

# Coarse-Grained Ions for Nucleic Acid Modeling

Daniel M. Hinckley<sup>†</sup> and Juan J. de Pablo<sup>\*‡¶</sup>

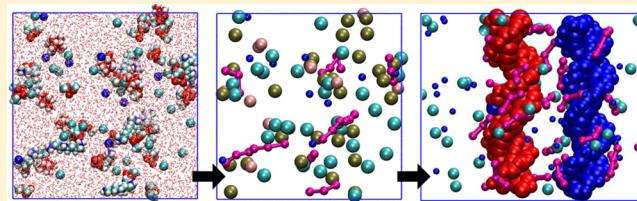
<sup>†</sup>Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, Wisconsin 53706, United States

<sup>‡</sup>Institute for Molecular Engineering, University of Chicago, Chicago, Illinois 60637, United States

<sup>¶</sup>Materials Science Division Argonne National Laboratory, Argonne, Illinois 60439, United States

## Supporting Information

**ABSTRACT:** We present a general coarse-grained model of sodium, magnesium, spermidine, and chlorine in implicit solvent. The effective potentials between ions are systematically parametrized using a relative entropy coarse-graining approach [Carmichael, S. P. and M. S. Shell, *J. Phys. Chem. B*, **116**, 8383–93 (2012)] that maximizes the information retained in a coarse-grained model. We describe the local distribution of ions in the vicinity of a recently published coarse-grained DNA model and demonstrate a dependence of persistence length on ionic strength that differs from that predicted by Odijk–Skolnick–Fixman theory. Consistent with experimental observations, we show that spermidine induces DNA condensation whereas magnesium and sodium do not. This model can be used alongside any coarse-grained DNA model that has explicit charges and an accurate reproduction of the excluded volume of dsDNA.



of persistence length on ionic strength that differs from that predicted by Odijk–Skolnick–Fixman theory. Consistent with experimental observations, we show that spermidine induces DNA condensation whereas magnesium and sodium do not. This model can be used alongside any coarse-grained DNA model that has explicit charges and an accurate reproduction of the excluded volume of dsDNA.

## 1. INTRODUCTION

Ions are critical to biology, creating potential gradients across cell membranes and acting as active sites for redox chemistry in enzymes. Ions also bind transiently to biological molecules, often modifying their biological activity. This is especially evident in double-stranded DNA (dsDNA), where ions have been implicated in minor groove narrowing,<sup>1</sup> dsDNA toroid formation,<sup>2</sup> and the stabilization of G-quartets.<sup>3</sup> Recently, multivalent ions have been shown to produce heterogeneous dynamics during viral packaging.<sup>4,5</sup> A correct representation of both DNA and its local ionic environment is necessary to predict biological form and function. The local ionic environment can be discerned by X-ray diffraction;<sup>1</sup> however, the crystals needed for such studies may not be representative of the ionic environment in solution. Solution NMR can resolve the local ionic distribution surrounding dsDNA;<sup>6</sup> however, it lacks the resolution of X-ray crystallography. The difficulty of experimentally querying the local environment motivates the use of molecular models and simulations.

All-atom (AA) simulations are well-suited for studying the local ionic environment. In contrast to mean-field approaches such as the Debye–Hückel (DH) theory,<sup>7</sup> Manning condensation,<sup>8</sup> or numerical solutions to the Poisson–Boltzmann (P–B) equation,<sup>9</sup> AA simulations can capture fluctuations of the local ionic environment in the vicinity of a biological molecule. Numerous simulation studies have examined this behavior.<sup>10–13</sup> However, the predictions of such studies often differ, as the properties of ions are very sensitive to the details of the force field. Recent efforts have sought to improve the ion parameters employed in these force fields.<sup>14–16</sup> Joung and Cheatham adjusted the Lennard-Jones (LJ) parameters between ions to eliminate salt crystals that formed in the

vicinity of biomolecules.<sup>14</sup> Subsequently Yoo and Aksimentiev improved the LJ parameters for a number of ion pairs to eliminate unphysical aggregation of dsDNA oligomers.<sup>15</sup> These recent improvements suggest that simulations with prior AA force fields might have incorrectly predicted the local ionic environment surrounding DNA. Additionally, AA simulations have in general been limited to relatively short time and length scales.

Coarse-grained (CG) models of DNA and ions have the potential to access the length and time scales relevant to studies of truly macromolecular DNA. A variety of CG models of DNA have been proposed in the literature;<sup>17–29</sup> the discussion here focuses on dsDNA models that include explicit ions. The SIRAH CG DNA model<sup>21</sup> was coupled with the WT4 CG water/ions model and demonstrated ion-dependent minor groove narrowing.<sup>30</sup> The 3SPN CG DNA model<sup>17</sup> was supplemented with ions to study the melting of small molecule–DNA hybrid dimers;<sup>31</sup> however, no details regarding the ion parametrization were provided. A detailed parametrization of the 3SPN.1 model was subsequently presented<sup>32</sup> with analytical ion–ion pair potentials that model the distance-dependent dielectric, hydration barriers, and excluded volume. Two one-site-per-nucleotide models<sup>20,27</sup> have been parametrized with ions using systematic coarse-graining methods (Molecular Renormalization Group<sup>20</sup> and Iterative Monte Carlo (IMC)<sup>27</sup>) and were shown to properly predict the scaling of dsDNA persistence length with varying ionic strength. De Biase and co-workers developed an implementation of 3SPN.1 with ions using IMC for simulating single-

Received: April 12, 2015

Published: September 29, 2015

stranded DNA (ssDNA) in nanopores.<sup>33,34</sup> Recently, Korolev et al. used IMC and atomistic simulations to parametrize a coarse-grained DNA model, which was in good agreement with experimental dsDNA persistence lengths in explicit monovalent salt.<sup>29</sup>

These past DNA–ion models have been model-specific; each has been developed for a particular CG DNA model, thereby precluding transfer of the CG ions to other DNA models. This also makes it difficult to confirm claims of deviations from Odijk–Skolnick–Fixman (OSF) theory<sup>35,36</sup> that have been observed with a CG DNA model.<sup>37,38</sup> Therefore, it is desirable to develop a general, transferable CG model of ions that could be used across different platforms. Such a CG ions model could potentially be applied to existing models that carry explicit charges<sup>25</sup> and to DNA models that currently lack explicit electrostatics.<sup>22</sup>

The objective of this work is to systematically develop a CG model for biologically relevant ions in implicit solvent that is independent of the underlying DNA CG model. This is accomplished by performing relative entropy coarse-graining to match results from AA simulations of surrogate small molecules instead of DNA. The resulting parameters are validated by examining the predictions of local and global behavior, including local ionic distributions, ssDNA and dsDNA persistence length, and the potential of mean force between dsDNA molecules. This manuscript is organized as follows: First, we present the experimental design and simulation methods. Then, we validate our model by demonstrating good agreement with radial distribution functions from AA simulations. We compare both predicted local densities of ions and ssDNA and dsDNA persistence length to experiments and simulations by others. We demonstrate the utility of the model for simulation systems containing dsDNA by showing DNA condensation in the presence of multivalent ions. Lastly, we discuss additional CG DNA model refinements that may be needed when used with the proposed general ions model and discuss the ion model's limitations.

## 2. METHODS

**2.1. Ions.** We develop a model including sodium ( $\text{Na}^+$ ), magnesium ( $\text{Mg}^{2+}$ ), spermidine ( $\text{Spd}^{3+}$ ;  $\text{NH}_3^+-(\text{CH}_2)_4-\text{NH}_2^+(\text{CH}_2)_3-\text{NH}_3^+$ ), and chlorine ( $\text{Cl}^-$ ) ions. Other ions, such as potassium, manganese, putrescine, and spermine, are not treated here. However, our approach can be applied in a straightforward manner to obtain effective potentials involving these and other ions.

For the development of a CG model of ions that is independent of the representation of DNA, it is necessary to make several assumptions. The first assumption is that local ionic environment around dsDNA is determined by (1) interactions between the ions and the phosphate groups along the DNA backbone and (2) the excluded volume created by the bases and sugar sites. This assumption may seem unreasonable, especially given the nonuniform radial distribution functions presented by refs 33 and 34. However, distributions of ions around base and sugar sites in this model and that work are qualitatively consistent (see Figure S1 in the Supporting Information). The second assumption is that the phosphate–ion interactions can be obtained by performing simulations with a surrogate small molecule instead of simulating a complete DNA oligomer. Yoo and Aksimentiev used dimethylphosphate ( $\text{DMP}^-$ ) and magnesium hexahydrate ( $\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$ ) as surrogates for phosphates in the DNA

backbone and the magnesium ion, respectively. Optimizing AA interaction parameters using these surrogates molecules remedied discrepancies between AA simulation and experiment for larger macromolecular arrays. Encouraged by their successful use of surrogate small molecules, we also use  $\text{DMP}^-$  and  $\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$  as surrogates when obtaining our effective potentials.

**2.2. Mapping Function and Effective Potentials.** The construction of the CG force field requires the definition of a mapping function  $M$  transforms fine-grained coordinates  $\mathbf{r}$  to coarse-grained coordinates  $\mathbf{R}$ , e.g.,  $\mathbf{R} = M(\mathbf{r})$ .  $\text{Na}^+$  and  $\text{Cl}^-$  are represented by a 1:1 mapping as a single site with point charges of +1 and -1, respectively;  $\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$  is represented with a single site carrying a charge of +2 placed at the location of the water-coordinated  $\text{Mg}^{2+}$  ion.  $\text{DMP}^-$  was mapped to a single phosphate site placed at the center of mass of the phosphorus atom and its four covalently bonded oxygens.  $\text{Spd}^{3+}$  is represented with three sites linked by bonds, each carrying a charge of +1, that are placed at the locations of the amine nitrogen.

The nonbonded effective potentials between ions are modeled with a Coulombic interaction and a correction term,  $U_{\text{corr}}$ , as follows

$$U_{\text{ion-ion}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon(T)r_{ij}} + U_{\text{corr}}(r_{ij}) \quad (1)$$

where  $q_i$  and  $q_j$  are the charges of the  $i$ th and  $j$ th ions,  $\epsilon_0$  is the dielectric permittivity of vacuum,  $\epsilon(T)$  is the solution dielectric, and  $r_{ij}$  is the intersite separation. The solution dielectric constant is given by

$$\epsilon(T) = 249.4 - 0.788T/K + 7.20 \times 10^{-4}(T/K)^2 \quad (2)$$

which represents a polynomial fit to the data of Fernandez et al.<sup>39</sup> that was successfully used in previous DNA CG modeling efforts.<sup>18,25,40</sup>

The function  $U_{\text{corr}}$  is represented using cubic splines. Cubic splines are uniquely defined by the points, also called knots, that are interpolated in a continuously differentiable manner and the boundary conditions at the terminal knots. Cubic splines have two advantages: first, they are linear in their parameters (the knots), thus ensuring convergence to the optimal set of parameters as detailed in section 2.3. Second, when a sufficiently large number of knots is used, it is possible to reproduce most, if not all, of the features of true effective potential. This is not the case when the effective potentials are defined using analytical expressions that are defined a priori, as done in other work.<sup>32,41,42</sup> Cubic splines are not without disadvantages; in particular, it can be difficult to ascribe physical meaning to all features that arise in the effective potentials. We use cubic splines to model the pair correction,  $U_{\text{corr}}$ , as well as the bond and bend angle potentials.

Interactions between ionic and nonionic sites are represented using a repulsive Lennard-Jones (LJ) function

$$U_{\text{rep}}(r) = \begin{cases} 4\epsilon_{\text{rep}} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \epsilon_{\text{rep}} & r < r_c \\ 0 & r \geq r_c \end{cases} \quad (3)$$

where  $\epsilon_{\text{rep}} = 0.239 \text{ kCal/mol}$ ,  $\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}$ , and  $r_C = \sqrt[6]{2} \sigma_{ij}$ . As  $\sigma$  does not explicitly enter into the effective potentials, we approximate  $\sigma$  using the location of the soft-core repulsion that arises after systematic coarse-graining. In the present study, the LJ diameters of  $\text{Na}^+$  and  $\text{Cl}^-$  are set to 2.494 and 4.478 Å, respectively, consistent with the work of Joung and Cheatham.<sup>14</sup> The radius of the  $\text{Mg}^{2+}$  is 4.0 Å, more than twice the diameter from AMBER, due to the presence of the hydrating waters. Lastly, the diameters of the phosphate and  $\text{Spd}^{3+}$  beads are set to 4.5 and 3.3 Å, respectively.

**2.3. Relative Entropy.** Optimization of the CG parameters of the ion–ion effective potentials was performed using relative entropy coarse-graining (RE-CG).<sup>43</sup> Reference AA simulations are targeted in an iterative approach that seeks to find the effective potential that best preserves the information in the AA ensemble. By minimizing the relative entropy, optimal effective potentials can be obtained.

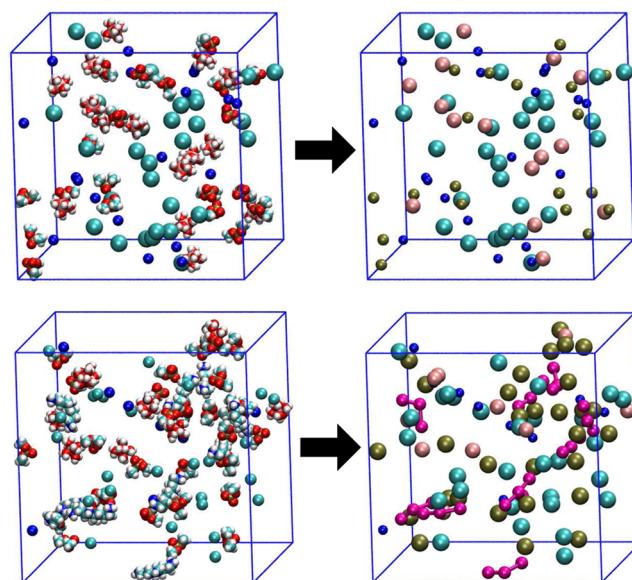
In this work, we follow the RE-CG algorithm as presented by Carmichael and Shell.<sup>44</sup> The knots of the initial nonbonded effective potentials,  $\lambda$ , are initially zeroed except for a linear ramp at close separations to model the excluded volume of each ion. The location of the linear ramp was unique to each pair interaction and was placed at the distance suggested by the AA radial distribution functions (RDFs). The initial bonded knots were obtained using Boltzmann inversion. For bonds between  $\text{Spd}^{3+}$  sites, linear ramps were added to eliminate artificially long bond lengths. Natural boundary conditions ( $\partial^2 U / \partial r^2 = 0$  at the lower and upper knots) were used. A Newton–Raphson iteration scheme, coupled with reweighting, was used to update the location of the knots until converged values were obtained. Simulations were reweighted until the effective fraction of frames contributing to the ensemble averages of derivatives was less than 0.5. In contrast to the dynamic adjustment of step size performed ref 44, the step size parameter  $\chi$  was held constant at 0.05. Twenty iterations of relative entropy optimization were performed to optimize the parameters between all ionic species except  $\text{Spd}^{3+}$ . Then, these effective potentials were held constant while 20 iterations of the effective potentials involving  $\text{Spd}^{3+}$  were optimized. We use “iterations” to refer to new CG simulations performed during optimization and not the reweighting step.<sup>44</sup> The source code used to perform our relative entropy coarse-graining is available upon request.

We note that effective potentials obtained from RE-CG or any other systematic coarse-graining method are dependent on the state point of the reference trajectory, as observed previously.<sup>45</sup> In principle, this would make the effective potentials presented here appropriate only at the conditions of the target ensemble ( $T = 300 \text{ K}$ ). In practice, we find that the effective potentials accurately reproduce AA RDFs at different ionic strengths (Figure S2). Including the temperature-dependent solution dielectric (eq 2) improves agreement when the effective potentials are used at temperatures other than the temperature of the target AA simulations. However, even with this correction, the agreement could be improved, and the effective potentials are not exactly transferable to other state points. Nonetheless, the differences are acceptable given the coarse-grained nature of our model.

**2.4. All-Atom Simulations.** The AA simulations targeted by RE-CG were performed using GROMACS 4.5.<sup>46</sup> We used the AMBER force field<sup>47</sup> with the Joung and Cheatham ion parameters and the nonbonded fix (NBFIX) corrections proposed by Yoo and Aksimentiev<sup>15</sup> for  $\text{DMP}^-$ ,  $\text{Na}^+$ ,  $\text{Cl}^-$ ,

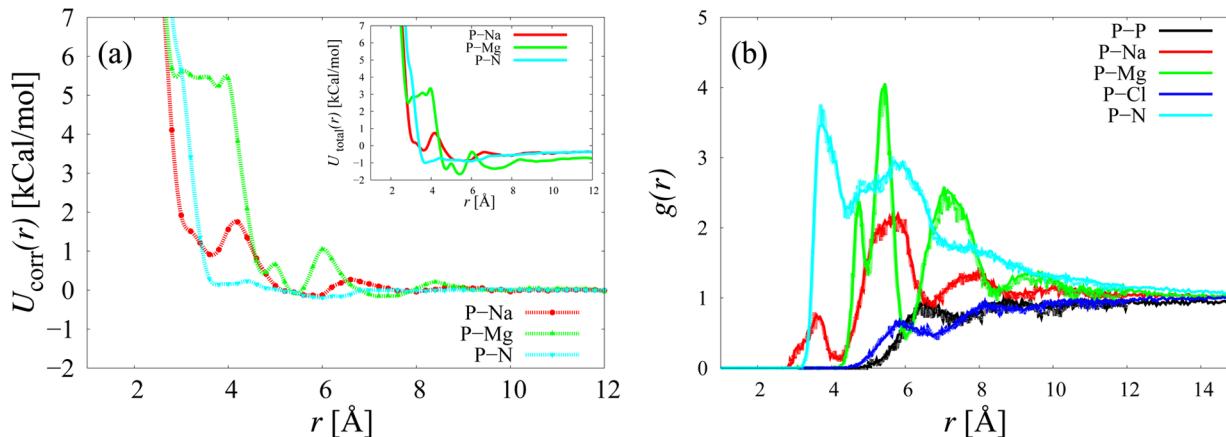
$\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$ . We used an additional NBFX correction between  $\text{DMP}^-$  and polyamine; although this correction was optimized for simulations with spermine<sup>4+</sup>,<sup>16,48</sup> we assume here that the correction can be applied to simulations of spermidine<sup>3+</sup>. The NBFX corrections universally decreased the affinity of cations to  $\text{DMP}^-$  (see Figure S6). Though the CHARMM force field<sup>49</sup> with the NBFX correction is not used for parametrization, preliminary simulations did not reveal significant differences between the radial distributions of the two models (see Figure S5).  $\text{Spd}^{3+}$  was modeled using the General Amber Force Field (GAFF)<sup>50</sup> with partial charges assigned using Antechamber<sup>51</sup> and Gaussian09<sup>52</sup> with the HF/6-31G\* basis set. AcPyPe<sup>53</sup> was used to convert the AMBER input files into a format suitable for running in GROMACS. Water was represented with the TIP3P model.<sup>54</sup>

Two different simulation boxes were simulated. The first simulation box contained 20  $\text{DMP}^-$ , 15  $\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$ , 20  $\text{Na}^+$ , and 30  $\text{Cl}^-$ . This trajectory was used to refine parameters between all CG ion sites except those of  $\text{Spd}^{3+}$ . The parameters between  $\text{Spd}^{3+}$  sites ( $\text{N}^+$ ) and the remainder of the ions were refined using simulations of the second simulation box, which contained 30  $\text{DMP}^-$ , 10  $\text{Spd}^{3+}$ , 10  $\text{Mg}\cdot(\text{H}_2\text{O})_6^{2+}$ , 10  $\text{Na}^+$ , and 50  $\text{Cl}^-$ . Each simulation box also contained approximately 7000 water molecules. Figure 1 displays the AA configuration and the



**Figure 1.** Cubic boxes used in the all-atom (left) and coarse-grained simulations (right). The top set of boxes were used when parameterizing interactions not involving spermidine<sup>3+</sup>. The bottom set of boxes were used in simulations for parametrizing ion–spermidine<sup>3+</sup> interactions. Explicit water has been omitted for clarity from the cubic boxes representing the all-atom systems simulated. In coarse-grained simulations, the color code is as follows: tan—phosphate; blue—sodium; cyan—chlorine; pink—magnesium; magenta—spermidine.

corresponding CG mapping for both of these systems. A 10 Å cutoff was used for the van der Waals and Coulombic interactions, and electrostatics were treated using the Particle Mesh Ewald technique.<sup>55</sup> Bond lengths were constrained using the LINCS algorithm,<sup>56</sup> and a time step of 2 fs was employed. The modified Berendsen thermostat<sup>57</sup> and Parrinello–Rahman barostat<sup>58</sup> were used to maintain a temperature and pressure of 300 K and 1 bar, respectively.



**Figure 2.** (a) Effective potentials obtained from relative-entropy coarse-graining. For clarity, only the effective potentials between phosphate sites and counterions are shown. (Inset) The total pair potentials that act between phosphate and counterions. (b) Radial distribution functions between phosphate sites and all other ionic site types. Solid lines represent the all-atom simulations; dotted lines represent the coarse-grained systems.  $T = 300$  K.

Simulations were performed as follows: after energy minimization the system was equilibrated for 100 ps at 300 K in the NVT ensemble. An additional 100 ps of equilibration occurred at 300 K and 1 bar in the NPT ensemble. After the box size had equilibrated (corresponding to a box length of 60 Å), 50 ns of simulations were performed in the NVT ensemble with snapshots saved every 2 ps. The simulations were observed to converge after 30 ns (see Figure S4). After the simulations were completed, histograms of relevant bond distances, bend angles, and pair distances between all four ion types were constructed for use in RE-CG. The construction of these histograms was facilitated by the MDAnalysis Python toolkit.<sup>59</sup>

**2.5. Coarse-Grained Simulations.** All CG simulations were performed using the LAMMPS molecular dynamics package.<sup>60</sup> The effective potentials obtained from RE-CG were represented using a tabulated pair style that linearly interpolated between 501 points in bonded potentials and 551 points in nonbonded effective potentials. The Particle-Particle Particle-Mesh solver was used to calculate the electrostatic interactions with a real space cutoff of 12 Å. This cutoff was consistent with the cutoff used in the effective potentials. As all ion simulations were performed with an implicit solvent, the Grønbæch-Jensen Farago (G-JF) Langevin thermostat<sup>61,62</sup> was used. We note that this thermostat produces a distribution of kinetic energies that is systematically lower than the Boltzmann distribution; however, the thermostat produces the correct Boltzmann distribution of positions that determine our effective potentials. A time step of 10 fs was used in all simulations. All tabulated potentials used in this work can be found in the USER-3SPN2 LAMMPS package, which is freely available online.<sup>63</sup>

When simulations of ions were performed with DNA, the 3SPN.2 CG DNA model<sup>25</sup> was used. That model has been described in detail in the literature, and here we merely point out that it uses a 3-site-per-nucleotide (3SPN) description of DNA with spherical sites representing the sugar, phosphate, and base. Angle-dependent interactions penalize deviations from the structure of B-DNA; interactions between adjacent base sites in a single strand capture base stacking interactions, and interactions between base sites on opposite strands capture base-pairing and base-stacking interactions. Electrostatics are modeled using Debye–Hückel (DH) theory, with each phosphate assigned a  $-0.6$  charge, assuming Manning

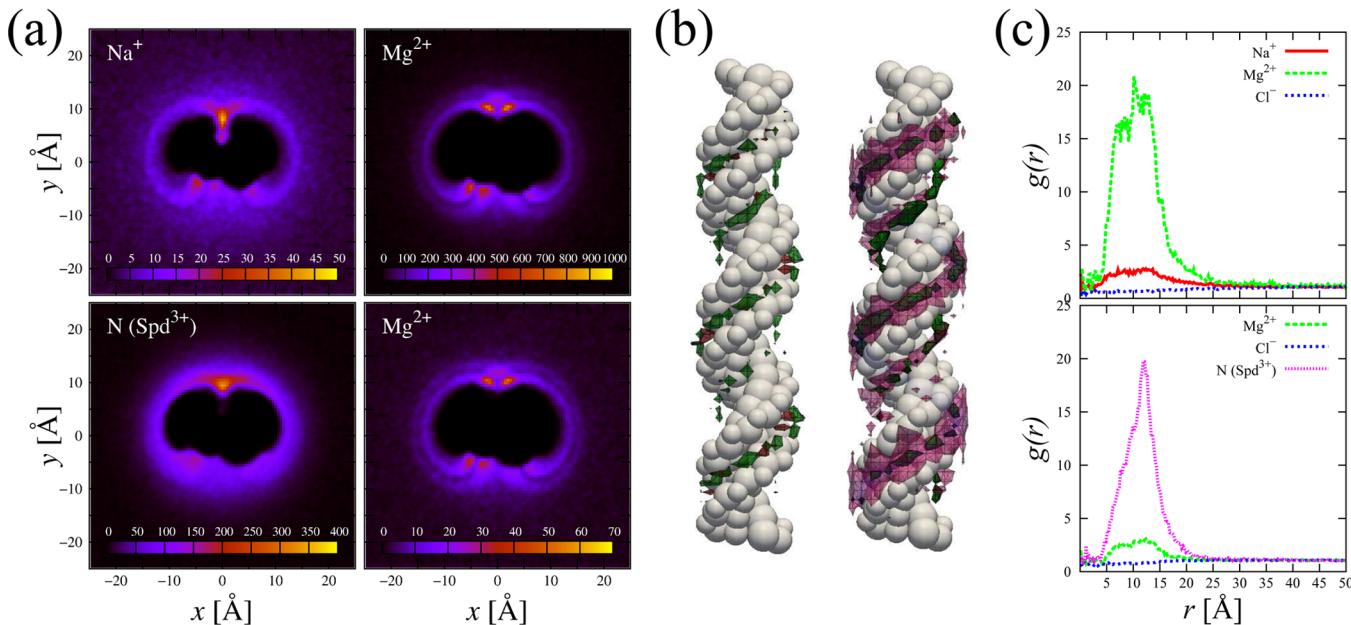
condensation<sup>8</sup> along a single-strand of DNA. Manning condensation predicts that each phosphate along dsDNA will carry a charge of  $-0.24$ ; a comparison of these two effective charges will be performed in section 3.3. The 3SPN.2 model describes the widths of the major and minor grooves, the persistence lengths of ssDNA and dsDNA, and the thermodynamics of DNA hybridization.<sup>25</sup> It has been used to study the influence of sequence, length, and ionic conditions on oligomer hybridization<sup>64</sup> and the role of intrinsic curvature in nucleosome positioning.<sup>40,65</sup> Our use of 3SPN.2 is motivated by our familiarity with the model; however, the ions model presented here should be appropriate for simulation with dsDNA models that properly represent the excluded volume of dsDNA and carry charges at the location of the phosphates. Examples of such models include those presented in refs 23, 29, and 66.

### 3. RESULTS

**3.1. Effective Potentials.** The optimized effective potentials are displayed in Figure 2a. (The polyamine of  $\text{Spd}^{3+}$  is indicated by  $\text{N}^+$ ). For clarity, only  $U_{\text{corr}}(r)$  for effective potentials between phosphate and cations (see eq 1) is displayed; the complete set of effective potentials can be found in the Figure S7.  $U_{\text{corr}}$  corrects the Coulombic potential to account for the effects of hydration and steric overlap. The use of cubic splines permits the reproduction of subtle effects that could not be easily resolved with analytical expressions.

The effective potentials reproduce essentially all features of the radial distribution functions from AA simulations. This is illustrated in Figure 2b, where excellent correspondence is seen between radial distribution functions from AA and CG simulations (the two curves lie almost exactly on top of each other). The covalent bonds between the three  $\text{N}^+$  sites that constitute  $\text{Spd}^{3+}$  cooperatively interact with the phosphate site, giving rise to peaks in the radial distribution functions for  $\text{P}^- - \text{N}^+$  and  $\text{P}^- - \text{Mg}^{2+}$  that have similar height. Despite the relative weakness of the  $\text{P}^- - \text{N}^+$  effective potential when compared to  $\text{P}^- - \text{Mg}^{2+}$ ,  $\text{Spd}^{3+}$  out-competes  $\text{Mg}^{2+}$  to bind with dsDNA, as discussed in the following section.

**3.2. Local Ionic Environment.** Achieving consistency with targets from AA simulations is necessary but does not constitute a sufficient validation of the CG model presented



**Figure 3.** (a) Helically averaged number densities of ions around dsDNA after normalization by simulation box densities. The top heat maps correspond to 40 mM NaCl and 5 mM MgCl<sub>2</sub>. The bottom heat maps correspond to 40 mM MgCl<sub>2</sub> and 5 mM Spd<sup>3+</sup>. The minor groove is facing up. (b) Isosurfaces of cation number density. (Left) 40 mM NaCl and 5 mM MgCl<sub>2</sub>; red, Na<sup>+</sup> = 0.0005 Å<sup>-3</sup>; green, Mg<sup>2+</sup> = 0.001 Å<sup>-3</sup>. (Right) 40 mM MgCl<sub>2</sub> and 5 mM SpdCl<sub>3</sub>; green, Mg<sup>2+</sup> = 0.0005 Å<sup>-3</sup>; magenta, Spd<sup>3+</sup> = 0.001 Å<sup>-3</sup>. (c) The cylindrically averaged radial distribution functions of ions around the dsDNA axis. (Top) 40 mM NaCl and 5 mM MgCl<sub>2</sub>; (Bottom) 40 mM MgCl<sub>2</sub> and 5 mM SpdCl<sub>3</sub>. DNA is 30 bp poly(A)–poly(T). T = 300 K.

here. Also note that improvements to AA force fields are continuously proposed to remedy observed deficiencies as they arise.<sup>15,67</sup> To further validate the ions model, we examined the local distributions of ions in the vicinity of dsDNA. CG simulations were performed on a 30 bp poly(A)–poly(T) dsDNA in a 200 × 200 × 200 Å cubic box. The dsDNA alters the concentration of ions in its immediate vicinity; therefore, we calculated a bulk concentration that facilitates comparison with experimental results and the simulation work of others. This bulk concentration is defined using the simulation volume outside of a cylinder with 50 Å radius centered on the helical axis of the dsDNA, similar to that of ref 68. Two separate ionic conditions were simulated: 31 mM NaCl and 9 mM MgCl<sub>2</sub> and 30.5 mM MgCl<sub>2</sub> and 7.5 mM SpdCl<sub>3</sub>; additional Na<sup>+</sup> or Cl<sup>-</sup> ions were added to neutralize the system. These ionic conditions had bulk concentrations of 5 mM Mg<sup>2+</sup> and 40 mM Na<sup>+</sup> and 5 mM Spd<sup>3+</sup> and 40 mM Mg<sup>2+</sup>, respectively. The dsDNA was held fixed in space, and the local distributions of ions around the central 20 nucleotides were recorded and subsequently averaged along the helical axis.

The local two-dimensional density of each ion relative to that of the simulation box is displayed in Figure 3a. Both Na<sup>+</sup> and Mg<sup>2+</sup> preferentially bind to the minor groove; Na<sup>+</sup> binds deeply into the groove whereas Mg<sup>2+</sup> binds on the surface. In contrast to the monatomic counterions, the binding of Spd<sup>3+</sup> sites is rather diffuse, making it difficult to draw conclusions from the density maps alone.

To more accurately discover the location of counterions, especially Spd<sup>3+</sup> sites, we constructed three-dimensional isosurfaces for each of the cations in both systems. Interestingly, we observed that the Spd<sup>3+</sup> defers from binding with the major groove and instead appeared to bind to the phosphates along the backbone. Spd<sup>3+</sup> exhibits a wide array of behavior: bridging the major groove, partially inserting within the minor groove, or

transient parallel binding to the backbone. Rarely do all three nitrogens of a single Spd<sup>3+</sup> interact with the same phosphate. Analysis of corresponding trajectories shows that binding events are transient, and that Spd<sup>3+</sup> is highly mobile, diffusing along the backbone contours of the dsDNA before being released back into solution.

Next, we compared our simulation results to those of others. In early Monte Carlo simulations of dsDNA with Mg<sup>2+</sup> and Spd<sup>3+</sup>, competitive binding between the two cations was observed.<sup>69</sup> At the conditions simulated, Mg<sup>2+</sup> out-competed Spd<sup>3+</sup> to bind to the phosphates. This is inconsistent with the RDFs presented in Figure 3c. These differences can be explained by examining the relative concentrations in the Monte Carlo simulations and the present work. Their simulations included fewer Spd<sup>3+</sup> ions than phosphate ions, thus making it entropically favorable for Mg<sup>2+</sup> to bind. In our simulations, the relatively high concentration of Spd<sup>3+</sup> minimizes the entropic cost of Spd<sup>3+</sup> binding.

Korolev and co-workers have performed AA MD simulations of DNA with polyamines, including Spd<sup>3+</sup>.<sup>11,70,71</sup> They observed that Spd<sup>3+</sup> interacts weakly with the major groove and interacts with DNA in a delocalized and dynamic manner.<sup>70,71</sup> This observation is consistent with our simulations, even though only we use the NBFIIX correction. Korolev et al. also showed that Na<sup>+</sup> has specific binding sites along DNA grooves,<sup>71</sup> a feature that cannot be resolved in our model, which does not include interactions between ions and base sites.

More recently, Yoo and Aksimentiev examined binding of Na<sup>+</sup> and Mg<sup>2+</sup> to dsDNA in AA simulations incorporating the NBFIIX correction.<sup>68</sup> They observed high local concentrations of both Na<sup>+</sup> and Mg<sup>2+</sup> in the vicinity of the major groove. We observe qualitatively similar distributions of ions around dsDNA; however, the magnitude of the concentrations is

smaller by more than an order of magnitude. These differences can be attributed to the absence of long-lived binding events.

**3.3. Persistence Length.** Although the local details of DNA–ion interactions may be lost due to the coarse-graining inherent in our modeling, global properties such as the dependence of persistence length on ionic strength should be well-resolved. To confirm this behavior, we performed persistence length calculations with 3SPN.2 in monovalent salt using 68 bp poly(A) ssDNA and the 68 bp dsDNA sequence used in our prior work.<sup>25</sup> Simulations were performed in a  $400 \times 400 \times 400 \text{ \AA}$  simulation box with NaCl concentrations ranging from 1.0 mM to 1.0 M. Eight simulations, each 75 million steps in length, were performed at each concentration with the first five million steps excluded from analysis. Analogous simulations were performed using the Debye–Hückel (DH) approximation with different effective charges assigned to each phosphate bead, as discussed below. The persistence lengths of ssDNA and dsDNA were calculated using different algorithms: ssDNA can be considered a flexible chain, for which the persistence length can be calculated as

$$l_p = \frac{\langle R_e^2 \rangle}{2L} \quad (4)$$

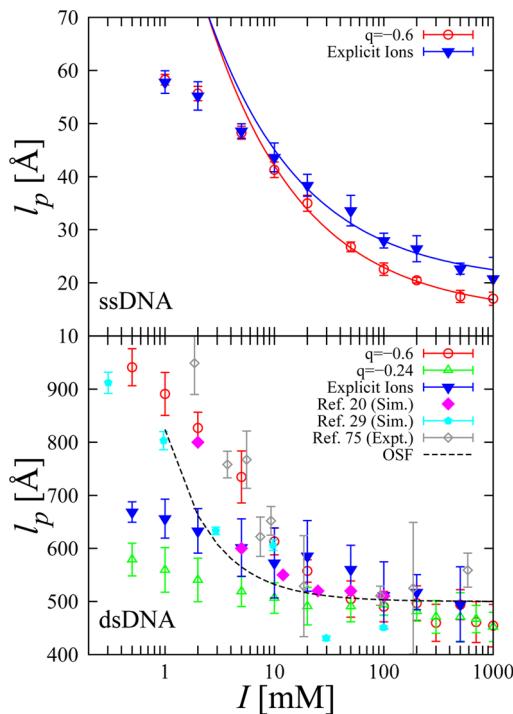
where  $R_e$  is the end-to-end distance,  $L$  is the contour length, and the angle brackets represent a long-time average. In contrast, dsDNA is a semiflexible polymer and is more accurately described as a wormlike chain. The persistence length can be calculated from the decay of the autocorrelation function

$$\langle \mathbf{u}(s) \cdot \mathbf{u}(s + \Delta s) \rangle = e^{-\Delta s/l_p} \quad (5)$$

where  $s$  is the position along the polymer contour length, and  $\mathbf{u}(s)$  is a vector tangent to the helical axis at  $s$ . The helical axis is approximated using tetrads of nucleotides spaced three nucleotides apart using the method presented by Kahn.<sup>72</sup> To eliminate any effect caused by the helical nature of dsDNA and fraying of the duplex ends, we used vectors separated by ten bases and omitted the 10 bases nearest the duplex ends from the analysis.

First, we examined the scaling of the ssDNA persistence length  $l_p$ . Tinland and co-workers observed<sup>73</sup> that the persistence length of ssDNA scales with ionic strength  $I$  according to the power law  $I^{-0.5}$ . We performed two sets of simulations; the first with explicit ions and the second with DH and an effective phosphate charge,  $q$ , of  $-0.6$ . The calculated persistence lengths are shown in Figure 4. The persistence lengths from explicit ion simulations display  $I^{-0.5}$  scaling above  $I = 10 \text{ mM}$ ; however, deviations occur at lower ionic strengths. These deviations have two origins: First, the ssDNA simulated here is relatively short, making finite length effects pronounced as the ionic strength decreases and the Debye length becomes very large. Second, simulations using explicit ions require counterions to preserve charge neutrality in the simulation box. As a result, the total concentration of cations cannot drop below 1.7 mM, which keeps us from effectively sampling very small ion concentrations in explicit salt. Note that simulations with and without explicit ions display the same deviations at low ionic strength, suggesting that the deviation is dominated by finite length effects.

The simulations using DH in lieu of explicit ions also scale as  $I^{-0.5}$ , validating the invocation of Manning condensation<sup>8,74</sup> to reduce  $q$  from  $-1.0$  to  $-0.6$  in prior modeling work.<sup>25</sup> The



**Figure 4.** (Top) Scaling of ssDNA persistence length with ionic strength for a 68 bp poly(A) with implicit (red) and explicit (blue) treatments of the ionic environment. The solid lines represent a best fit line to the data at 10 mM using the functional form  $A I^{-0.5} + B$ . (Bottom) Persistence length of a 68 bp dsDNA at various ionic strengths of NaCl. Results are presented for implicit ion simulations with varying effective charges on the phosphates ( $q = -0.24, -0.6$ ). Error bars represent one standard deviation, as calculated from eight different simulations. Also shown are the persistence lengths calculated with the explicit ion model presented here. These results are compared to other simulations with explicit ions,<sup>20,29</sup> experiment,<sup>75</sup> and OSF theory.<sup>35,36</sup>  $T = 300 \text{ K}$ .

divergence of explicit and implicit ion persistence lengths at increasing ionic strength highlights a limitation of DH; it is not strictly valid at high ionic strengths.

The scaling of the dsDNA persistence length with ionic strength is shown at the bottom of Figure 4. As observed with ssDNA, we see deviations from theory at low ionic strength. The dependence of  $l_p$  on ionic strength at very low concentrations is less pronounced with explicit ions than analogous 3SPN.2 simulations performed with implicit salt. This divergence is unsurprising given that the effective charge per phosphate in the implicit salt simulations ( $q = -0.6$ ) was assigned based on the linear charge density of ssDNA. Although such assignment produces good agreement with explicit ion simulations for ssDNA, Manning condensation predicts  $q = -0.24$  in dsDNA. Additional simulations using this value of  $q$  reveal a dependence on ionic strength that is weaker than in experiment.<sup>75</sup> The persistence length from simulations with explicit ions display scaling between that predicted by the implicit ion treatments. The scaling is roughly consistent with predictions from the other CG DNA models that include explicit ions;<sup>20,29</sup> the discrepancies that do arise may be attributed to the shorter length of dsDNA used in this work. OSF theory predicts that the persistence length of a semiflexible polyelectrolyte can be separated into nonelectrostatic and electrostatic contributions with the latter becoming negligible at high ionic strength. The persistence lengths calculated here do

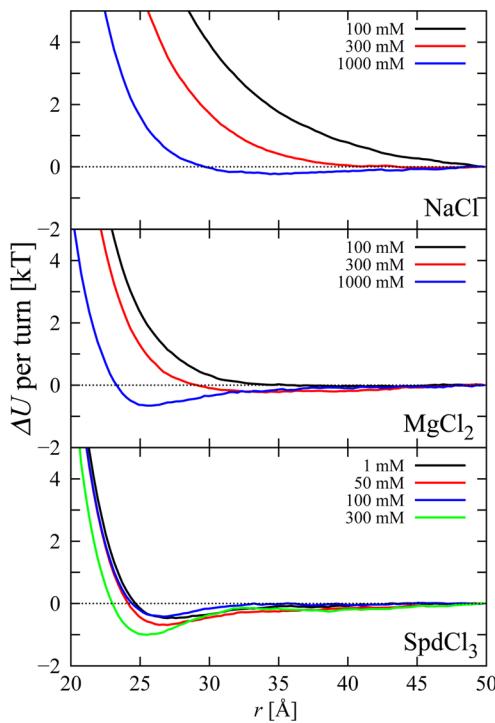
not appear to approach a limiting value at high ionic strength, thus challenging the assumption of a constant, nonelectrostatic contribution. Saveliev and co-workers recently performed CG DNA simulations with explicit ions<sup>37,38</sup> similar to those presented here. They observed that electrostatics contributes more strongly to persistence length than predicted by OSF theory, observing that dsDNA persistence length decreases by 25% from 0.1 to 1 M ionic strength. The limited explicit ion simulations presented here do not directly corroborate the results of Saveliev et al.; however, the ions model we present should prove useful in further studies of this behavior.

**3.4. dsDNA Potential of Mean Force.** It has been observed experimentally that attractive interactions between dsDNA can be induced in the presence of multivalent counterions, resulting in DNA condensates.<sup>76</sup> Luan and Aksimentiev recently performed AA simulations of dsDNA attraction using the AMBER force field.<sup>77</sup> Using periodic boundary conditions, they simulate effectively infinite dsDNA and calculated the potential of mean force (PMF) between the two dsDNA molecules. They observed that high concentrations of MgCl<sub>2</sub> resulted in Mg<sup>2+</sup> binding to minor grooves, causing a bridging effect that generated attractive forces between dsDNAs. However, this observation is inconsistent with experimental observations that demonstrate that only trivalent and higher ions, such as Spd<sup>3+</sup> or Cobalt hexamine<sup>3+</sup>, condense dsDNA.

To further validate the present CG ion model, we performed PMF calculations analogous to those of ref 77. Two separate poly(A)-poly(T) dsDNA molecules, consisting of two turns of dsDNA with bonds wrapped around the periodic box, were simulated in a box with *x* and *y* dimensions of 200 Å. Harmonic constraints with a force constant of 1.0 kCal mol<sup>-1</sup> Å<sup>-2</sup> were placed between the centers of mass of the two dsDNA. Simulations were performed for 20 M steps with distances recorded every five ps. The resulting data were postprocessed using the Grossfield implementation of the weighted histogram analysis method (WHAM).<sup>78</sup>

The PMF between dsDNA in NaCl, MgCl<sub>2</sub>, and SpdCl<sub>3</sub> are displayed in Figure 5. The reported concentrations of NaCl and MgCl<sub>2</sub> were selected to correspond with those used in ref 77. In contrast, the reported concentrations of SpdCl<sub>3</sub> are bulk concentrations, calculated as described in section 3.2. In NaCl, increasing the ionic strength increases the screening of the repulsive interaction between the dsDNA. At no concentration do monovalent counterions induce attraction between dsDNAs, as was expected based on experimental observations.<sup>76,79</sup> In MgCl<sub>2</sub>, Mg<sup>2+</sup> primarily screens charge and produces a very small attraction at 1 M. This attraction is on the order of one *kT* per turn, consistent with the PMF measured from all-atom simulations using the Yoo and Aksimentiev NBFIX correction.<sup>80</sup> The minimum of the coarse-grained PMF is roughly consistent with that observed in the AA simulations, although subtle features such as secondary minima are lost upon coarse-graining. This relatively weak attraction may not be strong enough to induce DNA condensation, and a definite conclusion cannot be drawn from the PMFs alone. These predictions differ from those in ref 77; however, effective potentials based on the uncorrected AMBER force field (e.g., without NBFIX) did produce potentials of mean force that are consistent with that work (see Figure S8).

In our simulations containing Spd<sup>3+</sup>, dsDNA condensation is observed to occur at a bulk concentration of only 1 mM. The onset of attraction at such low concentrations is consistent with



**Figure 5.** Potentials of mean force (PMFs) between effectively infinite dsDNA molecules in NaCl (top), MgCl<sub>2</sub> (middle), and SpdCl<sub>3</sub> (bottom). Concentrations of NaCl and MgCl<sub>2</sub> are for the entire simulation box, and SpdCl<sub>3</sub> describe the bulk (>50 Å from the helical axes). PMFs using the AMBER force field without the NBFIX correction can be found in Figure S8.  $T = 300$  K.

experimental observations;<sup>81</sup> however, we do not observe the destabilization of condensation at high concentrations that was seen in the experiment. Instead, the strength of attraction between the dsDNA becomes stronger at large, nonbiological concentrations. Osmotic stress and magnetic tweezers have been used to report the change in free energy upon condensation.<sup>82</sup> The reported change in Gibbs free energy was  $-2\text{ kT/turn}$ , smaller than the potential energy of attraction described here ( $\sim kT/\text{turn}$ ); however, a comparison between potential energy and free energy is tenuous. These experiments report interhelical spacing between 25–30 Å; consistent with our simulated PMFs.

The depth of the potential is not consistent with the attractive forces measured in the AA simulations of Dai and co-workers.<sup>83</sup> The attractive well in their work ( $\sim -6\text{ kT/turn}$ ) is smaller than that observed here ( $\sim 1\text{ kT/turn}$ ). This difference is mostly attributable to the use of the NBFIX correction for polyamine in our work, although geometry and ion concentrations may also contribute to the disagreement. A better comparison is with the recently published inter-dsDNA PMF in the presence of spermine<sup>4+</sup> (Figure 12b of ref 48). AA simulations incorporating the polyamine NBFIX correction yield an interhelical spacing of 27–28 nm, which is in excellent agreement with Figure 5. At 1 mM, the AA attractive well is slightly deeper than that of the CG simulations ( $1\text{ kT}$  vs  $0.6\text{ kT}$ ); the stronger attraction in the AA simulations is likely due to their use of a polyamine with a higher valence (spermine<sup>4+</sup>). This consistency observed between our work and AA simulations incorporating the polyamine NBFIX correction suggests that our CG representation of Spd<sup>3+</sup> can be used with

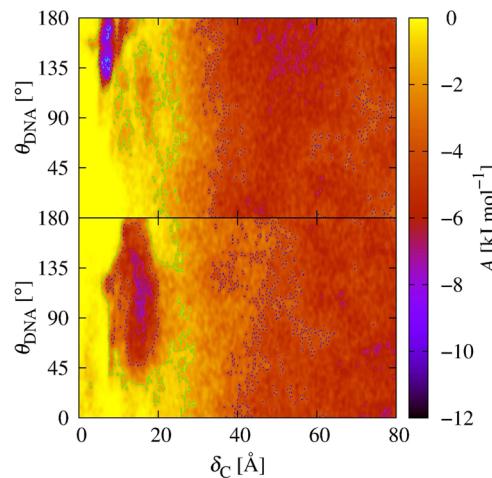
reasonable confidence to study systems where DNA condensation occurs.

**3.5. DNA Model Refinement.** When any model is altered, it is important to reconsider its predictions by revisiting the metrics used during the original model development. The CG ions model presented here provides a detailed description of the ionic environment surrounding DNA; therefore, those DNA phenomena that are known to be strongly dependent on ionic strength must be re-examined. We have already demonstrated in section 3.3 that adding CG ions to an otherwise unmodified CG DNA model changes the predicted persistence lengths. In this section, we examine the impact of the ionic representation on the DNA melting temperature, a phenomenon that is strongly dependent on the local ionic environment.

The melting temperature of dsDNA is a thermodynamics quantity, representing the balance between attractive Watson–Crick base pairing, entropy, and electrostatic repulsion between ssDNA molecules. Ions shield the phosphates of these complementary molecules, leading to a logarithmic dependence on ionic strength. The 3SPN.2 model uses DH theory and an effective charge,  $q$ , of  $-0.6$  on each phosphate site. The reduced charge on the phosphate was justified using Manning condensation,<sup>8</sup> which predicts that cations will condense along the DNA backbone until the effective repulsion along the backbone is equal to  $kT$ . When using this implicit treatment of counterions, 3SPN.2 was able to reproduce experimental melting temperatures across a wide range of ionic strengths. The question that arises now is whether including explicit ions results in melting temperatures consistent with previous results and experiments.

To address this question, we performed multiple-walker metadynamics simulations.<sup>84,85</sup> These simulations differ from those in our prior work<sup>25,40,64</sup> in that these simulations were performed in LAMMPS using the PLUMED 1.3 add-in.<sup>86</sup> All other simulation parameters, including number of walkers, hill height, and order parameters were as performed in our previous studies of DNA.<sup>25</sup> For completeness, we remind the reader that simulations are performed using two order parameters. The first is  $\delta_C$ , the distance between the centers of the two complementary ssDNA molecules, and the second is  $\theta_{\text{DNA}}$ , the angle between vectors connecting the 5' and 3' ends of each ssDNA. These order parameters allow resolution of hybridized (small  $\delta_C$  and  $\theta_{\text{DNA}} \approx 180^\circ$ ) and dehybridized states (large  $\delta_C$ ). As ionic effects are independent of sequence length and identity, we confine our simulations to a single 15 bp DNA oligomer (5'-TACTAACATTAACCA-3'). Our simulation protocol is as follows: first, simulations of 150 million time steps were performed using 3SPN.2 without ions at 10 K intervals centered on the melting temperature. After converged free energy surfaces were obtained, the added metadynamics bias was used as a starting point for two separate sets of simulations: one with the CG ions model described in this work, and the other with the DH using  $q = -1.0$  on the phosphate sites. To ensure stability when using explicit ions, the time step was halved to 10 fs and hills were deposited every 20 ps. The metadynamics simulations were run for an additional 30 million time steps and the bias was recorded.

If the electrostatic treatments are equivalent to those in the original 3SPN.2 model, then the added bias should be approximately uniform across the order parameter space. Representative examples of the added bias are shown in Figure 6. The added bias is not uniform for either set of simulations. In



**Figure 6.** Free energy surfaces representing the additional bias added to simulations with explicit ions (top) and with Debye–Hückel (DH) and  $q = -1.0$  (bottom) after obtaining convergence in simulations with DH and  $q = -0.6$ .  $T = 300$  K;  $I = 69$  mM.

simulations with explicit ions, a well appears at small  $\delta_C$  and  $\theta_{\text{DNA}} \approx 180^\circ$ . This indicates that, at small strand separations, the repulsion present with explicit ions is smaller than that experienced with DH and  $q = -0.6$ . This is not surprising; Manning condensation predicts phosphates in dsDNA should have effective charges of  $-0.24$  due to the high linear charge density. Additionally, a bias is also added at large  $\delta_C$ . Although the amount of bias at any one location is not as large as that added at small  $\delta_C$ , the cumulative effect of this bias is a reduction in melting temperature of approximately 4 K when compared to that of 3SPN.2, as shown in Table 1. In contrast,

**Table 1. Melting Temperatures of the 3SPN.2 CG DNA Model Using Three Different Representations of the Electrostatic Interactions As Calculated Using Metadynamics Simulation<sup>a</sup>**

ionic strength (mM)	expt (K)	DH ( $q = -0.6$ , K)	explicit (K)	DH ( $q = -1.0$ , K)
69	308.4	315.0	310.9	297.9
119	313.6	317.6	314.0	306.8
220	317.3	321.7	317.5	307.7
400	320.5	323.5	320.4	306.6

<sup>a</sup>The standard deviations of these melting temperature are between 1 and 4 K. DNA sequence is 5'-TACTAACATTAACCA-3'. DH = Debye–Hückel.

the simulations of 3SPN.2 with  $q = -1.0$  result in the majority of the bias being added at large values of  $\delta_C$ ; this bias is required to offset the larger barriers to hybridization that arise when the electrostatic repulsion between dsDNA is almost 3× stronger. This effectively reduces the melting temperatures by more than >15 K when compared to the original 3SPN.2 values.

These results confirm that a CG DNA model would need minor refinement after inclusion of explicit ions to recover the correct thermodynamics. In particular, interstrand attractions in DNA models with explicit charges of  $q = -1.0$  at the phosphates would need to be reduced. The close correspondence between the explicit ion simulations and  $q = -0.6$  validates the previous decision to invoke Manning condensation to reduce the effective charge of phosphates.<sup>25</sup>

#### 4. LIMITATIONS

Before concluding, it is important to revisit some of the limitations of the ions model presented here. They fall into three categories: model construction, simulation protocols, and model application.

To construct effective potentials from AA simulations, it was necessary to make a few assumptions. The first assumption was that an AA force field, specifically the AMBER force field with the NBFIX correction, correctly models the local electrostatic environment around dsDNA. This prediction appears reasonable, given the previous success of AA simulations describing the behavior of dsDNA and the local distributions of ions.<sup>68</sup> However, future improvements to the AA force field would require adjustment to the effective potentials to most accurately model DNA–ion interactions. The second assumption was that dimethylphosphate acts as a surrogate for the phosphate in the DNA backbone. The success of previous work by Yoo and Aksimentiev<sup>15</sup> suggests that this assumption is reasonable for Na<sup>+</sup>, Mg<sup>2+</sup>, and Cl<sup>-</sup>. However, it may not be appropriate for a small molecule, such as Spd<sup>3+</sup>, where it is reasonable to expect that the presence of polymeric DNA would alter the binding and flexibility of Spd<sup>3+</sup>. Nevertheless, we believe the observed consistency between the CG and AA PMFs validate this simplifying assumption. Perhaps more limiting is the assumption that backbone and base sites only act to exclude ions from the vicinity of the phosphates. The differences between RDFs between base sites and ions (see Figure S1) suggest that the predicted local distribution of ions should be viewed as only qualitatively consistent with the experiment results. The lack of attractive interactions between ions and base sites, as well as the accelerated dynamics inherent in coarse-grained, implicit solvent simulations, make this model inadequate for studies of specific ion binding. If information regarding the specific binding of ions to DNA is needed, AA simulations in explicit solvent provide a better alternative.

The CG simulations performed here involved configurations where ions are expected to bind transiently to DNA. In such binding events, the chemical potential in the bulk is altered. It is necessary to simulate the  $\mu$ VT ensemble using techniques such as GCMC to preserve the correct chemical potential(s). In all of our CG simulations, we simulate in the NVT ensemble neglecting the changes in chemical potential caused by condensation. Inserting and removing ions as done in GCMC disrupts charge neutrality and invalidates the Ewald-based methods used in LAMMPS. We note that alternative methods for  $\mu$ VT simulations with charged species have been developed.<sup>87,88</sup> Although simulating in the NVT ensemble may limit our validation simulations, it does not affect the effective potentials. Both AA and CG simulations were performed in the NVT ensemble, and the effective potentials are transferable to any ensemble. A more important concern would be the lack of explicit water. The effective potentials partially account for the local structure of water around ions by matching radial densities; however, the kinetic effects due to ordered water molecules are absent. Therefore, behaviors such as the response of condensed dsDNA to external stresses are expected to only be qualitatively consistent with experimental results.

The last limitations are related to the application of this model to existing coarse-grained DNA models. The effective potentials presented here were developed entirely independent of the DNA model. Therefore, the ions model can be applied to CG DNA models if the following caveats are met: (1) the CG

DNA model properly models the excluded volume of DNA (e.g., major and minor grooves in the case of dsDNA), and (2) the DNA model carries electrostatic charges at the location of the phosphate in the DNA backbone. Examples of CG DNA models that would be appropriate are found in refs 23, 29, and 66. As the ions model does not seek to resolve ion-base interactions, the specific properties of the bases are irrelevant. However, it is necessary to define the LJ or similar parameters to capture the repulsion between ions and nonphosphate sites. In this work, we apply arithmetic mixing rules to determine  $\sigma_{ij}$  between ions and uncharged sugar and base sites. The effective ion diameters described in section 2.2 should be suitable initial values.

Including this ions model in an existing CG DNA model will likely require refinement of the underlying DNA CG force field. This is particularly true in applications where it is important to capture quantitatively the thermodynamic transition between ssDNA and dsDNA, as highlighted in section 3.5. It is possible to achieve agreement with experiment when using the ions model with an unmodified DNA model, as shown in sections 3.3 and 3.4. Despite these successes, users should simulate benchmarks with and without ions to establish consistency and refine the model parameters when inconsistencies are found.

#### 5. CONCLUSIONS

A new, general coarse-grained model for ions has been presented that can be applied to existing coarse-grained DNA models that currently lack explicit ions. Importantly, the ions model includes CG parameters for modeling spermidine<sup>3+</sup>, thereby enabling simulations of dsDNA condensation. The model is parametrized using relative entropy coarse-graining that match distributions from all-atom simulations performed with the latest NBFIX corrections.<sup>15,16</sup> The performance of the ions model is demonstrated through examinations of the local ionic distribution, dependence of persistence length on ionic strength, and potential of mean force between dsDNA. Although the model is not predictive of ion binding, it does predict a dependence of persistence length on monovalent ionic strength, which is consistent with the simulations performed using other dsDNA models with ions. Furthermore, the potential of mean force between dsDNA is consistent with experimental observation and all-atom simulations.

The results shown here were generated with the 3SPN.2 CG DNA model; however, the coarse-grained ions model can, in principle, be used with any coarse-grained DNA model that carries electrostatic charges at the location of each phosphate and resolves the excluded volume of dsDNA, including the major and minor grooves. The DNA model may need to be slightly modified to preserve quantitative agreement with thermodynamic properties, such as melting temperature. Nevertheless, this coarse-grained ions model should find application in the DNA modeling community, enabling existing coarse-grained DNA models to study phenomena where dsDNA condensation occurs, including viral packing and dsDNA toroids.

#### ■ ASSOCIATED CONTENT

##### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00341.

Effective potentials developed herein, radial distributions indicating the effective size of each ionic species, and

potentials of mean force calculated with the AMBER force field without NBFIX. It also contains additional figures demonstrating the convergence of the all-atom simulations and the transferability of the effective potentials to different temperatures and concentrations ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [depablo@uchicago.edu](mailto:depablo@uchicago.edu).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. Jejoong Yoo for providing all-atom topology files of dimethylphosphate and magnesium-hexahydrate and the prepublication NBFIX correction for polyamine nitrogen. We acknowledge the computational support of the University of Chicago Research Computing Center. We also acknowledge the compute resources of the University of Chicago Research Computing Center (RCC), as well as those of University of Wisconsin–Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW–Madison and the Wisconsin Alumni Research Foundation and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. D.M.H. was funded by a Graduate Research Fellowship from the National Science Foundation (NSF) (Grant No. DGE-1256259). This work is supported by the Department of Energy, Basic Energy Sciences, Materials Science and Engineering Division.

## REFERENCES

- (1) Tereshko, V.; Minasov, G.; Egli, M. *J. Am. Chem. Soc.* **1999**, *121*, 3590–3595.
- (2) Hud, N. V.; Vilfan, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 295–318.
- (3) Davis, J. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 668–698.
- (4) Keller, N.; delToro, D.; Grimes, S.; Jardine, P. J.; Smith, D. E. *Phys. Rev. Lett.* **2014**, *112*, 248101.
- (5) Berndsen, Z. T.; Keller, N.; Grimes, S.; Jardine, P. J.; Smith, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8345–8350.
- (6) Denisov, V. P.; Halle, B. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 629–633.
- (7) Frank, H. S.; Thompson, P. T. *J. Chem. Phys.* **1959**, *31*, 1086–1095.
- (8) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924.
- (9) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.
- (10) Feig, M.; Pettitt, B. M. *Biophys. J.* **1999**, *77*, 1769–1781.
- (11) Korolev, N.; Lyubartsev, A. P.; Laaksonen, A.; Nordenskiöld, L. *Biophys. J.* **2002**, *82*, 2860–2875.
- (12) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14771–14775.
- (13) Várnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269–4280.
- (14) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (15) Yoo, J.; Aksimentiev, A. *J. Phys. Chem. Lett.* **2012**, *3*, 45–50.
- (16) Yoo, J.; Aksimentiev, A., arXiv:physics/1510.02056, arXiv.org e-Print archive. <http://arxiv.org/abs/1510.02056>, accessed 10/7/15.
- (17) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.
- (18) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Biophys. J.* **2009**, *96*, 1675–1690.
- (19) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *J. Chem. Phys.* **2010**, *132*, 035105.
- (20) Savelyev, A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20340–20345.
- (21) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725.
- (22) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. *J. Chem. Phys.* **2011**, *134*, 085101.
- (23) Edens, L. E.; Brozik, J. A.; Keller, D. J. *J. Phys. Chem. B* **2012**, *116*, 14735–14743.
- (24) Hsu, C. W.; Fyta, M.; Lakatos, G.; Melchionna, S.; Kaxiras, E. *J. Chem. Phys.* **2012**, *137*, 105102.
- (25) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. *J. Chem. Phys.* **2013**, *139*, 144903.
- (26) Fan, Y.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L. *PLoS One* **2013**, *8*, e54228.
- (27) Naômé, A.; Laaksonen, A.; Vercauteren, D. P. *J. Chem. Theory Comput.* **2014**, *10*, 3541–3549.
- (28) Maffeo, C.; Ngo, T. T.; Ha, T.; Aksimentiev, A. *J. Chem. Theory Comput.* **2014**, *10*, 2891–2896.
- (29) Korolev, N.; Luo, D.; Lyubartsev, A. P.; Nordenskiöld, L. *Polymers* **2014**, *6*, 1655–1675.
- (30) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 3793–3807.
- (31) Prytkova, T. R.; Eryazici, I.; Stepp, B.; Nguyen, S.-B.; Schatz, G. C. *J. Phys. Chem. B* **2010**, *114*, 2627–2634.
- (32) Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. *J. Chem. Phys.* **2011**, *135*, 165104.
- (33) De Biase, P. M.; Solano, C. J.; Markosyan, S.; Czapla, L.; Noskov, S. Y. *J. Chem. Theory Comput.* **2012**, *8*, 2540–2551.
- (34) De Biase, P. M.; Markosyan, S.; Noskov, S. *J. Comput. Chem.* **2014**, *35*, 711–721.
- (35) Odijk, T. *J. Polym. Sci., Polym. Phys. Ed.* **1977**, *15*, 477–483.
- (36) Skolnick, J.; Fixman, M. *Macromolecules* **1977**, *10*, 944–948.
- (37) Savelyev, A.; Materese, C. K.; Papoian, G. A. *J. Am. Chem. Soc.* **2011**, *133*, 19290–19293.
- (38) Savelyev, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2250–2254.
- (39) Fernandez, D. P.; Goodwin, A. R. H.; Lemmon, E. W.; Levent Sengers, J. M. H.; Williams, R. C. *J. Phys. Chem. Ref. Data* **1997**, *26*, 1125–1166.
- (40) Freeman, G. S.; Hinckley, D. M.; Lequieu, J. P.; Whitmer, J. K.; de Pablo, J. J. *J. Chem. Phys.* **2014**, *141*, 165103.
- (41) Lenart, P. J.; Jusufi, A.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2007**, *126*, 044509.
- (42) Jusufi, A.; Hynninen, A.-P.; Panagiotopoulos, A. Z. *J. Phys. Chem. B* **2008**, *112*, 13783–13792.
- (43) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108.
- (44) Carmichael, S. P.; Shell, M. S. *J. Phys. Chem. B* **2012**, *116*, 8383–8393.
- (45) Chaimovich, A.; Shell, M. S. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1901–1915.
- (46) Pronk, S.; Pál, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.
- (47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (48) Maffeo, C.; Yoo, J.; Comer, J.; Wells, D.; Luan, B.; Aksimentiev, A. *J. Phys.: Condens. Matter* **2014**, *26*, 413101.
- (49) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (50) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (51) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci,

- B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Á.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian Inc.: Wallingford, CT, 2009.
- (53) da Silva, A. W. S.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 367.
- (54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (55) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (56) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (57) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (58) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (59) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (60) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–42.
- (61) Grønbech-Jensen, N.; Farago, O. *Mol. Phys.* **2013**, *111*, 983–991.
- (62) Grønbech-Jensen, N.; Hayre, N. R.; Farago, O. *Comput. Phys. Commun.* **2014**, *185*, 524–527.
- (63) <https://uchic.ag/3spn2>.
- (64) Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J. *J. Chem. Phys.* **2014**, *141*, 035102.
- (65) Freeman, G. S.; Lequieu, J. P.; Hinckley, D. M.; Whitmer, J. K.; de Pablo, J. J. *Phys. Rev. Lett.* **2014**, *113*, 168101.
- (66) Snodin, B. E.; Randisi, F.; Mosayebi, M.; Sulc, P.; Schreck, J. S.; Romano, F.; Ouldridge, T. E.; Tsukanov, R.; Nir, E.; Louis, A. A.; Doye, J. P. K. *J. Chem. Phys.* **2015**, *142*, 234901.
- (67) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (68) Yoo, J.; Aksimentiev, A. *J. Phys. Chem. B* **2012**, *116*, 12946–12954.
- (69) Lyubartsev, A. P.; Nordenskiöld, L. *J. Phys. Chem. B* **1997**, *101*, 4335–4342.
- (70) Korolev, N.; Lyubartsev, A. P.; Laaksonen, A.; Nordenskiöld, L. *Nucleic Acids Res.* **2003**, *31*, 5971–5981.
- (71) Korolev, N.; Lyubartsev, A. P.; Laaksonen, A.; Nordenskiöld, L. *Biopolymers* **2004**, *73*, 542–555.
- (72) Kahn, P. C. *Comput. Chem.* **1989**, *13*, 185–189.
- (73) Tinland, B.; Pluen, A.; Sturm, J.; Weill, G. *Macromolecules* **1997**, *30*, 5763–5765.
- (74) Manning, G. S. *Macromolecules* **2001**, *34*, 4650–4655.
- (75) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6185–6190.
- (76) Bloomfield, V. A. *Biopolymers* **1997**, *44*, 269–282.
- (77) Luan, B.; Aksimentiev, A. *J. Am. Chem. Soc.* **2008**, *130*, 15754–15755.
- (78) Grossfield, A. WHAM. <http://membrane.urmc.rochester.edu/content/wham> (accessed May 26, 2014).
- (79) Bloomfield, V. A. *Curr. Opin. Struct. Biol.* **1996**, *6*, 334–341.
- (80) Li, W.; Nordenskiöld, L.; Zhou, R.; Mu, Y. *Nanoscale* **2014**, *6*, 7085–7092.
- (81) Besteman, K.; Van Eijk, K.; Lemay, S. *Nat. Phys.* **2007**, *3*, 641–644.
- (82) Todd, B. A.; Adrian Parsegian, V.; Shirahata, A.; Thomas, T.; Rau, D. C. *Biophys. J.* **2008**, *94*, 4775–4782.
- (83) Dai, L.; Mu, Y.; Nordenskiöld, L.; van der Maarel, J. R. *Phys. Rev. Lett.* **2008**, *100*, 118301.
- (84) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533–3539.
- (85) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- (86) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provati, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (87) Im, W.; Seefeld, S.; Roux, B. *Biophys. J.* **2000**, *79*, 788–801.
- (88) Vitalis, A.; Baker, N. A.; McCammon, J. A. *Mol. Simul.* **2004**, *30*, 45–61.