A Coarse-Grained Simulation Study of the Structures, Energetics, and Dynamics of Linear and Circular DNA with Its Ions

Aymeric Naômé,*,[†,‡,¶,§] Aatto Laaksonen,[¶,§,‖] and Daniel P. Vercauteren[†,‡]

[†]Laboratoire de Physico-Chimie Informatique, Unité de Chimie Physique Théorique et Structurale, University of Namur, 5000 Namur, Belgium
[‡]Namur Medicine & Drug Innovation Center (NAMEDIC), University of Namur, 5000 Namur, Belgium
[¶]Division of Physical Chemistry, Department of Materials and Environmental Chemistry, Arrhenius Laboratory, Stockholm University, 10691 Stockholm, Sweden
[§]Science for Life Laboratory, 17121 Solna, Sweden
[‖]Stellenbosch Institute of Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, 7600 Stellenbosch, South Africa

S *Supporting Information*

ABSTRACT: We have studied the structural, energetics, and dynamical properties of a variety of linear and circular DNA fragments using a solvent-mediated coarse-grained (CG) model of DNA with explicit ions recently developed by us [Naômé et al., *J. Chem. Theory Comput.*, **2014**, *10*, 3541−3549]. We particularly examined the treatment of electrostatics and determined that a large cutoff is necessary to properly reproduce the DNA flexibility. Moreover, it is crucial to include long-ranged electrostatic interactions: a Particle Mesh Ewald scheme at low resolution is sufficient to avoid structural artifacts. We calculated the ring closure probabilities, as *j*-factors, for DNA fragments of different lengths from equilibrium, as well as restrained molecular dynamics (MD) simulations. The latter force integration method provided accurate results without model fitting. We generated topology and energy maps for DNA minicircles of various lengths and helical densities, at low and high ion concentrations. A general trend for structure compaction is observed, driven by an increase in writhing as the ionic concentration increases. Finally, we applied a reconstruction procedure to generate detailed molecular structures from the various superhelical conformations generated by the CG MD of the DNA minicircles. These pre-equilibrated reconstructed atomistic structures can serve as starting material for atomistic simulations.

1. INTRODUCTION

The necessity to address biologically relevant systems consisting of thousands of molecules and millions of atoms across times reaching the microsecond scale has seen the blooming of a variety of modern coarse-grained (CG) models. These simplified models, which were used in the past mainly because of limited computer capabilities, and as first attempts toward present-day complex and accurate atomistic force fields (FFs), are currently the preferred tools to construct the *computational microscope* complementing experimental microscopy techniques. In addition to unveiling a new level of complexity, CG methods also get rid of the flood of details that are less relevant to the scales under consideration. Complex phenomena such as protein plasticity (folding, domain rearrangement, ...), membrane-mediated transactions (fusion, signaling, secretion), nucleic acids structuring, and molecular recognition are those, not to cite all, of which the understanding will be advanced thanks to CG modeling. For an overview of biomolecular CG modeling, the reader is referred to the reviews by Noid[1] and Ingólfsson et al.[2] A broader view

on the philosophy of coarse-graining can be found in a review by Kamerlin et al.[3]

Despite the plethora of published models for biomolecules— lipids in front, followed by proteins—only a handful involve DNA. Moreover, among those, only a few have led to continued development and applications. Three such successful models are, in order of number of articles published, the oxDNA model of Ouldridge and co-workers,[4−6] the 3SPN family models of de Pablo and co-workers,[7−12] and the SIRAH model of Pantano and co-workers.[13−16]

We should also acknowledge the original models of Linak et al.,[17] Edens et al.,[18] and Cragnolini et al.[19]

These CG models are classified as "top-down" models. They are empirically parametrized, in a trial-and-error manner, such that they match experimentally determined thermodynamical properties or structural features of macroscopic systems, such as the melting temperatures of oligomers and the structure of the double helix. The three aforementioned medium-resolution

models (in the range of 3−6 interaction sites per nucleotide) are able, to different extents, to describe DNA strands hybridization and melting thermodynamics and kinetics, with sequence complementarity. These models, at their current stage of development, are useful to the understanding of bare DNA systems and find application in DNA-based nanotechnologies. (Let us mention that the 3SPN model has been recently combined with a protein CG model.[20]) On the topic, we refer the reader to the recent review by Ouldridge.[21]

Another class of structure-based DNA CG models, namely, the "bottom-up" models, has developed in parallel. The approach consists in constructing a set of interactions that reproduces correlation functions from more-detailed atomistic simulations or data from experimentally determined structures. Typically, distributions of structural parameters defined between CG sites mapped onto the atomistic representation of DNA are the target features that the "bottom-up" CG models aim to reproduce. One advantage of the "bottom-up" approach resides in that it requires very little to no parametrization effort, since it is hierarchically derived from an underlying atomistic model. Compared to "top-down" models, there is thus a more rigorous, physics-based connection with the models with finer details. However, this is also its main weakness, since it only ensures that the input information is properly reproduced in subsequent CG simulations (i.e., the "bottom-up" models are highly state-dependent). An additional issue is that the model will inherit, at best, the quality of the fine-grained description, which offers no guarantee on how well the phenomena pertaining to higher time and size scales can be reproduced. This lack of horizontal (between different systems) and vertical (across scales) transferability is the reason why "bottom-up" models are lagging behind the "top-down" models, in terms of applicative power. Most models are indeed unable to describe both duplex and single-stranded DNA, discarding access to their interconversion which is at the heart of the majority of "top-down" models.

Despite these reservations, a few valuable "bottom-up" CG models have emerged with various fields of application. There exist a few techniques grounded on statistical mechanics to *systematically* derive such models, among which the (iterative) Boltzmann inversion ((I)BI)[22,23] and the Newton inversion (NI)[24,25] methods have already been successfully applied to a variety of molecular systems:[26] molecular liquids, ions in solution, liquid crystals, polymer melts, lipids, ionic liquids, and DNA. These methods essentially transform a set of distributions of CG structural parameters into a set of effective interaction potentials. For example, Trovato and Tozzini constructed a one-bead-per-nucleotide model based on the structural parameters of B-DNA crystallographic structures.[27] The intrastrand bond, angle, and dihedral are modeled with a conventional harmonic potential of which the force constant is extracted by fitting the effective potentials derived with IBI. The base-pairing and nonbonded terms are similarly fitted with a Morse potential that artificially allows the two strands to dissociate. This model is used to study the dynamics of plasmid structures. Sayar and co-workers studied the same type of systems with a two-site model.[28] In their case, the FF is constructed on a simpler direct BI of probability distributions extracted from all-atom (AA) MD simulations trajectories, followed by the fitting of the resulting potentials of mean force (PMFs) with a harmonic function. Maffeo et al. derived a single-stranded DNA model with tabulated interaction potentials using IBI that performs as well or better than the

"top-down" models of Ouldridge and de Pablo in a range of single-molecule experiments.[29] The NI method, which is more rigorously based on statistical mechanics considerations, basically considers the cross-correlation between the different structural distributions and therefore enables a better convergence of the corresponding interaction potentials. This approach was developed and applied by Lyubartsev and Laaksonen to various systems including DNA.[24,30] The model of Korolev et al.[31] consists of one bead per phosphate group and a central bead encompassing the sugar and nucleobase atoms of a two-base-pair (two-bp) unit. As in the work of Trovato and Tozzini[27] or Sayar et al.,[28] the final potentials for intramolecular distributions are fitted with a harmonic function. The nonbonded interactions are modeled by a Coulomb and a Lennard-Jones potential with the appropriate charges and radii. Savelyev et al.[25] also applied the NI method, but they used the parameters of fitting functions to the potentials as the observables subject to optimization instead of the values of the potential themselves. These two CG DNA FFs with explicit ions are able to describe the persistence length of double-stranded DNA and its ionic concentration dependence and offer good candidate models to study chromatine fibers, in combination with protein models. One last "bottom-up" model to mention is due to Maciejczyk and co-workers.[32,33] This CG FF differs from the previous ones, since it is constructed following the philosophy and methodology of the AA FFs.[34] The nonbonded terms in the form of Lennard-Jones centers, charges, and electric dipoles are positioned on beads mapped onto the four nucleotides. Their parameters are optimized to best reproduce the AA FFs Lennard-Jones energies for a set of configurations, as well as the quantum-mechanical electrostatic potential. The bonded terms are extracted from umbrella sampling of the internal degrees of freedom. They provided a model able, with minimal manual adjustment of several parameters, to fold short duplexes. One other parametrized model, which is similar in philosophy, is attributed to the work of Morris-Andrews et al.[35]

We have also contributed to the surging of new CG models. The DNA CG model that we have derived[36] used both the IBI and NI methods. This one-bead-per-nucleotide model with explicit ions shares the same qualities as those from Korolev et al.[31] and Savelyev et al.[37] However, it stands out from the others because of its greater systematic nature. Unlike the model of Korolev et al., we do not fit the final potentials to any functional form nor restrict the shape of the potentials to a predetermined shape, as observed in the work of Savelyev et al. They are used as such in a tabulated form, with no scaling factor or adjustable parameters.

In this paper, a collection of applications of our CG DNA model is presented. We first pinpoint the range of optimal simulation parameters that guarantees no artifacts. We then study aspects of the process of cyclization of small DNA fragments using both energetic and statistical approaches, taking advantage of restrained MD simulations. DNA minicircles are then characterized by shape, strain energy, and degree of supercoiling. Finally, a procedure of molecular details reconstruction is provided, robust enough to generate well-behaved atomistic structures of arbitrary sequence even from highly distorted supercoiled CG DNA minicircles.

2. BACKGROUND AND METHODS

2.1. Coarse-Grained Model. The DNA model used in this work is derived using a sequential Iterative Boltzmann

Inversion/Newton Inversion (IBI/NI) scheme. This systematic method consists of iteratively reconstructing a set of effective interaction potentials that will reproduce target structural distributions coming from more-detailed simulations. Although these methods are explained in detail in numerous papers to which the reader is referred (for IBI, see refs 22, 23, 36, and 38; for NI, see refs 24, 30, 36, and 38), we will outline the important features in the specific context of our DNA CG model.

The structural distributions our model is based on are extracted from all-atom (AA) MD simulations of DNA 18-mers of various sequences[39] based on the Amber ff99bsc0 FF[40−42] with the SPC/E water model[43] and Dang et al. ion parameters for KCl.[44] These detailed trajectories were first mapped with the chosen CG representation. In this work, every nucleotide and ion is mapped as a unique interaction site located at the center of mass of the atoms they represent. DNA CG sites bear a mass taken as the average of each of the four nucleotides (A, C, G, T) mass, i.e., 307.9 atomic mass units. The pair distance distributions were then computed between the CG sites according to a simple systematic interaction network. Two networks were selected, coined *10pot* and *16pot*, both comprising five intermolecular pair interactions between ions and DNA CG sites (K−K, K−Cl, Cl−Cl, R−K, R−Cl) and either 10 or 16 DNA intramolecular terms, respectively. These terms are composed of intrastrand pairs linking a CG site i to its neighbors $i+n$ along the same strand and interstrand pairs connecting a site i to its Watson−Crick partner j and to the upstream $j−n$ and downstream $j+n$ sites located on the complementary strand (Figure 1). From the set of target pair
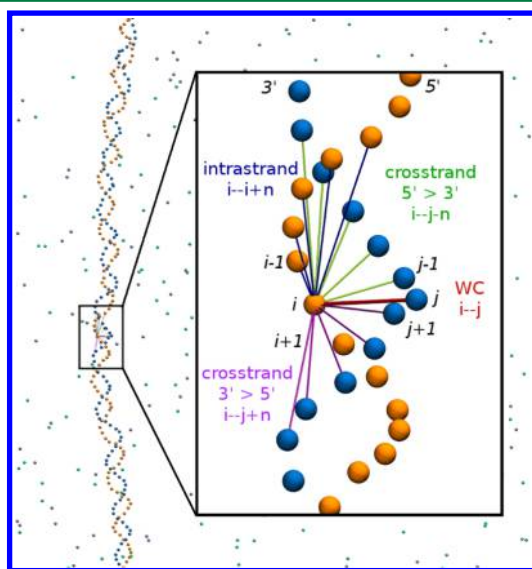
Figure 1. One-site-per-nucleotide CG DNA model. A bead represents the interaction center placed at the center of mass of a nucleotide. Every site interacts with other nucleotides through intrastrand and interstrand interactions. For the *10pot* and *16pot* FFs, $n = 3$ and 5, respectively.

distributions S^*, a corresponding set of initial effective interaction potentials was obtained as the potential of mean force (PMF) $V_{PMF} = -kT \ln S^*$. Practically, the reference distributions S^* are divided in slices of width r_α so that, to every S_α^* bin, corresponds a value of the discretized PMF ($V_\alpha = -kT \ln S_\alpha^*$). The effective potentials we are using are thus in the form of tables with a resolution of r_α, which is set to 0.05 Å. We

first performed a series of IBI runs using Metropolis Monte Carlo (MC) as a sampling method to generate new sets of $\{S_\alpha\}$. The set of potentials $\{K_\alpha\}$ is updated as

$$K_\alpha^{(n+1)} = K_\alpha^{(n)} + kT \ln\left(\frac{S_\alpha^{(n)}}{S_\alpha^*}\right)\lambda \tag{1}$$

with λ being a scaling parameter between 0 and 1. These mean field corrections typically provide reasonably converged calculated distributions in a few iterations and few MC steps per iteration. To push the convergence further, we conducted the second NI phase of the scheme. In the NI, $\{K_\alpha\}$ is updated in a thermodynamically consistent way without the assumption, as in the IBI, that the different pair interactions have independent distributions. After every iteration, the system of eq 2 must be solved with the partial derivatives given in eq 3, which are the covariance elements of all the pairs of S_α^* bins:

$$\langle S_\alpha^{(n)} \rangle - S_\alpha^* = \sum_\gamma \frac{\partial \langle S_\alpha^{(n)} \rangle}{\partial K_\gamma^{(n)}} \Delta K_\gamma^{(n)} + O(\Delta K_\gamma^{~2}) \tag{2}$$

$$\frac{\partial \langle S_\alpha^{(n)} \rangle}{\partial K_\gamma^{(n)}} = -\frac{(\langle S_\alpha^{(n)} S_\gamma^{(n)} \rangle - \langle S_\alpha^{(n)} \rangle \langle S_\gamma^{(n)} \rangle)}{kT} \tag{3}$$

Again, the potential $\{K_\alpha\}$ is updated as

$$K_\alpha^{(n+1)} = K_\alpha^{(n)} + \Delta K_\alpha^{(n)}\lambda \tag{4}$$

A very high convergence can be achieved with NI but more MC steps per iteration are necessary to properly sample the values in eq 3. The factor λ also must be kept below 0.2, to avoid instabilities of the potentials. In the MC simulations, electrostatics was explicitly considered (using Ewald summation) for the intermolecular pairs (ions and DNA). Therefore, the effective potentials obtained for those pairs must be combined with some treatment of electrostatics in later CG simulations. The IBI/NI procedure was carried out using the MagiC software[45] already described in more detail in the work of Naômé et al.[36] In that article, the CG model was also characterized and validated. It was shown that the reference distributions are well-reproduced in CG MD simulations and that a mechanical property such as the persistence length, and its dependence on the ionic strength, calculated for 300-bp DNA fragments, is in qualitative agreement with experimental measurements.

As already mentioned, the major strength of such a "bottom-up" model is that its physical properties are, to some extent, inherited from the detailed system upon which it is resting, and, consequently, parametrization is neither needed nor possible. Its strength is also its main weakness, since models derived with IBI/NI or other systematic methods such as force-matching[46,47] are strictly state-dependent (composition, temperature, pressure, configuration space sampled). In conjunction to our model, this does not only imply the fact that we are constrained to a B-DNA state with nucleotides with averaged chemical identity; it also means that we must be very cautious in drawing conclusions from cases where the ionic concentration deviates strongly from 150 mM and the temperature deviates from 300 K (pertaining to the AA simulations).

2.2. Coarse-Grained Molecular Dynamics. The CG MD simulations were performed using GROMACS v4.5.[48] The smoothed, interpolated, and switched potentials (500 points/nm) were stored in tables. Similar to that observed in the MC

simulations, the interaction potentials between charged CG sites comprise an electrostatic contribution: here, either in the form of a Coulomb $1/r$ potential or Particle Mesh Ewald (PME). The relative dielectric constant was taken as $\epsilon_r = 80$ to account for the screening effect of water. The force field (FF) used is *10pot* and the plain cutoff or direct space cutoff is $r_{cut} = 8.0$ nm, except when stated otherwise. For the PME simulations, a grid spacing of 1.6 nm was used ($\sim$885 000 grid points for the largest systems considered). MD simulations were performed in the NVT ensemble with a Nosé–Hoover thermostat (time constant $\tau_T = 0.5$ ps). A stable integration time step of 5 fs was used in all simulations.

All simulations in the "Persistence Length and Electrostatics" section, presented later in this work, are based on the *10pot* FF. Among the five different interaction networks considered in a previous article,[36] this is indeed the FF that gives the persistence length of DNA in closest agreement with the commonly accepted value of 50 nm.[49] The (*16pot*) FF is also retained, because it is the most extended network mapped onto the molecular structure of DNA (up to 5 bp upstream and downstream), thus probably reflecting the actual rigidity of the underlying atomistic model more closely. This second model is used in the "j-Factor" section, for comparison purposes, to assess the accuracy of a free-energy method and in the "DNA minicircles" section, to explore the variety of superhelical structures arising from a different higher strain energy density.

2.3. Persistence Length. The persistence length l_P quantifies the stiffness of linear polymer chains as the distance after which the orientation of the chain becomes uncorrelated.[50] It is also the unique mechanical parameter of the elastic rod model—or worm-like chain (WLC)—used to describe semiflexible polymers. DNA is thus depicted as a chain of inextensible jointed links with an isotropic bending rigidity that determines how much the angle between successive segments can fluctuate.[51,52] We evaluated l_P using the exponent decay approximation:

$$\langle \hat{\mathbf{l}}_i \cdot \hat{\mathbf{l}}_j \rangle_{|i-j|=n} = \exp\left(-\frac{n\|\mathbf{l}\|}{l_P} \right) \tag{5}$$

where $\hat{\mathbf{l}}_i$ is a unit vector along the segment i direction ($\hat{\mathbf{l}}_i = \mathbf{l}/\|\mathbf{l}_i\|$) and $\|\mathbf{l}\|$ is the average segment length. The direction vectors are chosen along the segments separating the midpoints between every 10th bp nucleotide ($i-i+10$) (roughly corresponding to a DNA turn, $|i-j| = 10$ in eq 5) in order to avoid the oscillation of the autocorrelation function due to the groove asymmetry.

The persistence length l_P was already calculated for 300 bp DNA fragments.[36] The *10pot* FF produced an asymptotic value of $l_P = 51$ nm at large K^+ concentrations, in a good agreement with the experimental value of $\sim$50 nm ($\sim$150 bp, $[Na^+] = 150$ mM).[49] We calculated here l_P for several other systems with shorter and longer DNA fragments (90–500 bp), for ionic concentrations up to 50 mM, again with both plain cutoff and PME electrostatics. All simulations performed are 1 μs long (2 $\times 10^8$ time steps) and the standard deviation of l_P were calculated from ten 100 ns subtrajectories. The sampled conformations provide well-behaving Gaussian probability distributions of l_P. For the sake of clarity, the standard deviations are not included in the graphs as error bars: they are only mentioned in the text if informative.

2.4. DNA Cyclization. An alternate way of determining the flexibility of our DNA model is to calculate the so-called *j*-factor, quantitatively characterizing its cyclization process.[53] Its value corresponds to the probability of having the two ends of a same polymer chain in contact, normalized by the probability of having two ends of different chains close to each other.[53,54] Experimentally, the *j*-factor is measured as the ratio of the rate of cyclization to the rate of bimolecular association, which relates to the standard free energies of cyclization and dimerization as

$$j \equiv \frac{K_{cyc}}{K_{dim}} = \exp\left[\frac{-(\Delta G^\circ_{cyc} - \Delta G^\circ_{dim})}{RT} \right] \tag{6}$$

For cyclization to take place, not only the ends must be in contact but also their axial orientation and helical twist must be aligned and in-phase. In terms of probabilities, the free energies in eq 6 can be expressed as

$$\Delta G^\circ = \Delta G^\circ_r - RT \ln[\mathbf{W}(0)\Gamma(0; 1)\Phi(0, 1; \tau_0)\, dV\, d\gamma\, d\tau] \tag{7}$$

where ΔG°_r is the change in standard free energy associated with covalent and noncovalent interactions between the terminal chain segments, $\mathbf{W}(0)$ is the volume probability density for two ends to be found in close contact, $\gamma(0;1)$ is the conditional probability density of having the two end segments in both close contact and coaxial alignment, and $\Phi(0,1;\tau_0)$ is the conditional probability density of having the two end segments in close contact, coaxial alignment, and proper torsional orientations τ_0.[54] In the limit of long DNA fragments, the orientation and torsional correlation of the end segments can be neglected and considered evenly distributed:

$$\Delta G^\circ_{cyc} = \Delta G^\circ_r - RT \ln\left[\mathbf{W}(0)\left(\frac{1}{2}\right)\left(\frac{1}{2\pi}\right) dV\, d\gamma\, d\tau\right] \tag{8}$$

Since the relative orientations and distances between the ends from different DNA fragments are always considered as uniformly distributed, the dimerization probability is dependent only on the chain ends concentration:

$$\Delta G^\circ_{dim} = \Delta G^\circ_r - RT \ln\left[N_A c^\circ\left(\frac{1}{2}\right)\left(\frac{1}{2\pi}\right) dV\, d\gamma\, d\tau\right] \tag{9}$$

where N_A is Avogadro's number and c° is the standard state concentration of chain ends. Taking into account that ΔG°_r is the same for cyclization and bimolecular reactions, the *j*-factor in eq 6 reads

$$j = \frac{\mathbf{W}(0)}{N_A} \tag{10}$$

In the frame of Gaussian chains statistics, which applies for long DNA fragments (>1 kbp), the *j*-factor becomes

$$j = \frac{(3/2\pi l_c^2)^{3/2}}{N_A} \tag{11}$$

with l_c being the contour length of the chain. This expression, which is the result of work by Jacobson and Stockmayer,[53] was shown to be in good agreement with the first experimental determination of the *j*-factor for long DNA fragments (>5 kbp) with overhangs.[55] However, it is not valid for shorter DNA fragments (<500 bp) that behave similar to elastic rods and are better described with the WLC model; the energy cost to bend and twist short stiff chains lowers the *j*-factor in eq 11. With the non-negligible correlation of the relative configuration of the chain ends, combining eqs 7 and 9 gives

$$j = \frac{4\pi \mathbf{W}(0)\gamma(0; 1)\Phi(0, 1; \tau_0)}{N_A} \tag{12}$$

We recognize eq 10 as j_0, the radial part of the j-factor, and the torsion independent part:

$$j_1 = \frac{2\mathbf{W}(0)\gamma(0; 1)}{N_A} = 2j_0\gamma(0; 1) \tag{13}$$

for which a uniform torsion angle distribution $\Phi(0,1;\tau) = (1/2)\pi$ is assumed.

Analytical expressions of the j_0 and j_1 components for semiflexible chains have been derived in the frame of a twistable WLC model by Shimada and Yamakawa[56] and consider the persistence length l_P to be a unique parameter:

$$j_0 = 28.01\left(\frac{l_c}{2l_P}\right)^{-5} \exp\left[-7.027\left(\frac{l_c}{2l_P}\right)^{-1} + 0.492\left(\frac{l_c}{2l_P}\right)\right] \times 10^{24}$$
$$\times \frac{l_P^3}{N_A} \tag{14}$$

$$j_1 = 4\pi^3\left(\frac{l_c}{2l_P}\right)^{-6} \exp\left[-\pi^2\left(\frac{l_c}{2l_P}\right)^{-1} + 0.514\left(\frac{l_c}{2l_P}\right)\right] \times 10^{24} \times \frac{l_P^3}{N_A} \tag{15}$$

The full j-factor comprising the torsional contribution cannot be expressed in a closed form but can be evaluated numerically and contains the DNA helical repeat h and a twisting force constant as additional parameters.[56] A program calculating the WLC j-factor is available from Vologodskii's group.[57] A widely used method to determine the j-factor of shorter DNA fragments is DNA ligase-catalyzed cyclization kinetics.[58] The results from those cyclization assays are fitted very well with the equations from the work of Shimada and Yamakawa.[59−62]

We have evaluated the j-factor of our CG models in three ways:

• First, simply by evaluating j_0 (eq 14) using the persistence lengths calculated from eq 5.

• Second, by extracting the value of $\mathbf{W}(0)$ from the radial probability densities $\mathbf{Q}(r)$ of the end-to-end distance r/l_c. In this method, $\mathbf{Q}(r)$ is computed by counting the fraction of structures, generated in the CG MD simulations, whose ends fall into a spherical shell bound by two spheres of radii r/l_c and $(r + \Delta r)/l_c$ and dividing by the volume of the shell. The histogram is then fitted with the interpolation formula of Becker et al.:[63]

$$\mathbf{Q}(r) = A \times 112.04\kappa^2 \times \exp\left(\frac{0.246}{\kappa - a\kappa}\right) \times \left(\frac{1 - cr^2}{1 - r^2}\right)^{5/2}$$
$$\times \exp\left(\frac{\sum_{i=-1}^{0}\sum_{j=1}^{3} c_{i,j}\kappa^i r^{2j}}{1 - r^2}\right)$$
$$\times \exp\left(-\frac{d\kappa\, ab(1 + b)r^2}{1 - b^2 r^2}\right) \times I_0\left(-\frac{d\kappa a(1 + b)r}{1 - b^2 r^2}\right) \tag{16}$$

κ being the fitting stiffness parameter, I_0 being a modified Bessel function of the first type, and the parameters are given as

$$a = 14.054, \quad b = 0.473$$

$$1 - c = [1 + (0.38\kappa^{-0.95})^{-5}]^{-1/5}$$

$$1 - d = \frac{1}{0.177/(\kappa - 0.111) + 6.40(\kappa - 0.111)^{0.783}}$$

The radial part of the j-factor (j_0) is then taken as the extrapolated value of $\mathbf{Q}(r)$ at zero distance and converted to the proper concentration units by dividing by N_A and l_c^3 (since $\mathbf{Q}(r)$ has the units of l_c^{-3}):

$$j_0 = \frac{\mathbf{W}(0)}{N_A} = \frac{\mathbf{Q}(0)}{N_A l_c^3} \tag{17}$$

• Third, by directly calculating the free energy ΔG_{cyc} of bringing the two ends of a DNA fragment in contact. Since it is illusory to obtain, from equilibrium MD, a proper conformational sampling of all possible configurations of a DNA chain, we conducted instead series of simulations in which the Cartesian coordinates of the two end residues are restrained by a harmonic potential, from fully extended to end-to-end contact. Since we only control the radial component of the cyclization process, the calculated free energy ΔG_{cyc} expressed in the formalism of eq 7 corresponds to

$$\Delta G_{cyc} = \Delta G_{ele} - RT\, \ln\left[\mathbf{W}(r_{con})\, dV \int_{-1}^{1} \{\gamma(r_{con}; \gamma)\right.$$
$$\left. \times \int_{0}^{2\pi} \Phi(r_{con}, \gamma; \tau)\, d\tau\} \, d\gamma\right]$$
$$= \Delta G_{ele} - RT\, \ln[\mathbf{W}(r_{con})\, dV] \tag{18}$$

where ΔG_{ele} is the free-energy change arising from the repulsion of the negative charges on both end segments as they come to contact distance (corresponding to ΔG_r° in eqs 8 and 9). The probability densities for angular $\gamma(r_{con};\gamma)$ and torsional alignment $\Phi(r_{con},\gamma;\tau)$ are left unaffected and sum to unity. The j-factor can thus be written as

$$j' = \exp\left(-\frac{\Delta G_{cyc}}{RT}\right) = \exp\left(-\frac{\Delta G_{ele}}{RT}\right)\mathbf{W}(r_{con})\, dV$$
$$\approx \exp\left(-\frac{\Delta G_{ele}}{RT}\right) j_0 N_A\, dV \tag{19}$$

where

$$j_0 = \frac{j'}{\exp\left(-\frac{\Delta G_{ele}}{RT}\right) N_A\, dV}$$

and $dV = (4/3)\pi r_{con}^3$ is the volume of the contact sphere. We used $\mathbf{W}(r_{con}) \gtrsim \mathbf{W}(0) = j_0 N_A$. Since we are practically considering a contact radius, the determined j_0 will overestimate the theoretical value at zero distance. The free energy Δg_{cyc} calculated as the work w_{open} obtained by integrating the force F going from a closed to an extended configuration in n successive λ-windows:

$$-\Delta g_{cyc} = \Delta g_{open} = -w_{open} = -\int_{closed}^{open} F\, dq = -\frac{\sum_{\lambda=0}^{n} F_\lambda}{n} \tag{20}$$

The closed state is characterized by a contact distance r_{con} and the open state by the end-to-end distance R of the relaxed linear DNA fragment. Because of the rotational restriction on the fragment ends, the free energies Δg_{cyc} must be corrected to account for the missing entropy:

$$\Delta G_{cyc} = \Delta g_{cyc} - T\Delta S$$

$$T\Delta S = T[S_f - S_i] = RT[\ln(4\pi r_{con}{}^2) - \ln(4\pi R^2)] \quad (21)$$

The end segment repulsion free energy ΔG_{ele} is determined from the linear regression of the PMF fitted between 20% and 80% of the contour length l_c. We considered the cyclization of fragments from 90 bp to 300 bp in $C_{K^+} = 50$ mM.

2.5. DNA Minicircles. Our model was further applied to DNA minicircle constructs. We have investigated the energetics of bending linear DNA into a circle of various lengths and degrees of undertwisting or overtwisting, and relaxing into a supercoiled structure. The structures were characterized according to their writhe (Wr) and twist number (Tw) at several ionic concentrations. The property Tw indicates how many times the two strands wrap around each other (i.e., the number of turns) and Wr, which is specific to closed curves, measures the helix coiling upon itself. For topologically closed structures, the sum of those two properties is a constant positive integer that is referred as to the linking number (Lk):[64,65]

$$Lk = Tw + Wr \quad (22)$$

It implies that a DNA circle with a Lk value differing from $Lk_0 = Tw_0 = N/h$ in the corresponding relaxed linear structure (where N and h are the number of base pairs and the helical pitch, respectively) will develop a supercoiled conformation manifested by the partitioning of the imposed Lk into a new Tw and creation of Wr, to accommodate the introduced torsional strain:

$$\Delta Lk = \Delta Tw + Wr \quad (23)$$

A relaxed DNA circle that is at least a persistence length (~150 bp) in size, for instance, will have ~168 bp, corresponding to $Lk = Lk_0 = 16$ since the commonly accepted helical repeat of relaxed B-DNA is 10.5 bp per turn.

The helical density (σ) is conveniently used as a value normalized to the circle size:

$$\sigma = \frac{\Delta Lk}{Lk_0} = \sigma_{Tw} + \sigma_{Wr} = \frac{\Delta Tw}{Lk_0} + \frac{\Delta Wr}{Lk_0} \quad (24)$$

Relaxed minicircles strictly have $Wr_0 = 0$, but the choice of the bp midpoints as the vertices of the discretized chain used to numerically evaluate the writhe produces a residual writhe. That is because this definition of the curve does not coincide with the centerline of the cylinder enclosing the DNA structure. The twist and writhe were calculated according to Clauvelin et al. with the code GeoToC.[66] The helical turn repeat h was determined by fitting the autocorrelation profile of $(i - (i + 1))$ segments (every bp step) with the formula[28]

$$\left\langle \hat{\mathbf{l}}_i \cdot \hat{\mathbf{l}}_j \right\rangle_{|i-j|=1} = \exp\left(-\frac{n\|\mathbf{l}\|}{l_P}\right)\left[a + (1-a)\cos\left(\frac{2\pi\|\mathbf{l}\|}{h}\right)\right]$$

2.6. Back-Mapping. If CG models allow one to study larger systems over longer time scales, it comes at the cost of details. In cases where atomic insight is desired, there exist, however, methods to reconstruct detailed molecular structures, based on the coordinates of the CG sites.[67,68] Such a process is usually termed back-mapping, reverse transformation, or fine-graining. The generated detailed structures can then be used as equilibrated starting points for higher-resolution simulations, thus taking advantage of the preceding CG simulation as an equilibration phase otherwise not possible to perform at the

atomistic level. A back-mapping procedure typically consists of two stages: first, generating an atomistic structure from the CG site coordinates, then relaxing the atomistic structure. Back-mapping methods can be organized according to the relative importance granted to the first stage. In decreasing order, we find approaches having recourse to conformer structure databases,[69,70] molecular fragments,[71] geometric rules,[67,72] and random placement.[68] The choice of a method highly depends on the resolution of the CG model to backmap. The less detail a CG representation has, the greater efforts must be exerted on the relaxation stage, since many different atomistic structures can possibly be mapped to a single collection of CG sites. Our model is clearly a low-resolution model, with ~20 heavy atoms reunited in one CG site. We have chosen to apply the method of Rzepiela et al.,[68] combining random atom placement, tethering around the DNA CG sites, and simulated annealing (SA) with the Amber ff99 FF[40,41] to relax the initial configuration into a low-energy structure. Because of the high initial forces at the beginning of the SA simulations, it is necessary to inspect the chiral centers of the sugar moiety that might end up in the wrong absolute configuration. In order to reconstruct and maintain the proper Watson−Crick (WC) pairing and groove orientation together, we have inserted additional distance restraints between key atoms and exclusions of nonbonded interactions. The method has been validated by reconstructing the atomic details from a collection of CG mapped DNA structures selected from the Protein Data Bank (PDB).

3. RESULTS AND DISCUSSION

3.1. Persistence Length and Electrostatics. In our previous work,[36] we had concluded that plain cutoff electrostatics with a long enough cutoff was a safe alternative to the more demanding PME scheme. We nuance here these conclusions by exploring the behavior of numbers of other systems with shorter and longer DNA fragments (90−500 bp), for ionic concentrations up to 50 mM, again with both plain cutoff and PME electrostatics.

We have inspected a possible chain length dependence of l_P. Figure 2A (solid lines) shows that
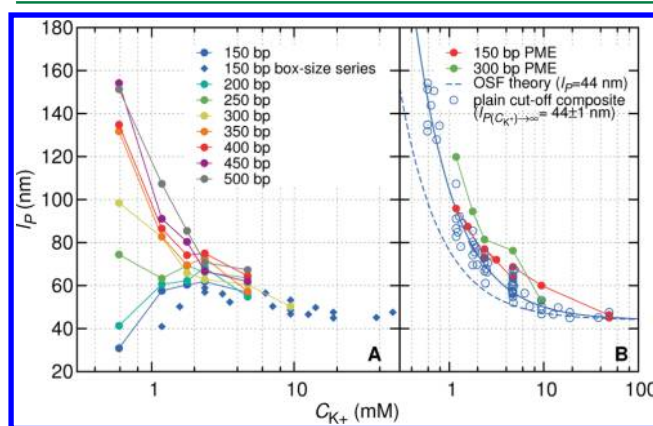
Figure 2. (A) Persistence length versus ion concentration series for DNA fragments of different sizes (150−500 bp) with plain cutoff treatment of electrostatics. (B) Reciprocal fit of a composite set of l_P values from plain cutoff simulations. The PME series for 150 and 300 bp are also drawn.

(1) the persistence length of 51 nm extracted for 300 bp fragments with 8 nm plain cutoff electrostatics has not reached the asymptotic value yet for C_{K^+} = 4.7 mM,

(2) there are huge variations between different fragment lengths for ion concentrations below ~4 mM, and

(3) shorter fragments (i.e., 150 and 200 bp) display a higher flexibility at lower concentrations.

To test a possible increased flexibility of smaller fragments due to chain-end effects, we have calculated l_P for 100 bp portions taken from the extremity and inward. It appears that chain ends are more flexible (l_P decreased by 11 nm at C_{K^+} = 1.18 mM with plain cutoff; see Figure S1 in the Supporting Information) but could not alone cause the dramatic 100 nm drop observed between long fragments and short fragments (Figure 2A).

We then envisaged a possible effect of the ion imbalance, the deviation of the ratio K^+/Cl^- from unity due to the presence of the polyanionic DNA chain and the small system size. If this ratio is close to 1 for high concentrations and all DNA fragment lengths (1.03 and 1.12 for 150 bp and 500 bp in C_{K^+} = 4.7 mM), at the lowest K^+ concentration, there is a large variation (from 1.33 to 6). We have performed series of simulations of a 150 bp fragment with five different box sizes (V_0, $V_0/2$, $3V_0/8$, $V_0/4$, $V_0/8$) and five different quantities of ions (n_{K^+} = 1200, 2400, 3600, 4800, and 9600, corresponding to K^+/Cl^- ratios of 1.33, 1.14, 1.09, 1.07, and 1.03, respectively). The corresponding l_P values are visible in Figures 2A (blue diamonds) and 3
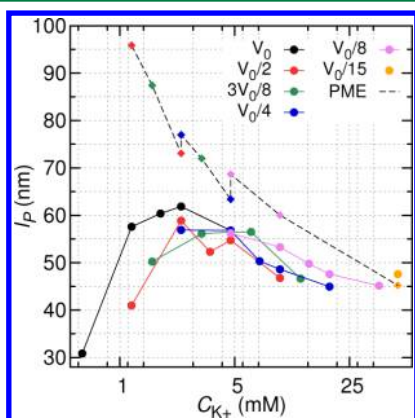

Figure 3. Ionic concentration dependence of the persistence length for different system sizes (V_0 = 150 nm³) of a 150 bp fragment. Solid curves are for plain-cut off simulations. The single data points for $V_0/15$ correspond to n_{K^+} = 6796 and K^+/Cl^- = 1.05. The standard deviations of l_P values vary from 11 nm to 3 nm, with a mean value of 5 nm.

(solid circles). The spread of l_P calculated at 4.7 mM K^+ from the five different box sizes (or K^+/Cl^- ratio) does not exceed 2 nm with a mean value of 56.2 nm (Figure 3). Values for l_P at 2.4 mM and 9.5 mM (three data points) have a spread of 5 and 7 nm and a mean value of 59.2 and 49.6 nm, respectively. The trend is, in substance, the same for every box size: l_P grows as C_{K^+} increases to 3−5 mM and then decays with a reciprocal law to the asymptotical high ionic concentration value, as expected and observed for fragments >250 bp.

The apparent chain length dependence observed at low ionic concentrations was eventually diagnosed as an artifact of the plain cutoff treatment of electrostatics. Indeed, performing simulations with PME on the 150 bp fragment systems restored

the proper reciprocal dependence of l_P with increasing ionic concentrations as seen from Figure 3 (dashed curve). However, a slight length dependence at low concentrations remains, as seen from a comparison of 150 bp and 300 bp fragments (Figure 2B). This might be attributed to chain-end effects since the conclusions about the increased l_P for DNA extremities also holds for PME simulations (l_P decreased by 20 nm at C_{K^+} = 1.18 mM; see Figure S1 in the Supporting Information). In the limit of large ionic concentrations, chain-end effects disappear: l_P values calculated with PME at C_{K^+} = 50 mM for seven fragments ranging from 90 bp to 150 bp virtually show the same values (mean, 45.5 nm; standard deviation, 0.6 nm). Moreover, l_P values determined with both electrostatics methods converge as the concentration increases (see Figure 3 (orange data points), as well as Figure S1 in the Supporting Information): at C_{K^+} = 50 mM, the PME value is 45.2 and 47.6 nm with plain cutoff. We have examined, in more detail, the parameters of the minimal electrostatics method giving a consistent flexibility (for a 90 bp fragment at C_{K^+} = 50 mM). A cutoff equal to that of the short-range effective potentials (r_{cut} = 3.2 nm) gives a much too high rigidity with a plain cutoff scheme and a coarse PME grid size (1.6 nm). Doubling and quadrupling the resolution (0.8 and 0.4 nm) brings a drastic but costly improvement. Instead, increasing the cutoff distance to 6.0 nm provides similar results independent of the method and PME resolution. Pushing the cutoff up to 8.0 nm causes an additional decrease in l_P by ca. 2 nm (Figure 4). From these
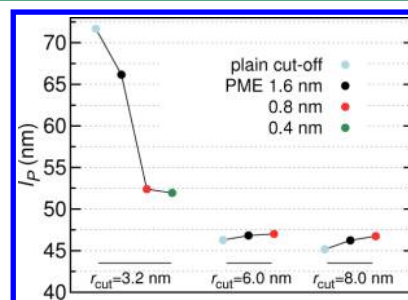
Figure 4. Persistence lengths of a 90 bp fragment in C_{K^+} = 50 mM with plain cutoff or PME treatment of electrostatics, with various cutoff radii and grid resolution.

results, it appears that the long-range detail of the charge distribution along the DNA chain is of prominent importance to render the proper flexibility of the chain. Clearly, the shortest cutoff, even with long-ranged electrostatics with a mesh size comparable to the base pair separation (0.4 nm), gives a chain that is too rigid (51.9 nm). Enough of these details seems to be encompassed within a larger cutoff of 6.0 nm, with little effect of the electrostatics scheme and grid resolution on the flexibility. It is an indication that, beyond this point, a more accurate representation of the bulk ions is not critical for the chain flexibility, at least for a 90 bp fragment and high ionic strength—indeed, a handful of DNA CG models treating the ionic solution implicitly as a continuum manage to provide relevant flexibility and its concentration dependence.[7,8,35] Also, the charge density in the bulk is so low (0.06 ion per nm³ at C_{K^+} = 50 mM) that even a truncated plain cutoff scheme grasps most of the electrostatic interactions.

We provide a definitive value of l_P = 44 nm for the *10pot* set of potentials with a plain electrostatic cutoff of 8.0 nm by fitting a composite data set gathering all calculated values for

fragments of length 350−500 bp and all values for shorter fragments with concentrations above 4 mM (Figure 2B), which is the range of composition where plain cutoff treatment seems to be valid.

Long fragments (400−500 bp) with the highest salt concentration studied (4.7 mM) tend to spontaneously collapse on themselves after a few hundred nanoseconds. (For those systems, we have therefore only measured l_P for configurations where DNA is linear.) Such an effect is quite unexpected, since monovalent cations do not cause DNA compaction in normal aqueous solution[73] (however, this is not true in a crowded environment[74,75]).

This artificial condensation also originates from the truncated treatment of the electrostatic interactions since switching to PME rapidly causes the aggregated structures to unfold. We hypothesize that, above a certain threshold ion concentration, and in the absence of long-ranged electrostatic repulsion, the attractive short-range forces due to the effective potentials overcome the dispersive thermal energy stirring the simulation box and particles start to aggregate locally, leading to the formation of the observed condensed structures. Regardless of the flawed origin of those observations, there are numerous experimental evidence of such ion-driven DNA aggregation in the case of multivalent cationic species. For instance, the polyamines spermine ($z = +4$) and spermidine ($z = +3$) are strong aggregation agents[76,77] as well as divalent cations (especially transition metals), but to a lesser extent.[78,79] We have identified three groups of resulting structures: plectonemes, "bretzel" knots, and "thread-in-eye" knots (see Figure 5). Sitko et al.[78] reported the formation of so-called *tennis*
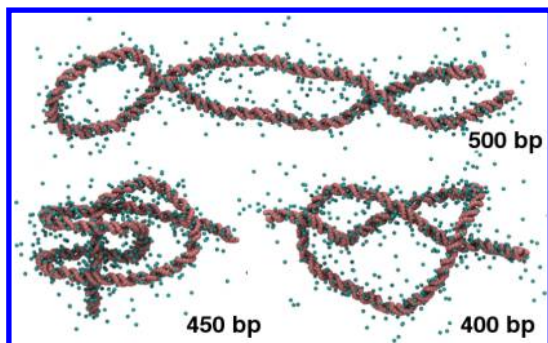

Figure 5. Representative structures of condensed DNA for C_{K^+} = 4.7 mM: (top) plectoneme, (left) thread-in-eye, (right) bretzel with display of the cations within a distance of 8 nm from DNA.

racquet looped structures for 800 bp DNA fragments in millimolar Ni^{2+} concentrations through a mechanism referred to as electrostatic zipper. Interestingly, we have observed such process types, as a prelude to the formation of the plectoneme structures: DNA adopts a highly curved hairpin conformation: one of the ends encounters the stem, both segments align coaxially and finally slide end to end. The plectoneme can further evolve toward a knotted structure for which the effective contact diameter of DNA equals the electrostatic cutoff (i.e., 8.0 nm). Note that because there is no attractive interactions between DNA CG sites, the only force dragging them close to each other is the DNA-K^+ interaction. Even if the driving force for the compactions we have observed is due to an artifact of the model, the ability of the DNA chain to form highly curved structures is encouraging for future use of the FF to model protein−DNA association and DNA condensation.

Therefore, we must complement and temper the precedent observation that a proper accounting of bulk ions is not crucial to determine the flexibility of DNA chains. Electrostatic interactions at both long and short distances give rise to complicated correlation effects and a truncated representation may bias such effects, which leads, as we have experienced, to an unexpected and erroneous collapse of DNA chains. From the abnormal phenomena that truncated electrostatics creates at both low (dramatic size-dependent flexibility) and high ionic concentrations (DNA aggregation), we consequently advise very cautious use of the plain cutoff scheme and we can only strongly encourage the use of the PME scheme. The introduced extra computational burden is moreover limited as the necessity of a large direct cutoff already demands about one-third of the computing workload through the neighbor search for pair list generation. A minimum direct cutoff of 6 nm with a grid size in the nanometer scale is advised (see Figure 4).

3.2. *j*-Factor. Since DNA flexibility is best characterized experimentally by the *j*-factor, we have determined these values directly from equilibrium simulations and from the ring-closure free energy. Because equilibrium simulations do not provide enough (any) sampling of the end-to-end contact conformations, the first method involves a functional fitting of the radial probability densities $\mathbf{Q}(r)$ of the end-to-end distance vectors with the WLC model interpolation formula of Becker et al.[63] (eq 16). $\mathbf{Q}(r)$ was evaluated for seven fragment sizes (90−150 bp) in C_{K^+} = 50 mM for and both FFs (*10pot* and *16pot*) for internal segments down to 50 bp in size. This method returns *j*-factors almost identical to those determined from eq 14 with l_P values obtained with the chain correlation exponential decay formula (eq 5) (see Figures 6 and 7).
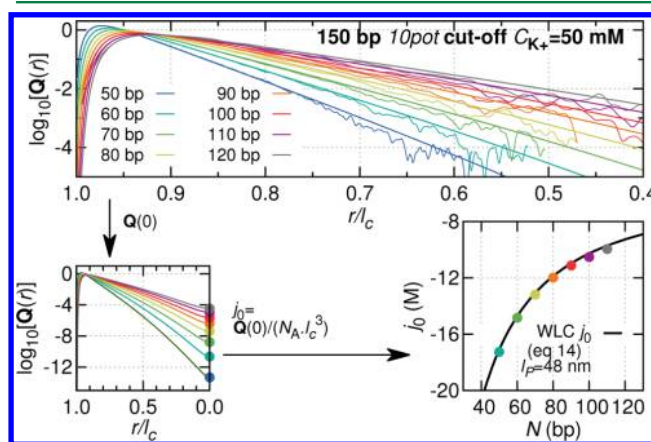

Figure 6. End-to-end distance radial probability densities for a 150 bp fragment (discarding the first and last 10 bp) and internal segments of smaller size with the *10pot* FF in C_{K^+} = 50 mM. The radial part j_0 of the *j*-factor is determined from $\mathbf{Q}(0)$ (eq 17).

In order to calculate the ring closure free energies, we conducted, for every size of DNA fragment, two independent sets of 51 simulations of 10 ns each with harmonic position restraints on the first and last residues of one strand, gradually moving their positions from a closed circle to a linear polymer. We have thus obtained the profiles for the restoring force exerted on the ends of fragments 90−300 bp in size in 50 mM K^+ for both the *10pot* and *16pot* sets of interaction potentials. The force was numerically integrated from the closed circle ($r_{con} \approx 1.15$ nm) up to the relaxed state ($R, F = 0$) to provide the PMF and thus the free energy of ring closure Δg_{cyc} (eq 20)
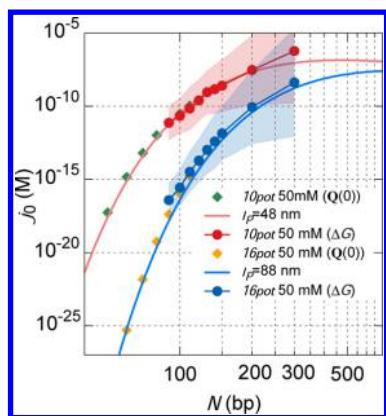
Figure 7. *j*-Factors calculated from the free energies of ring closure ΔG_{cyc} (restrained simulations) (circles) or from the extrapolated value of the end-to-end contact distance probability density $\mathbf{Q}(0)$ (equilibrium simulations) (diamonds) for the *10pot* and *16pot* FFs in C_{K^+} = 50 mM, plain cutoff and PME. The colored zones represent the uncertainties on the ΔG_{cyc} data. Radial components J_0 of the WLC *j*-factor (eq 14) are drawn as solid lines for the relevant l_P values.
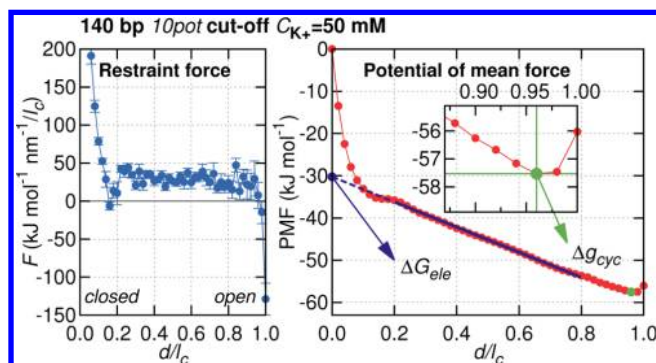
Figure 8. (left) Restraint force exerted on the first and last nucleotide of a 140 bp DNA fragment with the *10pot* FF in C_{K^+} = 50 mM as a function of the restrained distance *d* expressed as fraction of the contour length l_c. (right) The integrated force provides the PMF from which the minimum, at d/l_c = 0.96, corresponds to the free energy of ring closure Δg_{cyc}. The free energy associated with the electrostatic repulsion of the end segments ΔG_{ele} is obtained from the linear regression of the PMF. The *j*-factors obtained from eq 19, after entropy and electrostatics corrections, are j' = 2.5 × 10^{-14} M and j_0 = 1.9 × 10^{-9} M.

(see Figure 8). The error on the free energies were calculated by integrating the errors on each of the 51 individual average forces along the ring closing path. These errors were themselves determined by the block averaging method.[80] The missing configurational entropy was added to Δg_{cyc} (eq 21) to provide ΔG_{cyc} and, thus, j' (eq 19). A last operation to remove the contribution from electrostatic repulsions gives the final set of j_0, which is plotted together with the theoretical values for the wormlike chain model (WLC) (eq 14) with l_P = 48 and 88 nm, which are the values determined from chain orientation correlation (see $V_0/15$ in Figure 3) (Figure 7). As expected, the determined values overestimate the actual value of j_0, because of the approximation in eq 19. We did not perform these restrained simulations with PME electrostatics, since the one end moving away from the other will gradually again feel its presence, because of the periodic conditions.

If the equilibrium method involves a model fitting, the free-energy method has the advantage of being parameter-free.

Since the uncertainty on the determined values is large (see the shaded areas in Figure 7), we have also calculated the *j*-factors for the *16pot* FF, to assess and verify the accuracy of the method.

3.3. DNA Minicircles. We have inspected both structural and energetics aspects of DNA minicircles. Figures 9A and 9B show the static mechanical energy surfaces for bending and twisting linear DNA into a closed circular structure and for the subsequent relaxation through the formation of superhelical structures. The starting circular structures have an initial twist Tw equal to their linking number Lk = N/10 + δ, where N is the number of bp (90−500 bp) and δ is the number of turns added or removed before closure (−5 to +5). The strain energy is taken as the difference between the tabulated bond potential energies of the regular plane circular and linear structures with an helical pitch of 10 bp. The relaxation energy is the difference between the bond energy of the relaxed circular structure and the strain energy minus the bond energy of the relaxed linear fragment. The relaxed energies were averaged over the 20 last ns of 25 ns simulations. All minicircles rapidly relax (<500 ps) into supercoiled structures. To avoid self-crossing of the polymer chain (and, thus, changes in Lk) during the release of the initial high stress energy, a $1/r^6$ repulsion potential between DNA beads (V_{rep}(1.3 nm) ≃ RT) was added. The smaller minicircles (90−110 bp) should normally partly dehybridize and produce kinks for the underwound (δ < −1) and highly overwound topoisomers (δ > +3)[81−83] but our CG model is not designed to render such processes, since it constrains DNA to a helical B-like state (such a permissive model can be found, e.g., in ref 27). An extreme example is the 90 bp δ = +5 minicircle, which is so frustrated energetically that it remains strictly circular (see Figure 9B). A wide variety of configurations was generated: plectonemes with 2−5 loops, as well as three-, four-, or five-lobed interwound, and warped circles. Those structures have been classified according to their overall shape (**EC** = ellipsoid-circular, **WC** = warped circle, **P** = plectoneme, **L** = lobed), the number of primary crossings (ranging from 2 to 5), and secondary crossings (denoted as "+" and "++"). The inset structures in Figures 9B and C show the observed representative supercoiled folds with their classification code (see Figure S2 in the Supporting Information for the complete topology/structure assignment tables). The **3L**, **4L**, and **5L** lobed-type structures display the shape of a trefoil, tetrahedron, and square pyramid, respectively. The "+" symbols indicate that one or several lobes/ears at the apexes of the aforementioned structures have additional crossings.

We have examined the degree of supercoiling of the minicircles more quantitatively by calculating the average writhe ⟨Wr⟩ over the 20 last ns. The ionic concentration has a clear effect on the writhe of large minicircles (200−500 bp) with positive ΔLk. For those topoisomers, the writhe is 5%−20% higher at 50 mM K⁺, compared to minimal salt conditions (MSC) (Figure 10A). Structurally, such trend translates into more supercoiled structures with additional lobes and crossings (see Figure S2 in the Supporting Information for the complete topology/writhe assignment table). By inspecting the components of the helical density (eq 24), we can compare how it is distributed for different sizes of minicircle as a function of the total helical density σ (Figure 10A). Small circles first exclusively absorb the change in linking number by modifying the twist (Figure 10B), resulting in a warped wheel structure (100, 200, 400 bp). Additional stress then contributes to the writhe that displays a jump and rapidly reaches a plateau for
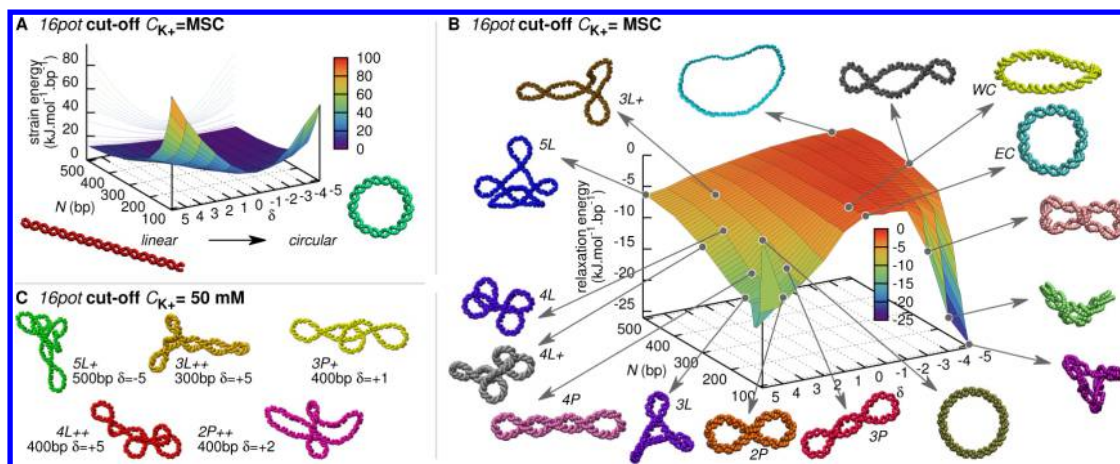
Figure 9. (A) Mechanical energy for bending and twisting linear DNA fragments into closed circles and (B) relaxing the structures. The interaction network used is *16pot* with minimal salt conditions (MSC). Inset structures with classification codes (see text) represent typical supercoiled fold. (C) Structures (with codes) only found with the higher salt concentration C_{K^+} = 50 mM.
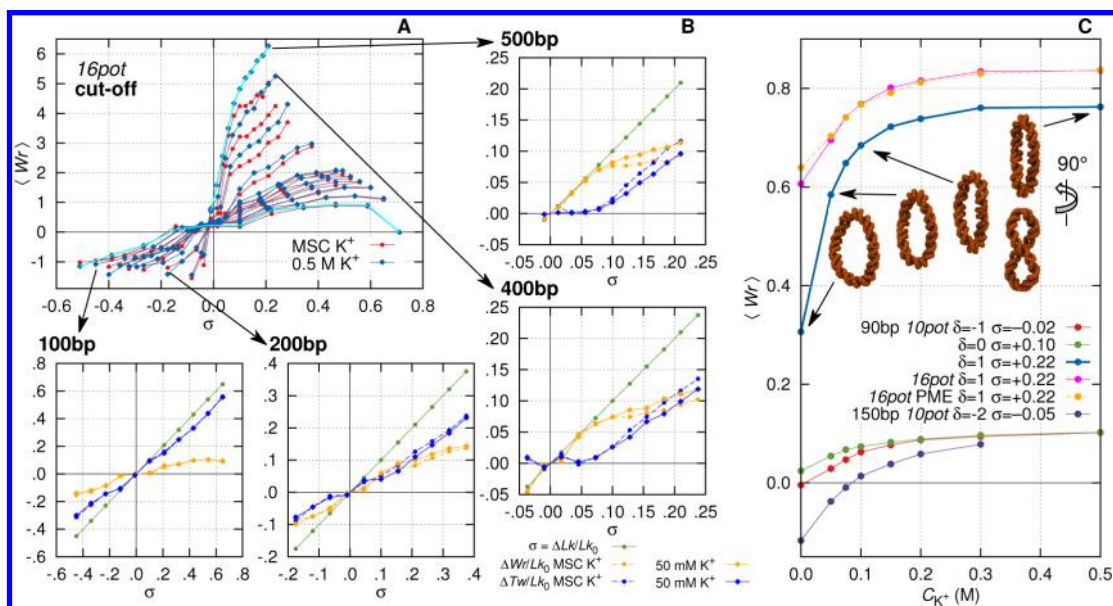
Figure 10. (A) Average writhe $\langle Wr \rangle$, with respect to the helical density σ (h = 11 bp), for various sizes of the minicircle and two ionic concentrations. (B) Tw and Wr components of the helical density (σ) for selected minicircles. (C) Evolution of Wr with the ionic concentration for different σ values.

minicircle up to 140 bp. This initial jump coincides with the buckling transition resulting in a figure-8 plectoneme. For larger circles (150−400 bp), the initial twisting regime is superseded by a higher growth in writhe that even absorbs some of the accumulated twist (300−500 bp). For the 500 bp minicircle, all the change in linking number is converted to writhe until, as for all circles, a plateau corresponding to a compact supercoiled structure is reached and only the twist continues to grow. Similarly, undertwisting causes the same effect with opposite sign. If evidence of the nonlinear response in torsional stiffness of DNA has been reported both in experiments[84,85] and simulations,[28,86] the low resolution in σ does not permit us to observe any discernible asymmetry in the partitioning of Tw and Wr. The buckling transition occurs around $|\sigma_{Wr}| \simeq 0.05$. Both *10pot* and *16pot* FFs display very similar writhe and twist density partitioning as a function of the minicircle size, indicating that the ratio of their bending and torsional characteristic lengths are comparable.

In order to expose the salt-induced buckling transition of minicircles observed experimentally in the 10^{-1} M regime for underwound structures (178 bp, $\Delta Lk = -1,-2$, $\sigma = -0.06,-0.12$)[87] and reproduced in implicit solvent AA MD simulations,[82,88] we have conducted a series of simulations with increasing salt concentration up to 500 mM. We have not observed any buckling characterized by negative writhing for the slightly underwound structures ($\sigma = -0.02$ (90 bp), -0.05 (150 bp), -0.07 (200 bp), lower σ are already buckled at MSC). In contrast, all minicircles grow positive writhing with increasing salt concentration. The effect of salt is more visible on slightly overwound circles (Figure 10C). The 90 bp minicircle with $\delta = +1$ ($\sigma = +0.22$), which is already slightly buckled at MSC, forms a flatter and flatter figure-8 plectoneme as the concentration increases. It is visible from the inset structures in Figure 10C, where the distance between nucleotide i and $i + N/2$ at the crossing point becomes smaller

and smaller. Plain cutoff or PME electrostatic treatments quantitatively give the same results.

3.4. Back-Mapping of DNA Minicircles. The approach of Rzepiela et al.[68] has the advantage of being thoroughly applicable to any type of structure-based CG model. The method is implemented in a modified version of GROMACS and takes as input the topology file of the CG structure to backmap simply composed of the CG sites list, the conventional topology file of the atomistic structure to reconstruct with an extra "mapping" section including the correspondence of atoms and CG sites, the CG structure file, and a starting atomistic structure. An initial atomistic structure with the atoms related to a CG site randomly distributed within a sphere of radius 0.3 nm around the CG site coordinates is generated (step **0**). All these atoms are harmonically restrained to the position of their corresponding CG site with a force constant K = 1000 kJ mol^{-1} nm^{-2}. Having the method coupled to GROMACS allows one to choose between the different FF libraries available to construct the needed atomistic topology. It is also straightforward to provide additional geometrical restraints and switch off nonbonded interactions between selected atoms. The atomistic topology files of the minicircles were created with GROMACS from linear DNA structures and the extremities manually sealed by adding the missing FF terms.

A sequential scheme able to reliably produce well-behaved atomistic structures of supercoiled minicircles proceeds as follows. The very first step consists of relaxing without restraints (except the tethering to the CG sites) the random atomistic structure into a raw molecular system (step **1**). The SA scheme consists of a 60 ps stage at 1800 K, followed by 20 ps of cooling to 1200 K and 20 ps of cooling down to 300 K. A stochastic dynamics integrator with a time step of 2 ps is used to best distribute the heat across the entire DNA structure in a solvent-free environment. We consider the same triple cutoff for the nonbonded interaction treatment as in ref 68. The output structure displays nucleotides pointing in all directions in and out of the DNA helical structure. In addition, a large share of the sugar–phosphate backbone chiral centers have ended up in the wrong absolute configurations. Therefore, we use the chirality plug-in[89] of the molecular visualization program VMD[90] to check and correct the configuration of the C1′, C3′, C4′ stereocenters (step **2**). We also fix the arrangement of the diastereotopic methylene hydrogens of C2′ (H2′1, H2′2) and C5′ (H5′1, H5′2). Even though the latter fix is not crucial, since swapping these hydrogens does not result in any FF energy difference, it is convenient to have consistent atom naming for structural analyses. The diastereotopic nonbridging oxygens bound to phosphorus (O1P, O2P) are also fixed, for the additional reason that O2P is involved in a distance restraint determining the proper groove orientation. The resulting structure is then used as the starting point for successive SA simulations with different restraints and nonbonded interactions exclusions. We consider a set of groove (GR) restraints enforced through harmonic distance restraints to force every nucleotide to adopt an *anti* conformation. A proper intrahelical stacked and paired state is promoted by a set of stacking (ST) restraints between successive nucleotides along one strand and WC restraints. The force constants associated with the distance restraints are given in Table 1. At first, only the WC restraints between every second bp are removed (WChalf). To facilitate rotation and flipping of the nucleobase thymine, the nonbonded interactions between the bulky methyl group (C7) and N1/N9, C1′, C2′, C4′, O4′

Table 1. Groove (GR), Watson-Crick (WC), and Stacking (ST) Distance Restraints

group	atom pair	distance (nm)	K (kJ mol^{-1} nm^2)
GR (i–i)	O2P–H8$_A$	0.29	1000
	O2P–H5$_C$	0.38	1000
	O2P–H8$_G$	0.29	1000
	O2P–C7$_T$	0.37	1000
WC (i–j)	N1$_A$–N3$_T$	0.30	1000
	N1$_G$–N3$_C$	0.29	1000
	N6$_A$–O4$_T$	0.28	500
	H2$_A$–O2$_T$	0.32	500
	N2$_G$–O2$_C$	0.30	500
	O6$_G$–N4$_C$	0.27	500
ST (i–(i+1))	C5–C5	0.41	2000

atoms are disrupted (step **3**). The second half of the WC restraints is then imposed (step **4**). The GR restraints and C7 exclusions are further removed in two consecutive SA runs (steps **5** and **6**). A restraint-free energy minimization is performed between every SA simulation. A final 200 ps equilibration step at 300 K with a reduced CG site tethering of K = 200 kJ mol^{-1} nm^{-2} concludes the procedure (step **7**), as summarized in Table 2. Figure 11A depicts structures from the

Table 2. Back-Mapping Procedure

step	restraints[a,b]	nonbonded exclusions[c]
0	random placement	
1	relaxation without restraints	
2	chirality check	
3	GR+WChalf+ST	C7
4	GR+WC+ST	C7
5	WC+ST	C7
6	WC+ST	
7	200 ps WC+ST equilibration at 300 K	

[a]Legend: GR = groove, WC = Watson−Crick, ST = stacking. [b]In addition to the position restraints at the CG sites. [c]In addition to those originating from bonds defined in the FF.

random placement to the last reconstructed state of an ideal 90 bp minicircle with Lk = 9. The idealized structure shown as a reference was generated with AmberTools.[91] The O2P, C7, and electronegative sites pointing to the major groove are highlighted to follow the improvement in the quality of the reconstruction. The same reconstruction scheme was successfully applied to the final structures of selected CG MD simulations (Figure 11B), with both a random sequence and a 15 bp A-tract repeating sequence $(A_4T_4(CG)_3C)_6$. The atomistic structures mapped back onto pre-equilibrated CG coordinates can find a certain interest as a starting point to study how DNA denaturates under bending and torsional stress as well as to apprehend the influence of the sequence on its stability, in the manner of the recent work of Harris and co-workers.[83,88] If the robustness of the procedure is obvious from its ability to properly reconstruct molecular details at the level of base pairs for distorted minicircles, its accuracy has also been validated by inspecting the root-mean-square deviations of the heavy atoms positions of back-mapped coarse-grained structures with respect to their high-resolution X-ray diffraction coordinates from the PDB. By adapting the constraints, the
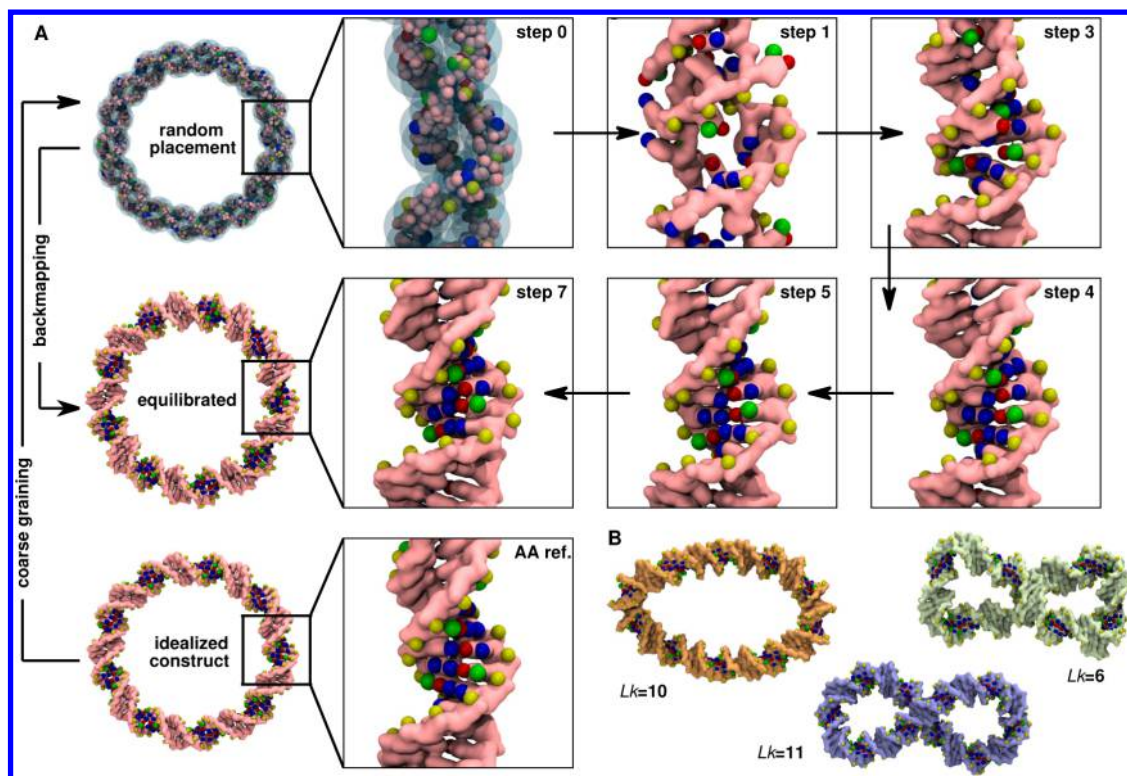
Figure 11. (A) Reconstruction procedure based in Table 2 applied to a flat circular 90 bp minicircle with Lk = 9. (B) Reconstructed 90 bp minicircles with Lk = 10 (orange), Lk = 11 (blue), and Lk = 6 (green). Electronegative sites pointing to the major grooves are highlighted as red (oxygen) and blue (nitrogen) spheres. The thymine methyl group C7 is shown in green, and nonbridging phosphate oxygen O2P is shown in yellow.

method is readily applicable to A-DNA, kinked, and opened structures (see the Supporting Information).

4. CONCLUSION AND PERSPECTIVES

We have performed coarse-grained (CG) molecular dynamics (MD) simulations on various systems consisting of linear and circular DNA fragments of various sizes with explicit KCl salt at different concentrations. The model used is a structure-based CG force field constructed from the set of distance distributions extracted from high-quality all-atom MD trajectories of multisequence 18 bp DNA fragments. The set of effective solvent mediated potentials was derived using a systematic iterative inverse Boltzmann−Newton inversion method. Here, this model, which has been detailed in a previous paper,[36] was further applied to a wider range of DNA systems. By calculating the persistence length for different fragment sizes with a plain cutoff electrostatics scheme, we have identified two artifacts:

(1) At low concentrations, smaller fragments are more flexible and the smallest fragments considered are also more flexible than at higher concentrations.
(2) At high concentrations, there is a collapse of longer fragments.

These observations have convinced us of the importance of a long-ranged electrostatics scheme when dealing with poly-electrolytes and explicit ions, since using the Particle Mesh Ewald (PME) scheme restored the proper expected length-independent flexibility of DNA and prevents unphysical aggregation. We concluded on the safe and good practice of using PME with a minimal direct cutoff of 6 and a grid resolution of 1.6 nm for reciprocal space. We further characterized the flexibility of DNA fragments by calculating the ring closure or cyclization probability, called the *j*-factor,

with two other methods. We extracted these values by fitting the radial probability density of the end-to-end distance obtained from equilibrium simulations with an interpolation formula for the wormlike chain model. The other method, which does not involve any assumed flexibility model, consists of calculating the free energy of ring closure from a series of restrained MD simulations. The results obtained from the different methods are self-consistent. After cyclization, we naturally approached small circular DNA constructs. We calculated the strain and relaxation energies for a range of minicircles of different sizes and helical densities. The supercoiled circles exhibit a large variety of plectoneme and lobed structures for which we computed the writhe number. Increasing the ionic concentration has the effect of producing more-compact structures with an increased writhe component. However, we did not observe ion-concentration-driven buckling of negatively writhed minicircles. We then managed to set up a robust sequential procedure to reconstruct molecular details of arbitrary sequences into the supercoiled CG structures of various DNA minicircles. These reconstructed structures can be used a starting material for studies of minicircles in explicit solvent and give access to denaturation and sequence-dependent phenomena that are not within the reach of the CG model.

This comprehensive study, which has shown the capabilities and limitations of our DNA-ions CG model, has confirmed, to us, the relevance of combining it with a protein model to gain insight into, for example, nucleosome core particles aggregation, DNA wrapping around histones, and, generally, any system where electrostatic effects are significant and an explicit representation of charged entities is essential.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Persistence length along a 300 bp DNA fragment for different ionic concentrations and electrostatics treatment. Tables with the structure types, writhe, and writhe density for minicircles of various sizes and helical density. Validation of the back-mapping procedure by reconstructing atomic details into coarse-grained DNA structures derived from the Protein Data Bank. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00113.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: aymeric.naome@unamur.be.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Noid, W. G. J. Chem. Phys. 2013, 139, No. 090901.

(2) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. Wiley Interdiscip. Rev.: Comput. Mol. Sci. 2014, 4, 225−248.

(3) Kamerlin, S. C. L.; Warshel, A. Phys. Chem. Chem. Phys. 2011, 13, 10401−10411.

(4) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. J. Chem. Phys. 2011, 134, No. 085101.

(5) Šulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. J. Chem. Phys. 2012, 137, No. 135101.

(6) Ouldridge, T. E. Coarse-Grained Modelling of DNA and DNA Self-Assembly; Springer Thesis; Springer−Verlag: Berlin, 2012; p 176.

(7) Knotts, T. A., IV; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. J. Chem. Phys. 2007, 126, No. 084901.

(8) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. Biophys. J. 2009, 96, 1675−1690.

(9) Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. J. Chem. Phys. 2011, 135, No. 165104.

(10) DeMille, R. C.; Cheatham, T. E., III; Molinero, V. J. Phys. Chem. B 2011, 115, 132−142.

(11) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. J. Chem. Phys. 2013, 139, No. 144903.

(12) Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J. J. Chem. Phys. 2014, 141, No. 035102.

(13) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. J. Chem. Theory Comput. 2010, 6, 1711−1725.

(14) Machado, M. R.; Dans, P. D.; Pantano, S. Phys. Chem. Chem. Phys. 2011, II, 1−2.

(15) Zeida, A.; Machado, M. R.; Dans, P. D.; Pantano, S. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. 2012, 86, No. 021903.

(16) Dans, P. D.; Darré, L.; Machado, M. R.; Zeida, A.; Brandner, A. F.; Pantano, S. In Series on Advances in Bioinformatics and Computational Biology; Lecture Notes in Computer Science, Vol. 8213; Setubal, J. C., Almeida, N. F., Eds.; Springer International Publishing: Cham, Swizterland, 2013; pp 71−81.

(17) Linak, M. C.; Tourdot, R.; Dorfman, K. D. J. Chem. Phys. 2011, 135, No. 205102.

(18) Edens, L. E.; Brozik, J. A.; Keller, D. J. J. Phys. Chem. B 2012, 116, 14735−14743.

(19) Cragnolini, T.; Derreumaux, P.; Pasquali, S. J. Phys. Chem. B 2013, 117, 8047−8060.

(20) Freeman, G. S.; Lequieu, J. P.; Hinckley, D. M.; Whitmer, J. K.; de Pablo, J. J. Phys. Rev. Lett. 2014, 113, No. 168101.

(21) Ouldridge, T. E. Mol. Phys. 2015, 113, 1−15.

(22) Soper, A. K. Chem. Phys. 1996, 202, 295−306.

(23) Reith, D.; Pütz, M.; Müller-Plathe, F. J. Comput. Chem. 2003, 24, 1624−1636.

(24) Lyubartsev, A. P.; Laaksonen, A. Phys. Rev. E: Stat., Phys., Plasmas, Fluids, Relat. Interdiscip. Top. 1995, 52, 3730−3737.

(25) Savelyev, A.; Papoian, G. A. Biophys. J. 2009, 96, 4044−4052.

(26) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. Soft Matter 2013, 9, 2108−2119.

(27) Trovato, F.; Tozzini, V. J. Phys. Chem. B 2008, 112, 13197−13200.

(28) Sayar, M.; Avşaroğlu, B.; Kabakçıoğlu, A. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. 2010, 81, 041916.

(29) Maffeo, C.; Ngo, T. T. M.; Ha, T.; Aksimentiev, A. J. Chem. Theory Comput. 2014, 10, 2891−2896.

(30) Lyubartsev, A. P.; Laaksonen, A. Comput. Phys. Commun. 1999, 121−122, 57−59.

(31) Korolev, N.; Luo, D.; Lyubartsev, A. P.; Nordenskiöld, L. Polymers 2014, 6, 1655−1675.

(32) He, Y.; Maciejczyk, M.; Ołdziej, S.; Scheraga, H. A.; Liwo, A. Phys. Rev. Lett. 2013, 110, 098101.

(33) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. J. Chem. Theory Comput. 2014, 10, 5020−5035.

(34) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. J. Comput. Chem. 2010, 31, 1644−1655.

(35) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. J. Chem. Phys. 2010, 132, 035105.

(36) Naôme, A.; Laaksonen, A.; Vercauteren, D. P. J. Chem. Theory Comput. 2014, 10, 3541−3549.

(37) Savelyev, A.; Papoian, G. A. Proc. Natl. Acad. Sci. U. S. A. 2010, 107, 20340−20345.

(38) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. J. Chem. Theory Comput. 2009, 5, 3211−3223.

(39) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. Nucleic Acids Res. 2010, 38, 299−313.

(40) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J. Am. Chem. Soc. 1995, 117, 5179−5197.

(41) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. J. Biomol. Struct. Dyn. 1999, 16, 845−862.

(42) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Biophys. J. 2007, 92, 3817−3829.

(43) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. J. Phys. Chem. 1987, 91, 6269−6271.

(44) Dang, L. X. J. Am. Chem. Soc. 1995, 117, 6954−6960.

(45) Mirzoev, A.; Lyubartsev, A. P. J. Chem. Theory Comput. 2013, 9, 1512−1520.

(46) Izvekov, S.; Voth, G. A. J. Chem. Phys. 2005, 123, No. 134105.

(47) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. J. Chem. Phys. 2008, 128, No. 244115.

(48) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J. Chem. Theory Comput. 2008, 4, 435−447.

(49) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Proc. Natl. Acad. Sci. U. S. A. 1997, 94, 6185−6190.

(50) Hagerman, P. J. Annu. Rev. Biophys. Biophys. Chem. 1988, 17, 265−286.

(51) Kratky, O.; Porod, G. *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 1106−1122.

(52) Peters, J. P.; Maher, L. J. Q., III. *Rev. Biophys.* **2010**, *43*, 23−63.

(53) Jacobson, H.; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600−1606.

(54) Levene, S. D.; Crothers, D. M. *J. Mol. Biol.* **1986**, *189*, 61−72.

(55) Wang, J. C.; Davidson, N. *J. Mol. Biol.* **1966**, *15*, 111−123.

(56) Shimada, J.; Yamakawa, H. *Macromolecules* **1984**, *17*, 689−698.

(57) *j*-Factor calculation for homogeneous wormlike chain model; available via the Internet at: http://crab.chem.nyu.edu/jfactor/, accessed Aug. 16, 2014.

(58) Shore, D.; Langowski, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **1981**, *78*, 4833−4837.

(59) Shore, D.; Baldwin, R. L. *J. Mol. Biol.* **1983**, *170*, 957−981.

(60) Taylor, W. H.; Hagerman, P. J. *J. Mol. Biol.* **1990**, *212*, 363−376.

(61) Vologodskaia, M.; Vologodskii, A. V. *J. Mol. Biol.* **2002**, *317*, 205−213.

(62) Du, Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. V. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 5397−5402.

(63) Becker, N. B.; Rosa, A.; Everaers, R. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2010**, *32*, 53−69.

(64) White, J. H. *Am. J. Math.* **1969**, *91*, 693−728.

(65) Fuller, F. B. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 815−819.

(66) Clauvelin, N.; Olson, W. K.; Tobias, I. *J. Chem. Theory Comput.* **2012**, *8*, 1092−1107.

(67) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. *J. Chem. Theory Comput.* **2014**, *10*, 676−690.

(68) Rzepiela, A. J.; Schäfer, L. V.; Goga, N.; Risselada, H. J.; de Vries, A. H.; Marrink, S. J. *J. Comput. Chem.* **2010**, *31*, 1333−1343.

(69) Shih, A. Y.; Freddolino, P. L.; Sligar, S. G.; Schulten, K. *Nano Lett.* **2007**, *7*, 1692−1696.

(70) Stansfeld, P. J.; Sansom, M. S. *J. Chem. Theory Comput.* **2011**, *7*, 1157−1166.

(71) Hess, B.; Leon, S.; van der Vegt, N.; Kremer, K. *Soft Matter* **2006**, *2*, 409−414.

(72) Brocos, P.; Mendoza-Espinosa, P.; Castillo, R.; Mas-Oliva, J.; Piñeiro, A. *Soft Matter* **2012**, *8*, 9005−9014.

(73) Wilson, R. W.; Bloomfield, V. A. *Biochemistry* **1979**, *18*, 2192−2196.

(74) Livolant, F.; Leforestier, A. *Prog. Polym. Sci.* **1996**, *21*, 1115−1164.

(75) Zinchenko, A. A.; Yoshikawa, K. *Biophys. J.* **2005**, *88*, 4118−4123.

(76) Pelta, J.; Livolant, F.; Sikarov, J.-L. *J. Biol. Chem.* **1996**, *271*, 5656−5662.

(77) Raspaud, E.; Olvera de la Cruz, M.; Sikorav, J. L.; Livolant, F. *Biophys. J.* **1998**, *74*, 381−393.

(78) Sitko, J. C.; Mateescu, E. M.; Hansma, H. G. *Biophys. J.* **2003**, *84*, 419−431.

(79) Hackl, E. V.; Kornilova, S. V.; Blagoi, Y. P. *Int. J. Biol. Macromol.* **2005**, *35*, 175−191.

(80) Hess, B. *J. Chem. Phys.* **2002**, *116*, 209−217.

(81) Lankas, F.; Lavery, R.; Maddocks, J. H. *Structure* **2006**, *14*, 1527−1534.

(82) Harris, S. A.; Laughton, C. A.; Liverpool, T. B. *Nucleic Acids Res.* **2008**, *36*, 21−29.

(83) Mitchell, J. S.; Laughton, C. A.; Harris, S. A. *Nucleic Acids Res.* **2011**, *39*, 3928−3938.

(84) Selvin, P. R.; Cook, D. N.; Pon, N. G.; Bauer, W. R.; Klein, M. P.; Hearst, J. E. *Science* **1992**, *255*, 82−85.

(85) Bryant, Z.; Stone, M. D.; Gore, J.; Smith, S. B.; Cozzarelli, N. R.; Bustamante, C. *Nature* **2003**, *424*, 338−341.

(86) Kannan, S.; Kohlhoff, K.; Zacharias, M. *Biophys. J.* **2006**, *91*, 2956−2965.

(87) Bednar, J.; Furrer, P.; Stasiak, A.; Dubochet, J.; Egelman, E. H.; Bates, A. D. *J. Mol. Biol.* **1994**, *235*, 825−847.

(88) Mitchell, J. S.; Harris, S. A. *Phys. Rev. Lett.* **2013**, *110*, 148105.

(89) Schreiner, E.; Trabuco, L. G.; Freddolino, P. L.; Schulten, K. *BMC Bioinf.* **2011**, *12*, 190.

(90) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

(91) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T. E., III; Darden, T.; Duke, R.; Gohlke, H.; Goetz, A.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.; Paesani, F.; Ro, D.; Roitberg, A.; Sagu, C.; Salomon-Ferrer, R.; Seabr, G.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Kollman, P. *AMBER 14*; University of California: San Francisco, CA, 2014.