th C2' | *N*th C3' | *N*th O3' | 1st P |
| *N*th C3' | *N*th O3' | 1st P | 1st O1P |
| *N*th C3' | *N*th O3' | 1st P | 1st O2P |
| *N*th C3' | *N*th O3' | 1st P | 1st O5' |
| *N*th H3' | *N*th C3' | *N*th O3' | 1st P |
| *N*th O3' | 1st P | 1st O5' | 1st C5' |

For simulation with explicit solvent, each model has been solvated in an $L_x \times L_y \times L_z$ TIP3P water box, where $L_x = L_y = 200$ Å and $L_z = 3.375N$ Å, then neutralized by $K^+$ and 0.15-1M KCl has been added in the bulk.

### B.2. Molecular Dynamics Simulation

After $10^4$ steps of energy minimization, each model was simulated for 10 ns at 2 fs time step (with SETTLE method for hydrogen). Nonbonded interactions were cut off at 12 Å (with switching from 10 Å), and particle mesh Ewald method was used for electrostatics. Langevin thermostat at 300 K (damping: 1 per ps) and Nosé–Hoover Langevin piston barostat at 1 atm (period: 2 ps, decay: 1 ps) were applied. For the barostat, constant area setting was used, keeping $L_z$ (and $L_x$) constant and adjusting only $L_y$ to control the pressure. The atomic positions were recorded at 10 ps intervals and used for analysis.

For simulations with periodic box length with different values of $L_z$ in Figure 6, the $z$-coordinate of each atom in the DNA model was rescaled and then solvated into the water box with the adjusted value of $L_z$.

### B.3. Analysis

The DNA structure in each snapshot was processed by 3DNA.[42] To consider also the base-pair step crossing the periodic boundary (i.e., the $N$th to the first base pair), a periodic image of the first residue (shifted by $L_z$) was added at the end of each strand, and the resulting $(N + 1)$ bp long segment was processed by 3DNA. The subsequent analysis then used the helical-axis positions and helical-axis vectors providing $\mathbf{h}_i$ in eq 18. The distance between adjacent base-pair steps was obtained as the distance between the corresponding helical-axis positions, and other distances along the chain are calculated as the sum of those adjacent distances within the section. These values were averaged over three simulation trials except the initial 1 ns (Figures 4, 5, and 7) or 5 ns (Figure 6) for each case.

To calculate $\langle H \rangle$ in Figure 6c, we average $\langle \mathbf{h}_i \cdot \mathbf{h}_j \rangle$ over all pairs $i$ and $j$ such that the corresponding base pairs are distance $d$ apart. More precisely, the distance $d$ is divided into intervals $[d - \Delta d/$


https://doi.org/10.1021/acs.jpcb.3c03887
J. Phys. Chem. B 2023, 127, 8257−8267

2, $d + \Delta d/2$) with width $\Delta d = 0.1$ Å, and $d$ takes values of multiples of $\Delta d$. To estimate the persistence length $a_p$, the data points $\langle H \rangle$ in the subinterval $3 \leq d \leq 100$ [Å] (the same setting as in Kameda et al.[20]), accumulated over the three trials for each case, are fit by function $f(d) = \exp[-d/a_p]$ using the method of least squares weighted by the number of samples in each bin.

## AUTHOR INFORMATION


### Corresponding Authors

**Radek Erban** − *Mathematical Institute, University of Oxford, Oxford OX2 6GG, U.K.;* orcid.org/0000-0001-8470-3763; Email: erban@maths.ox.ac.uk

**Yuichi Togashi** − *Department of Bioinformatics, College of Life Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan; RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo 650-0047, Japan;* orcid.org/0000-0003-0032-2424; Email: togashi@fc.ritsumei.ac.jp

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.3c03887


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council, grant number EP/V047469/1, awarded to R.E., and by the Japan Society for the Promotion of Science, KAKENHI grant number JP18KK0388 to Y.T. The computation was in part carried out using the computer resource offered under the category of General Projects by Research Institute for Information Technology, Kyushu University, and by the Advanced Research Computing (ARC) service at University of Oxford. In compliance with EPSRC's open access inititive, the data in this paper is available from http://dx.doi.org/10.5287/ora-j5ynyjg25.

## REFERENCES


(1) Vologodskii, A. *Biophysics of DNA*; Cambridge University Press: Cambridge, U.K., 2015.

(2) Jacobson, D. R.; Saleh, O. Counting the ions surrounding nucleic acids. *Nucleic Acids Res.* **2016**, *45*, 1596−1605.

(3) Mocci, F.; Laaksonen, A. Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is "worth its salt". *Soft Matter* **2012**, *8*, 9268−9284.

(4) Várnai, P.; Zakrzewska, K. DNA and its counterions: a molecular dynamics study. *Nucleic Acids Res.* **2004**, *32*, 4269−4280.

(5) Zavadlav, J.; Podgornik, R.; Praprotnik, M. Adaptive resolution simulation of a DNA molecule in salt solution. *J. Chem. Theory Comput.* **2015**, *11*, 5035−5044.

(6) Zavadlav, J.; Sablic, J.; Podgornik, R.; Praprotnik, M. Open-boundary molecular dynamics of a DNA Molecule in a hybrid explicit/implicit salt solution. *Biophys. J.* **2018**, *114*, 2352−2362.

(7) Dans, P. D.; Walther, J.; Gómez, H.; Orozco, M. Multiscale simulation of DNA. *Curr. Opin. Struct. Biol.* **2016**, *37*, 29−45.

(8) Korolev, N.; Nordenskiöld, L.; Lyubartsev, A. Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome. *Adv. Colloid Interface Sci.* **2016**, *232*, 36−48.

(9) Poppleton, E.; Matthies, M.; Mandal, D.; Romano, F.; Šulc, P.; Rovigatti, L. oxDNA: coarse-grained simulations of nucleic acids made simple. *J. Open Source Software* **2023**, *8*, No. 4693.

(10) Sengar, A.; Ouldridge, T.; Henrich, O.; Rovigatti, L.; Šulc, P. A primer on the oxDNA model of DNA: when to use it, how to simulate it and how to interpret the results. *Front. Mol. Biosci.* **2021**, *8*, No. 693710.

(11) Kovaleva, N.; Koroleva, I.; Mazo, M.; Zubova, E. The "sugar" coarse-grained DNA model. *J. Mol. Model.* **2017**, *23*, No. 66.

(12) Rolls, E.; Togashi, Y.; Erban, R. Varying the resolution of the Rouse model on temporal and spatial scales: application to multiscale modelling of DNA dynamics. *Multiscale Model. Simul.* **2017**, *15*, 1672−1693.

(13) Maffeo, C.; Aksimentiev, A. MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems. *Nucleic Acids Res.* **2020**, *48*, 5135−5146.

(14) Minhas, V.; Sun, T.; Mirzoev, A.; Korolev, N.; Lyubartsev, A.; Nordenskiöld, L. Modeling DNA flexibility: comparison of force fields from atomistic to multiscale levels. *J. Phys. Chem. B* **2020**, *124*, 38−49.

(15) Lee, S.; Rasaiah, J. Molecular dynamics simulation of ion mobility. 2. alkali metal and halide ions using the SPC/E model for water at 25°C. *J. Phys. Chem. A* **1996**, *100*, 1420−1425.

(16) Erban, R. Coupling all-atom molecular dynamics simulations of ions in water with Brownian dynamics. *Proc. R. Soc. A* **2016**, *472*, No. 20150556.

(17) Dobramysl, U.; Rüdiger, S.; Erban, R. Particle-based multiscale modeling of calcium puff dynamics. *Multiscale Model. Simul.* **2016**, *14*, 997−1016.

(18) Erban, R.; Chapman, S. J. *Stochastic Modelling of Reaction-Diffusion Processes*; Cambridge University Press: Cambridge, U.K., 2020.

(19) Erban, R. Coarse-graining molecular dynamics: stochastic models with non-Gaussian force distributions. *J. Math. Biol.* **2020**, *80*, 457−479.

(20) Kameda, T.; Suzuki, M. M.; Awazu, A.; Togashi, Y. Structural dynamics of DNA depending on methylation pattern. *Phys. Rev. E* **2021**, *103*, No. 012404.

(21) Lyubartsev, A. P.; Laaksonen, A. Molecular dynamics simulations of DNA in solution with different counter-ions. *J. Biomol. Struct. Dyn.* **1998**, *16*, 579−592.

(22) Korolev, N.; Lyubartsev, A.; Laaksonen, A.; Nordenskiöld, L. A molecular dynamics simulation study of oriented DNA with polyamine and sodium counterions: diffusion and averaged binding of water and cations. *Nucleic Acids Res.* **2003**, *31*, 5971−5981.

(23) Marko, J. F.; Siggia, E. Stretching DNA. *Macromolecules* **1995**, *28*, 8759−8770.

(24) Seol, Y.; Li, J.; Nelson, P.; Perkins, T.; Betterton, M. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6−7 μm. *Biophys. J.* **2007**, *93*, 4360−4373.

(25) Drozdetski, A. V.; Tolokh, I.; Pollack, L.; Baker, N.; Onufriev, A. Opposing effects of multivalent ions on the flexibility of DNA and RNA. *Phys. Rev. Lett.* **2016**, *117*, No. 028101.

(26) Chib, S.; Greenberg, E. Understanding the Metropolis-Hastings algorithm. *Am. Stat.* **1995**, *49*, 327−335.

(27) Wu, Y.-Y.; Bao, L.; Zhang, X.; Tan, Z. Flexibility of short DNA helices with finite-length effect: From base pairs to tens of base pairs. *J. Chem. Phys.* **2015**, *142*, No. 125103.

(28) Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15421−15426.

(29) Baumann, C. G.; Smith, S.; Loomfield, V.; Bustamante, C. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185−6190.

(30) Ortiz, V.; de Pablo, J. Molecular Origins of DNA Flexibility: Sequence Effects on Conformational and Mechanical Properties. *Phys. Rev. Lett.* **2011**, *106*, No. 238107.

(31) Zhang, C.-Y.; Zhang, N. Mechanical constraint effect on DNA persistence length. *Molecules* **2022**, *27*, No. 7769.

(32) Segers, M.; Voorspoels, A.; Sakaue, T.; Carlon, E. Mechanical properties of nucleic acids and the non-local twistable wormlike chain model. *J. Chem. Phys.* **2022**, *156*, No. 234105.

(33) Zavadlav, J.; Melo, M.; Marrink, S.; Praprotnik, M. Adaptive resolution simulation of an atomistic protein in MARTINI water. *J. Chem. Phys.* **2014**, *140*, No. 054114.

(34) Zavadlav, J.; Melo, M.; Cunha, A.; de Vries, A.; Marrink, S.; Praprotnik, M. Adaptive resolution simulation of MARTINI solvents. *J. Chem. Theory Comput.* **2014**, *10*, 2591−2598.

(35) Floyd, C.; Ni, H.; Gunaratne, R.; Erban, R.; Papoian, G. On stretching, bending, shearing and twisting of actin filaments I: variational models. *J. Chem. Theory Comput.* **2022**, *18*, 4865−4878.
(36) Gunaratne, R.; Floyd, C.; Ni, H.; Papoian, G.; Erban, R. On stretching, bending, shearing and twisting of actin filaments II: multi-resolution modelling. arXiv:2203.01284. arXiv.org e-Print archive, 2022. https://doi.org/10.48550/arXiv.2203.01284.
(37) Marantan, A.; Mahadevan, L. Mechanics and statistics of the worm-like chain. *Am. J. Phys.* **2018**, *86*, 86−94.
(38) Hautman, J.; Halley, J.; Rhee, Y. Molecular dynamics simulation of water between two ideal classical metal walls. *J. Chem. Phys.* **1989**, *91*, 467−472.
(39) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(40) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, No. 044130.
(41) Li, S.; Olson, W. K.; Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **2019**, *47*, W26−W34.
(42) 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.
(43) Kameda, T.; Awazu, A.; Togashi, Y. Molecular dynamics analysis of biomolecular systems including nucleic acids. *Biophys. Physicobiol.* **2022**, *19*, No. e190027.
(44) Marin-Gonzalez, A.; Vilhena, J. G.; Perez, R.; Moreno-Herrero, F. Understanding the mechanical response of double-stranded DNA and RNA under constant stretching forces using all-atom molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 7049−7054.
(45) Gore, J.; Bryant, Z.; Nöllmann, M.; Le, M.; Cozzarelli, N.; Bustamante, C. DNA overwinds when stretched. *Nature* **2006**, *442*, 836−839.




https://doi.org/10.1021/acs.jpcb.3c03887
*J. Phys. Chem. B* 2023, 127, 8257−8267