A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions

Robert C. DeMille,[†] Thomas E. Cheatham III,[‡] and Valeria Molinero*,[†]

Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112-0850, United States, and Department of Medicinal Chemistry and Department of Pharmaceutics and Pharmaceutical Chemistry, College of Pharmacy, University of Utah, 2000 South 30 East, Skaggs Hall 201, Salt Lake City, Utah 84112, United States

Received: July 27, 2010; Revised Manuscript Received: November 22, 2010

Solvation by water and ions has been shown to be vitally important for biological molecules, yet fully atomistic simulations of large biomolecules remain a challenge due to their high computational cost. The effect of solvation is the most pronounced in polyelectrolytes, of which DNA is a paradigmatic example. Coarse-grained (CG) representations have been developed to model the essential physics of the DNA molecule, yet almost without exception, these models replace the water and ions by implicit solvation in order to significantly reduce the computational expense. This work introduces the first coarse-grained model of DNA solvated explicitly with water and ions. To this end, we combined two established CG models; the recently developed mW-ion model [DeMille, R. C.; Molinero, V. *J. Chem. Phys.* **2009**, *131*, 034107], which reproduces the structure of aqueous ionic solutions without electrostatic interactions, was coupled to the three-sites-per-nucleotide (3SPN) CG model of DNA [Knotts, T. A., IV; et al. *J. Chem. Phys.* **2007**, *126*, 084901]. Using atomistic simulations of $d(CGCGAATTCGCG)_2$ as a reference, we optimized the coarse-grained interactions between DNA and solvent to reproduce the solvation structure of water and ions around CG DNA. The resulting coarse-grained model of DNA explicitly solvated by ions and water (mW/3SPN-DNA) exhibits base-pair specificity and ion-condensation effects and it is 2 orders of magnitude computationally more efficient than atomistic models. We describe the parametrization strategy and offer insight into how other CG models may be combined with a coarse-grained solvent model such as mW-ion.

1. Introduction

The DNA molecule has been the subject of many molecular simulation studies in an attempt to understand its varied physical phenomena at atomic resolution.[1−11] This includes the development of coarse-grained (CG) models that improve the efficiency of DNA modeling by using different levels of coarsening to focus on different aspects of the physics of the DNA molecule, such as plasmid supercoiling,[12] protein docking,[13] persistence length,[14,15] and denaturation.[16−18] These CG models neglect explicit solvation for the sake of computational efficiency, choosing instead to implicitly define the interaction of DNA with water and ions. The structure and dynamics of DNA, however, are strongly influenced by solvation and ions, and the exclusion of specific hydration and ionic interactions has been shown to have deleterious effects on the structure of DNA.[19−21] Significant advances have been made in coupling a particle-based implicit solvent with atomistic DNA, resulting in about 5 times increase in efficiency with respect to atomistic simulations.[22] Another CG DNA model focusing on helix formation includes a similar explicit solvent but neglects features such as base-specificity and the counterions.[23] Due to their exclusion of explicit water and/or ions, these various models suffer from limited applicability for studying the structure of DNA in solution and processes that involve solvation/desolvation of the DNA.

Being a polyelectrolyte, the polyanionic nature of DNA and the closeness between the phosphate groups along the strand lead to counterion association. Described theoretically by Manning,[24,25] counterion condensation has been subject to many experimental and simulation studies. For a thorough review, we refer the reader to refs 19 and 26−30 and references therein. The attraction of the negatively charged phosphate moieties of DNA to the cationic counterions in solution involves strong electrostatic interactions. When solvated by water and cations, the long-range influence of the polyanion is effectively diminished due to the shielding of the charges. We hypothesize that the physical picture of the solvated, counterion-condensed DNA can be captured with a coarse-grained model with short-ranged interactions. Along these lines, we have recently developed a coarse-grained model of water and ions, the mW-ion model,[31] where each water molecule and ion (Na and Cl in ref 31) is represented by a single particle interacting through very short-ranged potentials. The model is based on the monatomic water model mW,[32] which reproduces the structure and anomalies of liquid water, ice, clathrates, and low-density amorphous ice and the phase transformations between them.[32−37] The key for the success of the mW model is its description of "hydrogen-bonded" structures through the use of a combination of two- and three-body potentials that encourage tetrahedral configurations of the water molecules.[32] The mW-ion model uses a similar strategy, plus the introduction of Yukawa potentials to modulate the repulsion between same-charge ions, to reproduce the density, radial and angular distribution functions, equilibrium between contact and solvent-separated ions pairs, and relative mobility of ions with respect to water of NaCl aqueous solutions up to a concentration of at least 5 M.[31] There are no electrostatics

* Corresponding author. E-mail: Valeria.Molinero@utah.edu.
[†] Department of Chemistry.
[‡] Department of Medicinal Chemistry and Department of Pharmaceutics and Pharmaceutical Chemistry.

10.1021/jp107028n
Published on Web 12/14/2010

or other long-ranged interactions in the mW and mW-ion models; chargeless ions interact with mW water and other ions via short-ranged potentials. We refer the reader to ref 31 for the full details of mW-ion, its parametrization, and its unique ability among coarse-grained models to replicate solution structure without hydrogen atoms and electrostatic interactions. Of particular note is the computational efficiency of the mW-ion model; molecular dynamics simulations with the mW-ion model are 2 orders of magnitude faster than with rigid atomistic water and ion models with Ewald sums. The increased efficiency is due to the lack of hydrogen atoms, increased time step, and the short range (between 4.3 and 7 Å) of the interactions.[31,32]

A natural progression and challenge is to interface the mW-ion model to coarse-grained models of biological molecules. Atomistic simulations, although becoming more negotiable with more efficient hardware and software implementations, are still intractable for long simulations (μs−ms) involving explicit solvation. On the other hand, typical CG implementations involving implicit solvation of DNA are too simplified to capture important interactions mediated by specific ion and water interactions, such as protein recognition and drug interactions. In the present work, we combine the mW-ion model with the well-performing and efficient 3SPN.0 coarse-grained DNA model and force field of Knotts and co-workers[17] to produce a coarse-grained model of nucleic acids with an explicit representation of water and ions that we name mW/3SPN-DNA. The 3SPN model reproduces several key features characteristic of DNA, including (i) sequence specificity by the incorporation of four distinct nucleotides; (ii) B-form helicity, structure, and stability through bonded and nonbonded interactions among phosphate, sugar, and base moieties; and (iii) semiquantitative reproduction of melting behavior. An improved version of this model, the 3SPN.1 DNA model, added an attraction between sugar moieties in opposing strands, achieving quantitative melting and rehybridization behavior when compared to the experiments.[38−40] The attraction between sugars is intended to emulate the effect of the solvent; thus, we base our parametrization on the 3SPN.0 model to avoid double counting of the effect of the solvent by its introduction in implicit and explicit terms of the potential. We note, however, that the methodology we present is general and could be extended to more recent parametrizations of this or other DNA models with comparable degrees of coarse-graining.

The central aim of this paper is to develop and validate the mW/3SPN-DNA coarse-grained model of DNA with explicit solvation by coarse-grained water and ions. A main question is whether the solvation structure of DNA can be accurately reproduced with a coarse-grained model without electrostatics. Here we show that, despite the replacement of long-ranged with short-ranged interactions and the coarse representation of water and DNA, the integrated mW/3SPN-DNA model is able to reproduce the solvation structure seen in atomistic simulations and the relative residence times of water and ions on the DNA moieties. The results suggest that the approach of this study could be extended to develop accurate solvated coarse-grained models of other complex biological systems, such as membranes, proteins, and RNA.

The paper is organized as follows: section 2 presents the models, the parametrization strategy, and the simulation methods; section 3 presents the results, with particular emphasis on DNA structure and stability, solvation structure, and solvation dynamics. After a brief discussion of the computational efficiency of the mW/3SPN-DNA model, the conclusions are presented in section 4.

2. Models, Parmeterization, and Simulation Methods

A. Models. ***Coarse-Grained DNA.*** The 3SPN force field involves bonded and nonbonded terms, corresponding to eqs 2a−2g of ref 17, including an implicit representation of solvent through Debye−Hückel Coulombic screening of the interactions between phosphates. For further insight into the nuances of the force field, including the Gō-type construction of the nonbonded interactions, we refer the reader to the original publication. In our implementation of the 3SPN model, we omit the shielded Coulomb interaction between phosphate sites (eq 2g of ref 17) as we explicitly include the ions and water molecules. Our initial tests with the original 3SPN model involving implicit solvation showed the Coulombic interaction to contribute only approximately 1% to the total energy of the system, and exclusion of the Coulombic interaction appears to have minimal effect on the structure of the DNA duplex, especially at high NaCl concentrations (∼1 M or greater).

We have used the classic Dickerson DNA dodecamer duplex $d(CGCGAATTCGCG)_2$ for the atomistic and coarse-grained simulations of this work. The middle panel of Figure 1 shows the CG sites of the 3SPN model superimposed on an all-atomistic representation of the DNA dodecamer. We followed the prescription for building molecules with standard coordinates, gleaned from the standard form the B-form helix of DNA (PDB ID: 1BNA),[41] previously described in ref 17. We note that this standard B-form structure used as reference for the parametrization of the 3SPN.0 model differs from the actual crystallographic structure,[42] which presents a modulation of the distances between the two strands along the minor groove, while this distance is constant in the standard model. The implications of this difference are discussed in section 3.A below.

Coarse-Grained Water and Ions. The monatomic water mW model[32] and its extension to ionic solutions, the mW-ion model,[31] were recently developed to study the behavior of aqueous ionic solutions at length and time scales that are not easily accessed by atomistic simulations. The mW and mW-ion use the short-ranged interactions of the Stillinger−Weber (SW) potential[43] and rely on the interplay between two-body attraction terms (ϕ_2), which favor high coordination, and three-body repulsion terms (ϕ_3), which encourage tetrahedral configurations, to reproduce the structure of liquid water and the ion solvation structure found in atomistic simulations and experiments. The functional form of the SW potential is shown in eq 1,

$$\begin{aligned} E &= \sum_i \sum_{j>i} \phi_2(r_{ij}) + \sum_i \sum_{j\neq i} \sum_{k>j} \phi_3(r_{ij}, r_{ik}, \theta_{ijk}) \\ \phi_2(r_{ij}) &= A\varepsilon_{ij}\left[B\left(\frac{\sigma_{ij}}{r_{ij}}\right)^p - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^q\right] \exp\left(\frac{\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}\right) \\ \phi_3(r_{ij}, r_{ik}, \theta_{ijk}) &= \lambda_{ijk}\varepsilon_{ijk}[\cos\theta_{ijk} - \cos\theta_{0ijk}]^2 \times \\ &\quad \exp\left(\frac{\gamma\sigma_{ij}}{r_{ij} - a_{ij}\sigma_{ij}}\right) \exp\left(\frac{\gamma\sigma_{ik}}{r_{ik} - a_{ik}\sigma_{ik}}\right) \end{aligned} \quad (1)$$

where r_{ij} is the distance between particles i and j and θ_{ijk} is the angle subtended by the vectors between the positions of the $i-j$ and $i-k$ pairs of particles. The constants that define the potential are $A = 7.049\,556\,277$, $B = 0.602\,224\,558\,4$, $p = 4$, $q = 0$, and $\gamma = 1.2$. The adjustable parameters are the characteristic size, σ; the interaction strength, ε; the strength of the tetrahedral interactions, λ; the cutoff parameter, a; and the preferred angle, θ_o. The water−water parameters were adjusted in ref 32 to reproduce the experimental temperature of melting,

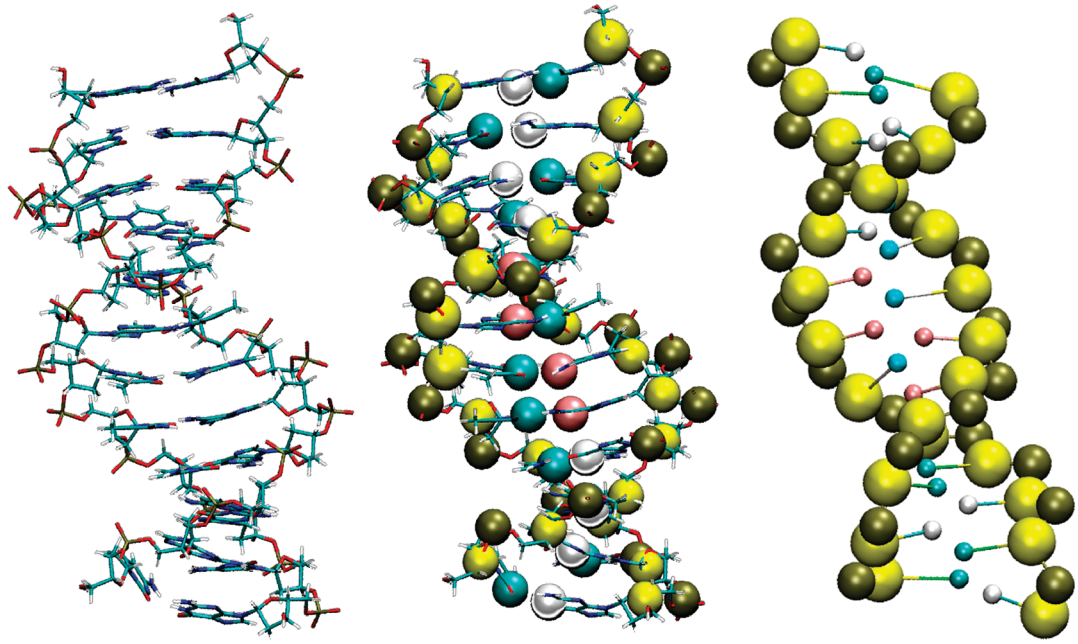

Figure 1. Representation of the d(CGCGAATTCGCG)$_2$ duplex for the atomistic model (left) and superimposition of 3SPN CG sites on top of the atomistic model (middle). (Right) A snapshot from a simulation of mW/3SPN-DNA used in this study, where the sizes of the bases have been reduced for clarity. Particle colors correspond to the following CG sites: G, white; C, dark blue; A, pink; T, light blue; P, olive green; and S, yellow.

density, and enthalpy of vaporization of liquid water. The water−ions and ion−ion parameters were tuned in ref 31 to reproduce the microscopic structure of NaCl−water solutions.

An additional shielded Coulomb Yukawa potential[44] is added to the interactions between cation pairs, as well as anion pairs,

$$E = C \exp(-\kappa r_{ij}) r_{ij}^{-1} \quad r_{ij} < r_{\text{cutoff}} \tag{2}$$

where the magnitude of C correlates with the strength of the shielded-Coulomb interaction and its sign indicates whether it is attractive or repulsive, κ is the screening length, and r_{cutoff} is a distance beyond which the contribution of this term to the energy is negligible. The values of r_{cutoff} in the mW-ion model do not exceed 7 Å, while the interactions described by eq 1 are even shorter ranged. The full details of the force field and the parameters used in the mW-ion model can be found in ref 31; these were used for water and NaCl in the present study.

Explicit Solvation of CG DNA. A combination of the two CG force fields requires careful attention to the cross interactions to provide the necessary balance between the stability of the DNA molecule and its solvation by water and ions. The model system used for parametrization of the solvation forces was a B-DNA model of the d(CGCGAATTCGCG)$_2$ duplex including 22 net-neutralizing Na "counterions" and an aqueous solution of 278 NaCl pairs in 15 000 mW molecules (representing ∼1 M NaCl). Simulations at a lower concentration of 200 mM NaCl were also performed, consisting of the DNA duplex solvated by 57 NaCl pairs in ∼15 400 waters. The interaction of water and ions with phosphate is described by eqs 1 and 2. The interaction of water and ions with the sugar and base sites involves a simplified form for the SW potential (eq 1) having only two-body interactions (i.e., $\lambda = 0$). The repulsive force between phosphate and chloride was additionally handled with a Yukawa potential (eq 2). A benefit of the CG model is that longer molecular dynamics integration time steps relative to all-atom simulations are possible while still conserving energy. The short-ranged nature of all interactions in the model, the decreased number of particles, and the longer time steps (5 fs) are responsible for its ∼90 times increased efficiency with respect to the atomistic model (see section 3.D). All interactions go effectively to zero between 4 and 9 Å. The full set of parameters used in this study for the mW/3SPN-DNA model is provided in Tables S1 and S2 of the Supporting Information.

Atomistic Models. All-atom (AA) simulations of explicitly solvated DNA were used as a reference for the parametrization of the mW/3SPN-DNA model. The initial AA model of d(CGCGAATTCGCG)$_2$ was built from the first model structure from the high-resolution NMR structures of Tjandra and co-workers (PDB ID: 1DUF),[45] and was solvated with TIP4P-EW water molecules[46] surrounding the DNA 15 Å in each direction into a truncated octahedral periodic box. Charge-neutralizing Na^+ ions were added, followed by the addition of 168 Na^+ and Cl^- ions for an approximate added salt concentration of ∼1 M[47] (33 NaCl pairs were used for a 200 mM salt solution). The position of each of the ions was randomized by swapping positions with random water molecules such that no ion was within 6 Å of another ion nor within 4 Å of the DNA (neglecting distance checks for ion−ion interaction across the periodic boundaries). When analyzing the AA trajectories, each was first mapped frame by frame into a CG trajectory using the same rules for assigning the CG sites prescribed by the 3SPN model: P at center of mass of the PO_4 group, sugar at center of mass of the deoxyribose unit, pyrimidine base sites C or T at the N3 position, and G and A at the N1 position for the purine bases. The AA water molecules were coarsened similarly by removing the hydrogen atoms.

B. Simulation Methods. The simulations of the coarse-grained mW/3SPN-DNA model were performed using the LAMMPS[48] molecular dynamics simulation package under isobaric−isothermal conditions (NpT) at $p = 1$ atm and at $T =$ 25 °C, except when otherwise indicated. The Nose−Hoover thermostat and barostat were used to control the temperature and pressure of the system, with time constants of 0.5 and 5.0

ps, respectively. We extended LAMMPS in order to incorporate the force field of the 3SPN.0 model. NaCl solutions of water, 0.2 or 1.0 M, were first equilibrated for 10 ns. Removing all water and ions in a cylinder centered in the simulation cell with rough dimensions 1.8 nm diameter by 4 nm height allowed the insertion of the dodecamer DNA strand. Equilibration of the system was then performed over 10 ns to allow the ions and water to completely solvate the mW/3SPN-DNA. The equations of motion were integrated with the velocity Verlet algorithm with a time step of 5 fs. Typical production runs were on the order of 100 ns.

The atomistic models of solvated DNA was simulated using a time step of 2 fs at a temperature of 300 K and 1 atm pressure using the Amber 10 package of molecular simulation programs.[49] The temperature was regulated with a Berendsen thermostat with a 5 ps decay constant.[50] Particle mesh Ewald[51] was used for electrostatic interactions with a cutoff of 9 Å and automated builds in a 1 Å buffered pairlist, cubic spline interpolation and an ~1 Å FFT grid. Trajectories comprising approximately 150 ns of simulation time for each ionic concentration were used for the data analysis presented here.

C. Parameterization. Three properties calculated from the AA trajectories with 1 M NaCl were used as targets for our parametrization efforts. We calculated radial distributions (rdf) of water and ions around the DNA CG sites, as well as the number of neighbors in the first shell in order to quantify the solvation structure around the DNA molecule. Additionally, as a measure of the strength of interactions between the CG sites with the mW-ion particles, residence times were calculated to determine the length of time that particles were associated with one another in their first solvation shells. These properties were calculated as we previously reported in the development and validation of the mW-ion force field.[31] The rdf, number of neighbors, and residence times were used as parametrization targets referenced against atomistic simulations with 1 M NaCl. The 1 M solution was chosen because it has a relatively high concentration of electrolyte, the regime for which the effective interactions between ions are expected to be shorter ranged and thus more amenable for its modeling through short-range potentials, while it is not high enough to shift the equilibrium between A- and B-forms of DNA. As molecules in CG models diffuse faster than in atomistic simulations, we did not attempt to reproduce the actual residence times for the ions and water in the first solvation shell of DNA but the *relative* times with respect to the mobility of water model in the system. In ref 31 we have shown that the ratio of diffusivity of ions with respect to water is well-reproduced by the mW-ion model, although the diffusion coefficients themselves are almost 3 times larger than in the atomistic simulations.

The primary method by which the force field was tuned to produce a better fit to AA data was in the adjustment of the parameters σ, ε, λ, and a of the solvent−DNA interaction potential. Upon examination of the atomistic data, we hypothesized the interaction of the water and ions with phosphate to play the most important role in the solvation forces. We initially focused on matching the solvation structure around the phosphate groups, leaving the interaction with the sugar and base groups for fine-tuning. Adjusting σ_{mW-P} and σ_{Na-P} in particular allowed the solvation shells of water and ions around mW/3SPN-DNA to closely align with the target data. Additional narrowing of the solvation shells, found to be necessary, for example, in the case of the interaction of mW with the CG phosphate moiety in order to match the atomistic data, was accomplished by lowering the value of the a parameter, directly affecting the length scale of the potential.

Adjustment of ε helped tune the number of neighbors in the solvation shells and allowed the residence time to more closely match the target. Again, we focused on the interaction between the solvating mW and ions with phosphate in our initial parametrization. Using sequential modifications to the value of ε_{mW-P} allowed the number of water neighbors surrounding mW/3SPN-DNA to be matched to the atomistic data. Generally, an increase in ε leads to an increase in the number of neighbors, as well as an increase in the residence time. This is, however, by no means a linear process; changing the parameters of one particular interaction often also influences others. The parameters of Tables S1 and S2 of the Supporting Information produce our best fit to the target data, though do not exhaustively explore the approximately 1000 tunable parameters available in the cross interactions of the water and ions with the DNA CG moieties.

D. Validation Measures. To assess the validity of the mW/3SPN-DNA model compared to atomistic data and to the implicitly solvated 3SPN.0 DNA model, we analyzed the structural and thermal stability of the DNA molecule, as well as the profile of water and ions surrounding its central axis for DNA in aqueous solutions of two ionic contents, 0.2 and 1 M NaCl. The structural stability of the DNA molecule is addressed in further detail below in section 3.A. The cylindrical distribution of water and ions around the curved central axis of DNA was calculated in the following fashion in order to show the short- and long-ranged structure of solvation of DNA. Disregarding the two terminal base pairs on either side of the DNA duplex of our study, we connected lines between consecutive base pairs. Next, the center of this line was translated to the origin, and the frame of reference rotated to set the line's terminal points at $\pm z_t$ (=1/2 the line distance), propagating the translation and rotation for the entire system of particles while maintaining the requirements of the periodic boundary conditions. For this arrangement, performed with each subsequent base pair for the entire trajectory, the radial distance $r_i = (x_i^2 + y_i^2)^{1/2}$ for all particles i with coordinates $|z| < z_t$ was binned every 0.1 Å to give a distribution of DNA particles, water, Na, and Cl around the actual central axis along the DNA oligomer. The data presented in section 3.B corresponds to averages over 100 ns simulation trajectories.

3. Results and Discussion

In developing an explicitly solvated model for DNA with CG water and ions, a critical aim was to retain the ability to accurately model the structure and thermal stability of the DNA molecule. Structure and stability results are compared to the explicitly solvated atomistic simulations and the implicit-solvation 3SPN.0 coarse-grained model. Additionally, the solvation structure of water and ions around the DNA molecule was replicated with the mW/3SPN-DNA model. The distribution of water and ions surrounding DNA, as well as a measure of the solvation dynamics—the residence times—further quantify the agreement between mW/3SPN-DNA and reference atomistic simulations. In what follows, we present the development and validation of the explicitly solvated coarse-grained DNA model in three subsections concerning (A) DNA structure and stability, (B) the solvation structure around DNA by water and ions, and (C) the residence times of water and ions on the DNA moieties.

A. DNA Structure and Stability. To measure the ability of mW/3SPN-DNA to maintain the helical structure of DNA, we compared the end-to-end distance of the d(CGCGAATTCGCG)$_2$ duplex with the atomistic simulations as well as the

TABLE 1: DNA Structural Stability of the Three Models Used in This Study for 1 M NaCl (data averaged 100 ns trajectories)

atomistic	3SPN.0	mW/3SPN-DNA
End-to-End Distance (Å)		
29.8 ± 0.1	33.4 ± 0.3	33.7 ± 0.2
Rmsd (Å) from Average Structure		
1.3 ± 0.3	2.8 ± 0.8	2.2 ± 0.4
Rmsd (Å) from Standard B-Form		
2.1 ± 0.4	4.6 ± 1.1	4.0 ± 0.5

3SPN.0 model. For this measurement, as well as any subsequent analysis of the DNA solvation structure, the first base pair on either end of the duplex was disregarded to preclude potential artifacts arising from the different solvation environment experienced at the termini (i.e., fraying), in accordance with previous studies.[19] Monitoring this end-to-end distance between the first guanine bases at the 5′ ends of the complementary strands gives a good comparison across the different models regarding the helical structure and compactness of the DNA duplex. This data is presented in Table 1 for the DNA in 1 M NaCl and in Table S3 of the Supporting Information for the 0.2 M NaCl solution. The standard structure the B-form duplex of DNA gives reference coordinates for the B-form Dickerson dodecamer[41] (which both the mW/3SPN-DNA and 3SPN.0 model use to build starting DNA structures) that yields a reference separation of ∼30.5 Å for these two G bases. The atomistic model maintains a slightly more compact structure with an end-to-end distance of 29.8 ± 0.1 Å in the 1 M and 29.7 ± 0.1 Å in the 0.2 M solution. The 3SPN.0 model shows a more elongated structure, averaging 33.4 ± 0.3 and 33.8 ± 0.2 Å, respectively. The slightly more elongated structure may be related to the small shifts in the backbone angles and dihedrals with respect to the reference standard B-form structure (Figures S1 and S2 of the Supporting Information). The mW/3SPN-DNA model replicates 3SPN.0 very closely (33.7 ± 0.2 Å in 1 M and 34.2 ± 0.2 Å in 0.2 M), showing that explicit solvation has no deleterious effect on the structure of the 3SPN DNA duplex. This suggests that the mW/3SPN-DNA model captures the subtle balance between the solvation forces of the duplex and the strength of DNA's stiffness inherent to the 3SPN.0 model.

To further compare the structure and stability of the DNA helix between the atomistic and coarse-grained models, we calculated the root-mean-squared deviations (rmsd) of representative 100 ns trajectories from both the trajectory-averaged structure and the standard B-form structure for each of the three models. The analysis was done in VMD 1.8.6,[52] and the rmsd data of the DNA helix can be found in Tables 1 and S3. The atomistic model shows the tighter fluctuations around its average structure, with an rmsd of about 1 Å. The coarse-grained models are softer and allow larger fluctuations, especially the 3SPN.0 model (∼3 Å rmsd). Addition of the explicit solvent with the mW/3SPN-DNA model damps these structural fluctuations, leading to an rmsd intermediate between the one of the fully atomistic model and that of the coarse-grained model with implicit solvation.

B-DNA presents major and minor grooves along the helix, where the backbones of the two strands are farther and closer apart, respectively. The upper panel of Figure 2 presents the minor groove average distance between the two strands along the dodecamer for the atomistic, mW/3SPN-DNA, and 3SPN models in 1 M NaCl, as well as the standard reference structure used to build and parametrize the coarse-grained models. We

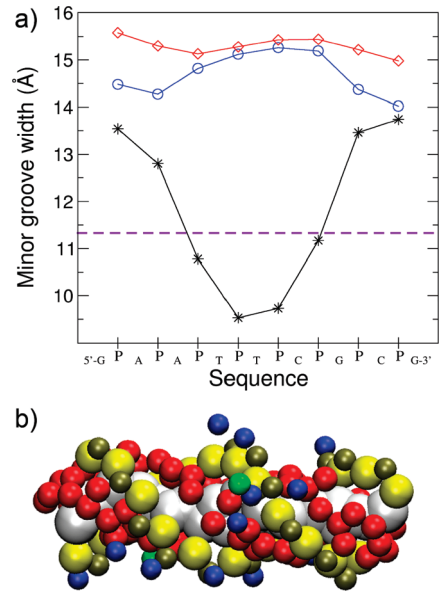

Figure 2. (a) Minor groove distances along the DNA duplex. Atomistic, 3SPN.0, and mW/3SPN-DNA data are symbolized by stars, diamonds, and circles, respectively. The constant minor groove distances of the reference model are shown by a dashed line. The coarse-grained models do not reproduce the thinning of the minor groove observed in the atomistic model and the experiment, presenting an almost constant distance similar to the reference B-DNA form structure used in the 3SPN.0 parametrization. (b) Snapshot of the solvated DNA dodecamer showing the water molecules (red) within 6 Å of both strands and Na (blue) and Cl (green) ions within 6 Å of either strand. DNA moieties are colored as in Figure 1, except for all bases are shown here in white for clarity.

evaluated the groove widths along the sequence by calculating the distance between a phosphate in the tagged strand and the closest phosphate in the other strand across the groove, averaging these quantities over the simulation trajectories at 25 °C (except for the reference B-form structure, which is a single configuration). Experimental results[42,53] and the atomistic simulations show a narrowing of the minor groove in the AATT regime of the dodecamer. This is not reproduced by the reference structure, which presents a constant minor groove width of 11.3 Å along the sequence of the duplex. Not surprisingly, a similarly constant trend is observed for the 3SPN.0 or mW/3SPN-DNA models constructed from the reference B-form structure. The coarse-grained model with implicit solvation shows a fairly constant distance along the sequence, with a value larger than the AA simulations and reference model. The larger distance may be due, in part, to the extension of the end-to-end distances in the coarse-grained DNA. The introduction of explicit water and ions into mW/3SPN-DNA decreases the phosphate−phosphate separation but also does not reproduce the sequence-specific narrowing of the AATT region. Different from the experiment and atomistic models, the minor and major grooves have almost the same width in the coarse-grained models. The resolution of the CG models is sufficient to differentiate the major and minor grooves; thus, the presence of grooves with similar widths is likely due to the form of the force field. Figure S2 of the Supporting Information presents distributions of the backbone dihedral angles for the atomistic and coarse-grained models that show that the CG DNA has unimodal distribution of dihedral angles (as commanded by the dihedral term in the 3SPN.0 potential), while the AA distribution of some dihedral angles is bimodal. The modulation of the width of the minor groove in the atomistic model is correlated to the number of water molecules (one and two) between the two strands along

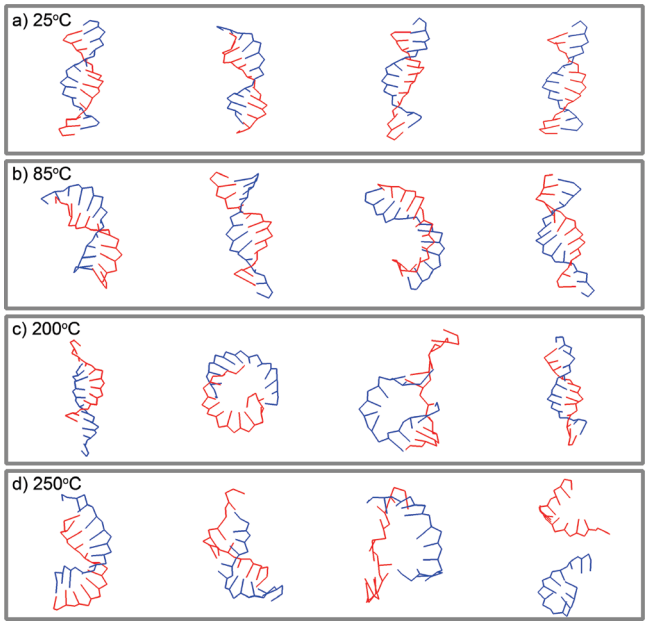

Figure 3. Representative snapshots from 100 ns mW/3SPN-DNA simulations of the Dickerson DNA dodecamer in 1 M NaCl at increasing temperatures. All simulations start from the equilibrated structure shown in the first snapshot of panel a. Note the presence of distorted structures at 85 °C (panel b) and the unzipping of the duplex at 200 °C (panel c), which, however, hybridizes back to a structure similar to its native form. At 250 °C (panel d) irreversible melting is witnessed after 10 ns. Complementary strands are colored in blue and red for clarity.

the sequence. The constant width of the minor groove in the mW/3SPN-DNA model results in about two water molecules between the strands, without significant variation along the sequence. This is observed in the lower panel of Figure 2 that shows a representative configuration of the dodecamer, the water molecules located between the strands, and the ions in the first solvation shell of the duplex. This suggests that an alternative pathway to reproduce the atomistic modulation of the groove width is to tune the water−base and water−sugar interactions. The results presented in subsection B below suggest that the representation of the sugars and, even more, the bases as single interaction sites is too coarse to adequately reproduce the details of their solvation structure.

The thermal stability of DNA was one of the main properties the 3SPN.0 model set out to reproduce. Through a sophisticated parallel tempering procedure, the 3SPN.0 model was parametrized to reproduce experimental melting temperatures of specific sequences. Although we did not include this as part of our parametrization of the mW/3SPN-DNA model, we have performed rudimentary tests to assess the model's stability as well as its ability to be denatured at higher temperatures. The experimental melting temperature of $d(CGCGAATTCGCG)_2$ is approximately 72 °C in 1 M NaCl.[54] In simulations at 85 °C, the explicitly solvated CG DNA distorts significantly by bending and untwisting the duplex; nevertheless, it does not dehybridize (see snapshots in panel b of Figure 3 and rmsd in Figure S8 of Supporting Information). To assess whether this lack of dehybridization is due to the presence of explicit solvent or insufficient sampling, we performed equivalent simulations of the implicitly solvated coarse-grained DNA sequence. A simulation of the 3SPN.0 sequence at 85 °C leads to unzipping within 10 ns. To assess the temperature at which the solvated coarse-grained DNA melts, we performed additional 100 ns simulations at 120, 150, 200, and 250 °C. Denaturation is not observed until 200 °C (panel c of Figure 3). Interestingly, denaturation is reversible at that temperature: the two strands are caged by the solvent and recombine within the simulation time scale. The coarse-grained solvated dodecamer irreversibly denatures at 250 °C. These results suggest that the melting temperature of the solvated model is between 200 and 250 °C, well above the melting temperature of the implicitly solvated duplex and the experimental value. The dehybridization of the solvated coarse-grained dodecamer starts in the middle region of the sequence. The average distances between complementary bases along the sequence at 25, 150, and 200 °C (presented in Figure S9 of the Supporting Information) show that while the CG ends remain at almost native distances up to 150 °C, the base pairing in the middle region weakens and extends considerably even before denaturation is observed. These results are consistent with the known ability of CG sequences to decrease fraying.

TABLE 2: Radial Distribution $g(r)$ Data for the 1 M NaCl Atomistic (AA) and mW/3SPN-DNA Models

	r (Å) of first shell peak	$g(r)$ value at peak	r (Å) of extent of first shell	first shell neighbors
		O−P		
AA	3.6	2.1	4.3	7.4
mW/3SPN-DNA	3.5	1.6	4.2	7.4
		Na−P		
AA	3.5	4.6	4.1	0.32
mW/3SPN-DNA	3.4	3.6	4.2	0.32
		Cl−P		
AA	6.0	0.5	6.7	0.13
mW/3SPN-DNA	6.3	0.7	7.4	0.25
		O−Sugar		
AA	4.6	0.9	6.2	16.5
mW/3SPN-DNA	4.3	1.0	5.9	15.4
		Na−Sugar		
AA	6.0	1.8	7.6	1.6
mW/3SPN-DNA	5.6	1.8	7.5	1.2

To determine whether the high melting temperature was due to direct (base−base) or indirect (water-induced) interactions, we performed an additional 100 ns simulation of the solvated model at 120 °C in which the base−base interactions between different strands were turned off, and no unzipping was observed. Our results suggest that the coarse-grained bases are too hydrophobic. The explicit inclusion of a water model that likes to associate with itself apparently disfavors the solvation of the hydrophobic base moieties. In principle, this could be corrected by an increase in the water−base attraction. In practice, such an increase in water−base attraction leads to a higher number of water molecules in the hydration shell of the bases that are already overestimated, as will be discussed in section B below. We conclude that the resolution of the bases should be increased, probably to no less than three coarse-grained particles per base, to better match the resolution of the solvent and allow for more favorable solvation free energies without an overall increase in the hydration number of the bases.

B. Solvation Structure. To quantify structural agreement in the solvation structure of the atomistic and coarse-grained DNA, we compared the radial distribution functions (rdfs) of atomistic solutions with those of mW/3SPN-DNA at 1 M NaCl concentration and focused primarily on the location of the first solvation shells of the CG sites. Additionally accessible from these rdfs is the number of neighbors in the first solvation shells, a measure that directly describes the solvation environment of the CG DNA molecule. Table 2 presents the number of first shell neighbors and other signatures from the rdfs for comparison of the

solvation structures of the AA and mW/3SPN-DNA simulations. The rdfs are presented in the Supporting Information (Figures S3−S7).

Hydration Structure. The atomistic simulations show water to have strong affinities to the phosphate group and some of the bases. The first shell surrounding the phosphate moiety contains an average of 7.4 water molecules up to 4.3 Å. Water associates neither closely nor strongly with the sugar groups of the atomistic DNA, giving rise to an extended first shell of water molecules surrounding sugars that are also closely associated with their bridging phosphate groups. An average of 16.5 water molecules can be found within 6.2 Å of the sugar moieties. Water has an interesting interaction with the bases in the atomistic simulations. Thymine, featuring, as a dione, two highly electrophillic sites, shows a very distinguished first shell (see rdfs, Figures S3−S7) extended to 5.2 Å. The other bases show peaks in the rdfs at around 5 Å, yet they do not differentiate a solvation shell as rigorously.

Solvation of the DNA duplex by water is very similar in the mW/3SPN-DNA model and in the AA simulations (Table 2). The interaction between mW water and the CG phosphate is equivalent in terms of the solvating first-shell neighbors: 7.4 at 4.2 Å for the CG, compared with 7.4 at 4.3 Å in AA. The position of the maxima and minima in the first peak of the water−phosphate rdf of the CG and AA models are identical within 0.1 Å. The agreement is almost equally good for the sugar−water interactions: the number of waters in the first shell is 15.4 for the CG and 16.5 for the AA, with a shift in minima and maxima within 0.3 Å. The slight deviations observed in the positions of the first peaks of the water−phosphate and water−sugar rdfs are a result of the coarseness of the model in representing the irregular hydrophilic environment of the atomistic moieties with a single bead. As expected, the more complex the atomistic group represented by a single particle is, the more difficult it is to reproduce its solvation rdf. Thus, it is not surprising that the agreement degrades for the solvation of the bases, represented in the coarse-grained model by a single particle. The mW/3SPN-DNA model exhibits a more featureless rdf of water with the bases compared to the atomistic data (see Supporting Information). For the three bases that do not present well-defined first shell solvation peaks (A, C, and G), the mW/3SPN-DNA model reproduces the water−base distances and the shape of the rdf but systematically overestimates the hydration shell: at 5.75 Å there are about eight water molecules in the hydration shell of all CG bases, while there are about six for the atomistic A, G, and C and ∼4 for T. There is a clear inverse correlation between the complexity of the actual solvation shell of the bases at the atomistic level and the ability of the coarse-grained model to reproduce it with a single site base. When the degree of coarsening of the DNA moiety is comparable to that of water in mW (e.g., phosphate), the agreement in solvation of the DNA by the coarse-grained model is remarkable. Considering the lack of hydrogens and electrostatics in the CG model of water and ions, this approach holds promise for expansion of explicit solvation by mW to other CG biomolecules and suggests that the best results will be obtained when the degree of coarsening of the biomolecule is comparable to the one of the mW model (i.e., beads effectively representing a few atoms).

Ionic Condensation. The condensation of sodium ions around the negatively charged DNA has been extensively studied,[19,24,26−29,55] and is witnessed in the atomistic simulations presented here. In spite of its lack of electrostatics, the coarse-grained model faithfully reproduces the position of the maximum, minimum, and number (0.32) of sodium neighbors in the first shell of the DNA's phosphate groups (Table 2). The sodium ions show very little affinity toward the sugar groups, exhibiting a very extended, featureless shell up to 7.5 Å with an average of 1.6 Na^+ neighbors in the AA and 1.2 in the mW/3SPN-DNA models. The ratio of water to ions in the first solvation shell of the sugars is about 10:1.

The bases T and G show a well-defined peak around 4.5 Å in the atomistic sodium−base rdfs, due to direct contact of the cation with the base. The high degree of coarsening for the bases, again, poses a challenge for an accurate representation of the cation−base rdf in the mW/3SPN-DNA model as the irregular, anisotropic, nucleophilic environment of each base is not sufficiently well-represented by a spherical potential. The mW/3SPN-DNA model cannot capture the Na−base contact peak well: although there is some interaction of Na with the bases at close range—particularly with G—the features of the sodium solvation of the bases are broader and not as defined as the atomistic data (see rdfs in Supporting Information). The success of the mW-ion in replicating the hydration structure of Na and Cl ions suggests that the lack of agreement in mW/3SPN-DNA is mainly due to a too coarse level of description of the bases. The implementation of explicit coarse-grained solvation to more detailed coarse-grained models of DNA such as, for example, that presented in ref 13 or 56 should improve the agreement.

Replicating the different affinities of sodium and chloride ions for DNA is a particular challenge due to the lack of electrostatic interactions in the mW-ion model. Our results for the 1 M solution show that the solvation shell around the phosphate ions is very well reproduced by the coarse-grained model. mW/3SPN-DNA reproduces the number of water, sodium, and chlorides around the phosphates (P) within 7% of the atomistic model. The average distances between the phosphates and the solvent moieties reflect a strong preference of P for water, which on average is even closer to the phosphate than the sodium counterion. The reason is probably a balance of Na−P direct attraction and the electrostatic repulsion that P experiences with Cl that, in turn, associates with Na (this is more readily observed in the cylindrical distribution of ions around DNA, presented below). The result is that the chloride ions are farther from DNA, due to the electrostatic repulsion of the negative charges. This is evident in the radial distribution Cl−P, where the peak moves toward larger distances in the atomistic simulations. This is well-captured by the mW/3SPN-DNA model.

Cylindrical Distributions of Ions around DNA. To further understand the solvation structure of water and ions around the DNA molecule, we calculated cylindrical distributions around the curved central axis of the DNA base pairs. The cylindrical distributions from the atomistic and mW/3SPN-DNA simulations with 1 M NaCl are presented in Figure 4. The mW/3SPN-DNA simulations replicate the features seen in the atomistic model, despite the lack of electrostatics. Measuring across the helix, the width of the coarse-grained DNA molecule is about 1 Å wider than that of the atomistic DNA (on the order of a C−C single bond). This is consistent with the slightly elongated end-to-end distance observed with the CG model. Atomistic data shows water to be closely hydrating the helix, with peaks corresponding to solvating base, sugar, and phosphate moieties at 5, 8, and 12 Å, respectively. The positions of these peaks and their intensities are well-reproduced by the mW/3SPN-DNA model.

The interaction of sodium is very similar between the atomistic and coarse-grained models. The mW/3SPN-DNA

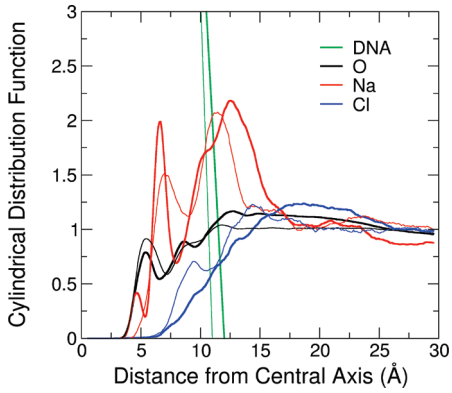

Figure 4. Cylindrical distribution of water and ions around the d(CGCGAATTCGCG) duplex, calculated for each segment along the curvilinear path of the center of sequential base-pairs (i.e., the central axis) and averaged over representative 100+ ns trajectories. Atomistic data are symbolized by thick lines and mW/3SPN-DNA by thin lines. All the DNA CG sites are represented in a single curve (green), with the atomistic showing a wider radius across the helix. Distribution of water (O) is shown in black, sodium (Na) in red, and chloride (Cl) in blue.

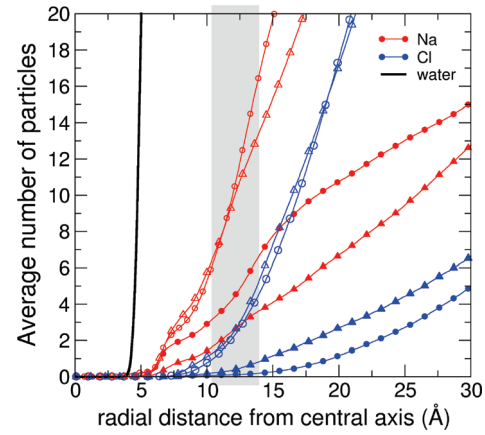

Figure 5. Average number of water, sodium, and chloride atoms surrounding DNA as a function of radial distance from its central axis for 1 M and 200 mM NaCl concentrations. Water, sodium, and chloride correspond to black, red, and blue curves, respectively. Atomistic data are circles and mW/3SPN-DNA triangles. Filled symbols correspond to 200 mM NaCl and empty symbols to 1 M NaCl. Note: the four data sets for water are indistinguishable on the scale of this graph and are therefore presented as a single curve for clarity. The gray box indicates the average location of the phosphate moieties of the DNA strands.

model does not quantitatively reproduce the Na−base distribution (see discussion above), yet the Na profile around DNA is dominated by the interaction with the phosphate and the sugar, which are both well-reproduced by the coarse-grained model (Table 2). Sodium solvation of the bases, sugar, and phosphate moieties appear as three peaks in the cylindrical distribution. A small inner peak at ∼4 Å, corresponding to the solvation of the bases by sodium, is not reproduced by the coarse-grained model. This is consistent with the poor representation of the sodium−base rdf discussed above. A second peak around 6 Å is due to the Na−sugar attraction and is well-reproduced by the mW/3SPN-DNA model (the atomistic peak is sharper, but it integrates to the same number of neighbors). The third and most pronounced peak at ∼12.5 Å is due to their interaction of sodium with the phosphates. The mW/3SPN-DNA model reproduces the distribution of sodium around DNA of the fully atomistic model in 1 M NaCl. A noteworthy feature of the sodium distribution is how short-ranged it is: the density of ions reaches the average value already within 5 Å of the DNA perimeter. This evidences a high degree of counterion condensation around the DNA, similar to previous studies reported in the literature.[19,24,26−28]

Chloride ions are repelled by the negatively charged atomistic DNA. This is evident in the cylindrical distribution by a depleted chloride density (i.e., the distribution function has a value less than 1) within the region occupied by DNA. On the other hand, the chloride anions are attracted to the sodium cations, leading to a layering of charged species of which the chloride forms the outer shell. This produces a modest increase in the density of Cl^- outside the outermost Na^+ peak (Figure 4). The same features (depletion within the DNA region and slight increase outside the sodium peak) are observed in the cylindrical distribution of chloride in the mW/3SPN-DNA model. The small peak around 8 Å in the depleted region of the coarse-grained model is due to a stronger Na−Cl attraction in mW-ion model than in the atomistic model. The Na−Cl interaction was purposefully decreased in the atomistic model[57] to avoid crystallization of the salt. The mW-ion model, which has a stronger Na−Cl interaction than the atomistic one, reproduces the experimental[58] association constant of NaCl in water.[31]

The agreement between atomistic and coarse-grained models is the best in concentrated electrolyte solutions. This is appreciated in Figure 5, which presents the integrated average number of solvent molecules surrounding the central axis of the DNA strand as a function of the distance to the helical axis. The agreement between the atomistic and mW/3SPN-DNA simulations is excellent for the 1 M solution, as seen also in Figure 4. Simulations at 0.2 M salt concentration show, both in the atomistic and coarse-grained models, qualitatively similar trends when compared to the 1 M solutions: there is less sodium condensation around the DNA in the dilute solution, and the chlorides are farther from the duplex. The change in ionic condensation from 0.2 to 1 M is not in accordance with what is expected by counterion condensation theory (valid at low electrolyte content),[24,25] which predicts the neutralization of the polyelectrolyte chain to be mostly invariant with salt concentration. Although the coarse-grained model reproduces the qualitative distribution of NaCl around the double helix for the 0.2 M solution, the agreement with the atomistic model is not quantitative as for the 1 M one: Figure 5 shows that at 0.2 M the CG model under-represents the amount of Na ions condensed onto the DNA and over-represents the distribution of chloride close to it. This degradation of the agreement in the more dilute solution is a limitation of the use of short-ranged potentials for the description of the interactions between ions.

The comparison between the water and ion distribution around the atomistic and mW/3SPN-DNA data is very favorable, especially for solvation by 1 M NaCl. This may be surprising, considering that DNA is a polyelectrolyte, the model has no charges or long-ranged interactions, and that this distribution is not a feature that we included in refining the parameters of the model. The good agreement in the concentrated solution is due to the lack of long-range ionic structure around the DNA, as the atomistic ions condense around the highly charged polyelectrolyte. When the polyelectrolyte is surrounded by water and condensed cations, the effective length of the interactions, described in the atomistic regime by long-ranged Coulombic interactions, is reduced. When the ionic concentration decreases, on the other hand, and the ionic distribution extends farther from the polyelectrolyte, the coarse-grained model still captures the qualitative distribution of ions but underestimates the strong depletion of Cl around the helix because being short-ranged it cannot account correctly for the less shielded, longer ranged,

TABLE 3: Residence Time (in ps) for the 1 M NaCl Atomistic (AA) and mW/3SPN-DNA Models

moieties	AA	mW/3SPN-DNA[a]
O−P	38	33
Na−P	210	100
Cl−P	26	33
O−sugar	43	43

[a] CG residence times were multipled by 2.5 to compare to AA data (see the text).

repulsion between the phosphates and the chloride. The mW-ion model that describes the solvent in this work shows good transferability from 0.66 to 5.6 M NaCl, the range of its original parametrization. It may be possible to parametrize the interactions to reproduce more faithfully the distribution of ions for dilute solutions; we expect that this would result in longer ranged interactions that would compromise the efficiency of the model.

C. Solvation Dynamics. We used the residence time of water and ions around the DNA molecule to determine how well the mW/3SPN-DNA model captures the solvation dynamics. This is an important feature of the solvation, as it may control the accessibility of other molecules (e.g., drug molecules, proteins, other nucleic acids) to DNA. Comparing the relative times that the water and ions are associated with the DNA moieties in the atomistic model gave us also a measure of the relative interaction strengths for use in developing the solvent−DNA parameters. These calculations are presented in Table 3 for the various interactions of water, ions, and the DNA backbone in the 1 M NaCl solution. In comparing atomistic data with that of the mW/3SPN-DNA model, we point out that CG models are inherently faster than their atomistic counterparts. This is due to the evolution of the CG system on a smoother potential energy surface. For the mW-ion model,[31] we found the diffusion of water in 1 M NaCl at 300 K to be 2.5 times faster than the atomistic (TIP4P-Ew). Thus, using the diffusivity of water to normalize the time scales, our target was to obtain residence times in the mW/3SPN-DNA model that are 2.5 times shorter than in the fully atomistic simulation. We therefore multiply the residence times we calculated for the mW/3SPN-DNA model by this factor so as to compare the dynamics of the two models directly. The residence times were computed, following the protocols of ref 31, as the average time it takes for the ion (water) to move to a distance further than the first minimum of the rdf of the DNA−ion (DNA−water). We computed residence times only for pairs that form well-defined first peaks in the rdf.

The shortest residence times are, not unexpectedly, between the anions: the residence time of Cl around phosphate is ∼30 ps. This is comparable to ∼35 ps, the average time it takes for water to diffuse over 7 Å, the width of the first peak of the Cl−P rdf. We note that this first peak encompasses distances for which Na or water mediates the two anions; we do not find contact pairs between chloride and phosphates.

The residence times for water on the phosphate and sugar are similar at around 40 ps. As the width of the water−phosphate peak in the rdf is about 4.2 Å, this residence time is about twice as long as the diffusional time for water in the bulk over the same distance, consistent with a strong water−phosphate attraction. The width of the water−sugar peak is about 7 Å; thus, these results indicate that water is not particularly attached to the sugars of DNA.

The longest residence times are for Na with the phosphate ions of the DNA, as is expected due to their strong attraction. The coarse-grained model underestimates the residence time for Na−P. Although mW/3SPN-DNA reproduces the number of neighbors and location of maxima and minima of the first peak, it predicts the peak to be less sharp than for the atomistic model. It should be explored in the future whether the number of neighbors and the scaled residence times for Na−P can be simultaneously reproduced by the coarse-grained model. Except for the Na−P pair, the scaled residence times of the mW/3SPN-DNA are within 10−20% of the atomistic; thus, we expect that processes dominated by desolvation of DNA should be qualitatively reproduced by the coarse-grained model.

D. Benchmarking. Comparison of mW/3SPN-DNA to all-atom simulations shows a sizable increase in computational performance. For the system sizes of this study, the explicitly solvated coarse-grained model is 90 times faster than the atomistic model (i.e., it takes 90 times more CPU time to compute a given simulation time, e.g., 1 ns, with the atomistic model). Most of the expense of the mW/3SPN-DNA simulations is from the explicit representation of water and ions, which, as we showed for mW-ion,[31] scales approximately as $N \log N$ (where N is the number of coarse-grained particles in the system). It is possible for our simulations with mW/3SPN-DNA to extend into the microsecond time frame with only 7 days of wall time across 48 Intel Xeon E5430 (2.66 GHz) processors for a double-stranded DNA solvated by ca. 16 000 molecules of water and ions.

4. Conclusions

In this work, we have combined a coarse-grained model of DNA with a coarse-grained model of water and ions to produce a coarse-grained model of DNA with explicit solvation. Using an established CG model of DNA—the 3SPN.0 model of Knotts et al.—allowed us to tap the wealth of knowledge already carefully developed into its force field and parameters. Our focus was mostly involved in adapting the mW-ion model—developed to introduce coarse-grained ions to the monatomic water (mW) model—to interact with the DNA molecule so as to retain its stability yet also exhibit the nuances of solvation. The multitude of parameters of the mW/3SPN-DNA model has not been fully explored, yet the results show a very good empirical match of the solvation structure to the atomistic reference. The model and parameters presented are our best empirical fit to the atomistic data.

Replicating the structure of the solution of ions and water surrounding the DNA molecule was a primary goal of parametrization and is measurable from the radial distribution functions with features such as the location of the solvation shells and number of solvating neighbors. Our experience with the mW-ion model allowed us to describe well the solvation of the phosphates with the coarse-grained water and ions, despite the use of extremely short-ranged forces (i.e., no electrostatics). The results suggest that the mW/3SPN-DNA model is able to capture the long-ranged structure of water and ions around the DNA molecule, as shown by the cylindrical distribution. Moreover, the explicit solvation of the CG DNA actually helps stabilize the molecule, as evidenced by the damping of its structural fluctuations and the increase in the melting temperature.

We found an inverse correlation between the ability of the coarse-grained model to describe the radial distribution functions and the degree of coarsening of the DNA moiety involved; the agreement between atomistic and mW/3SPN-DNA models was very good for rdfs involving phosphates, slightly less for the sugars, and the least for the bases. Despite the differences in size and complexity of the atomistic representations of phosphate, sugar, and the four bases, each of these is described by

a single particle in the present CG model. The very good agreement in the description of the water−water, water−salt, and water−phosphate structures indicates that the model could be improved by increasing the resolution of the bases and the sugars. An increase in the number of particles per coarse-grained base and sugar (to, for example, two and three for the bases, as in ref 56 or three for sugar as in the M3B model[59]) would marginally affect the efficiency of the simulations, as most of the computational cost is attributable to the solvent.

Ions and solvent play a crucial role in the melting transition: Schatz and co-workers have shown that the presence of explicit ions in coarse-grained simulations of DNA sharpens the melting transitions,[18,60] and De Pablo and co-workers have shown that addition of a mid-range sugar−sugar attractive term to the implicitly solvated coarse-grained DNA force field achieves quantitative melting temperatures and a good description of the renaturation process.[38−40] It may thus seem paradoxical that a model with explicit solvation, such as mW-DNA/3SPN, performs less well in the prediction of the melting temperatures than the implicitly solvated 3SPN. Our results strongly suggest that the high melting temperature of the mW/3SPN-DNA Dickerson dodecamer (between 200 and 250 °C, higher than for the implicitly solvated 3SPN model and the experiment) is due to a strong solvent-induced hydrophobic attraction between the bases in the presence of explicit water. An increase in the water−base attraction would lead to a lower melting point, by favoring the solvation of the single strands. Nevertheless, we find that the current parameters already overestimate the number of water molecules in the hydration shell of the bases, and this would be exacerbated with an increase in water−base attraction. This poses a representability problem: the model cannot simultaneously represent the melting temperature and the solvation of the bases. We conjecture that commensurate accurate melting temperatures and solvation could be obtained through a finer resolution of the bases (at least 3 sites per base) that would allow for the distinction of its internal hydrophilic and hydrophobic moieties and a better resolution of their shape (i.e., excluded volume) as experienced by a water molecule.

The results of this work show that a coarse-grained model with very short ranged interactions, no hydrogen atoms, and no electrostatics is able to reproduce rather quantitatively the distribution and relative residence times of water and ions around DNA. The strategy developed in this work could be applied to the coarse-grained solvation of proteins, RNA, and other biomolecules. We expect the solvation of proteins to be the most challenging, because of the fact that the density of ionic moieties is generally much lower in proteins than in DNA and that part of the success of the present model is based on the condensation of the counterions around the polyelectrolyte. We envision the explicitly solvated mW/3SPN-DNA model to have the greatest applicability in model systems where hydrophobic/hydrophilic interactions and solvation play a large role in the microscopic behavior and the role of the solvent cannot be ignored. Examples of this include interactions of DNA with drug molecules, surfaces, micelles, and proteins.

Acknowledgment. We thank Thomas A. Knotts IV and Terry Schmitt for sharing the 3SPN code and for insightful comments and discussions. We acknowledge the support of the National Science Foundation through Collaborative Research Grant CHE-0628257 (VM), a TeraGrid award of computer time TG-MCA01S027 (TEC3), and the National Institutes of Health GM-081411 (TEC3). We thank the Center of High Performance Computing at the University of Utah for allocation of computing time.

Supporting Information Available: Expanded data sets characterizing the stability of the DNA molecule, graphed radial distributions, the full parameter set of the mW/3SPN-DNA model, and analysis of simulations of DNA solvated in a 200 mM NaCl solutions. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III *Acc. Chem. Res.* **2000**, *33*, 889.
(2) Cheatham, T. E., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 360.
(3) Orozco, M.; Noy, A.; Perez, A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185.
(4) Mackerell, A. D., Jr.; Nilsson, L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194.
(5) Hashem, Y.; Auffinger, P. *Methods* **2009**, *47*, 187.
(6) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182.
(7) Kenward, M.; Dorfman, K. D. *J. Chem. Phys.* **2009**, *130*, 10.
(8) Malikova, N.; Cadene, A.; Marry, V.; Dubois, E.; Turq, P.; Zanotti, J. M.; Longeville, S. *Chem. Phys.* **2005**, *317*, 226.
(9) Bruant, N.; Flatters, D.; Lavery, R.; Genest, D. *Biophys. J.* **1999**, *77*, 2366.
(10) Locker, R.; Harvey, S. C. *Biophys. J.* **2005**, *88*, 229A.
(11) Olson, W. K.; Zhurkin, V. B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 286.
(12) Trovato, F.; Tozzini, V. *J. Phys. Chem. B* **2008**, *112*, 13197.
(13) Poulain, P.; Saladin, A.; Hartmann, B.; Prevost, C. *J. Comput. Chem.* **2008**, *29*, 2582.
(14) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *J. Chem. Phys.* **2010**, *132*, 035105.
(15) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.
(16) Drukker, K.; Schatz, G. C. *J. Phys. Chem. B* **2000**, *104*, 6108.
(17) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 12.
(18) Lee, O.-S.; Prytkova, T. R.; Schatz, G. C. *J. Phys. Chem. Lett.* **2010**, *1*, 1781.
(19) Cheatham III, T. E.; Young, M. A. *Biopolymers* **2000**, *56*, 232.
(20) Stefl, R.; Cheatham, T. E.; Spackova, N.; Fadrna, E.; Berger, I.; Koca, J.; Sponer, J. *Biophys. J.* **2003**, *85*, 1787.
(21) Spackova, N.; Cheatham, T. E.; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759.
(22) Basdevant, N.; Ha-Duong, T.; Borgis, D. *J. Chem. Theory Comput.* **2006**, *2*, 1646.
(23) Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2005**, *122*, 124906.
(24) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924.
(25) Manning, G. S. *Q. Rev. Biophys.* **1978**, *11*, 179.
(26) Spasic, A.; Mohanty, U. Counterion condensation in nucleic acid. In *Advances in Chemical Physics*; John Wiley & Sons Inc: New York, 2008; Vol. 139; pp 139.
(27) Volker, J.; Klump, H. H.; Manning, G. S.; Breslauer, K. J. *J. Mol. Biol.* **2001**, *310*, 1011.
(28) Muthukumar, M. *J. Chem. Phys.* **2004**, *120*, 9343.
(29) Record, M. T.; Zhang, W. T.; Anderson, C. F. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: A practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. In *Advances in Protein Chemistry*; Academic Press Inc: San Diego, 1998; Vol. 51; pp 281.
(30) Hud, N.; Polak, M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 293.
(31) DeMille, R. C.; Molinero, V. *J. Chem. Phys.* **2009**, *131*, 034107.
(32) Molinero, V.; Moore, E. *J. Phys. Chem. B* **2009**, *113*, 4008.
(33) Moore, E. B.; Molinero, V. *J. Chem. Phys.* **2009**, *130*, 244505.
(34) Jacobson, L. C.; Hujo, W.; Molinero, V. *J. Phys. Chem. B* **2009**, *113*, 10298.
(35) Jacobson, L. C.; Molinero, V. *J. Phys. Chem. B* **2010**, *114*, 7302.
(36) Moore, E. B.; Molinero, V. *J. Chem. Phys.* **2010**, *132*, 244504.
(37) Jacobson, L. C.; Hujo, W.; Molinero, V. *J. Am. Chem. Soc.* **2010**, *132*, 11806.
(38) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Biophys. J.* **2009**, *96*, 1675.
(39) Sambriski, E. J.; Ortiz, V.; de Pablo, J. J. *J. Phys.: Condens. Matter* **2009**, *21*, 034105.
(40) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Proc. Natl. Acad. Sci.* **2009**, *106*, 18125.
(41) Arnott, P. J. S.; Smith, C.; Chandrasekaran, R. Atomic coordinates and molecular conformations for DNA−DNA, RNA−RNA, and DNA−

RNA helices. In *CRC Handbook of Biochemistry and Molecular Biology*; 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, 1976; Vol. II; pp 411.

(42) Shui, X.; McFail-Isom, L.; Hu, G.; Williams, L. *Biochemistry* **1998**, *37*, 8341.

(43) Stillinger, F. H.; Weber, T. A. *Phys. Rev. B* **1985**, *31*, 5262.

(44) Carre, A.; Berthier, L.; Horbach, J.; Ispas, S.; Kob, W. *J. Chem. Phys.* **2007**, *127*, 114512.

(45) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 6190.

(46) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665.

(47) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020.

(48) Plimpton, S. J. *J. Comput. Phys.* **1995**, *117*, 1.

(49) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California, San Francisco, 2008.

(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684.

(51) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 4.

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

(53) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 6190.

(54) Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Broka, C.; Rice, J. A.; Itakura, K.; Breslauer, K. J. *Biochemistry* **1982**, *21*, 428.

(55) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89*, 3721.

(56) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 14.

(57) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020.

(58) Fuoss, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 34.

(59) Molinero, V.; Goddard, W. A. *J. Phys. Chem. B* **2004**, *108*, 1414.

(60) Prytkova, T.; Eryazici, I.; Stepp, B.; Nguyen, S.; Schatz, G. C. *J. Phys. Chem. B* **2010**, *114*, 2627.

JP107028N