Chemically accurate coarse graining of double-stranded DNA

Alexey Savelyev and Garegin A. Papoian[1]

Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

Edited* by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved October 1, 2010 (received for review February 12, 2010)

Coarse-grained (CG) modeling approaches are widely used to simulate many important biological processes involving DNA, including chromatin folding and genomic packaging. The bending propensity of a semiflexible DNA molecule critically influences these processes. However, existing CG DNA models do not retain a sufficient fidelity of the important local chain motions, whose propagation at larger length scales would generate correct DNA persistent lengths, in particular when the solution's ionic strength is widely varied. Here we report on a development of an accurate CG model for the double-stranded DNA chain, with explicit treatment of mobile ions, derived systematically from all-atom molecular dynamics simulations. Our model generates complex local motions of the DNA chain, similar to fully atomistic dynamics, leading also to a quantitative agreement of our simulation results with the experimental data on the dependence of the DNA persistence length on the solution ionic strength. We also predict a structural transition in a torsionally stressed DNA nanocircle as the buffer ionic strength is increased beyond a threshold value.

molecular simulations | polymer chain | one-step renormalization | chemical coarse graining | explicit ion electrostatics

To successfully model many biological processes involving DNA, atomistic or coarse-grained (CG) DNA force fields should reproduce the bending rigidity of a semiflexible DNA chain, which is a large-scale polymer property of critical significance (1, 2). In particular, the response of the DNA persistence length to the change in the surrounding ionic environment is of major biological importance. For example, even a slight change in the persistence length of a linker DNA segment connecting adjacent nucleosomes in the chromatin fiber, induced by the variation of the ionic strength of the solution, may result in significant conformational changes of a chromatin in a compact state (3). Current structure-based DNA models (4, 5) do not generate DNA persistence length values that are fully consistent with those measured experimentally in a wide range of ionic concentrations, $c \sim [0.1–100]$ mM (6–8). These and other DNA models differ in resolution in terms of representing both the DNA structure and also the surrounding ionic environment.

The discrete nature of mobile ions and spatial correlations among them may significantly affect structure, dynamics, and the electrostatic atmosphere of many biomolecules, such as DNA and RNA (9–15). Thus, explicit inclusion of the monovalent mobile ions into the CG model of DNA is desirable from the physical standpoint. Another vital aspect of DNA modeling is a choice of the parametrization protocol. Many prior one-bead DNA models were based on use of the phenomenological worm-like chain Hamiltonian (16, 17). In more detailed structure-based models, parameters are often derived by combining the statistical information from the available crystal structures with some specific experimental data (4). However, this parametrization approach, which is focused on matching a small number of integral experimental characteristics (for example, the melting temperature or free energies of denaturation), does not guarantee the fidelity of specific local motions. On the other hand, large-scale polymer chain behavior is generated by propagation in the scale of local motions.

In our recent works on modeling double-stranded DNA, we used a different strategy of parameterizing a DNA potential, on the basis of coarse graining of high-resolution all-atom (AA) force field, in particular, using refined AMBER Parmbsc0 parameters for nucleic acids (18). A similar technique was applied to coarse-grain electrolyte solutions, where recently developed AMBER 10 ionic parameters were used (19). This method, which we called molecular renormalization group coarse graining (MRG-CG), is based on simultaneously matching the moments of all physical "observables" that enter the CG Hamiltonian, between the CG and AA systems. Each of these observables is associated with a specific type of CG effective interaction, such as DNA bond or bending angle potentials. In this approach, the transition from the detailed AA to simplified CG representation is seen as a one-step renormalization, which is further elaborated in *Materials and Methods*.

In the present work we merge our prior models for the double-stranded DNA and NaCl bulk solution, derived systematically from atomistic AMBER10 molecular dynamics (MD) simulations with the MRG-CG technique. Our goal is to accurately capture the local dynamics of both the DNA and mobile ions, including *coupling* between DNA motions and the ionic environment fluctuations. The model of the *two-bead* DNA chain surrounded by Na^+ and Cl^- ions is depicted in Fig. 1. Although it is structurally similar to our previous DNA model (18), where electrostatic interactions were treated implicitly, by using simplified mean-field theory, here the force-field parameters need to be rederived because mobile ions are explicitly present. The procedure for estimating the initial parameters for all types of effective interactions, including structural DNA constraints, interionic interactions, and the interactions between DNA beads and mobile ions, is outlined in *Materials and Methods*. All trial interaction parameters were then subject to optimization with the MRG-CG technique. In contrast to many other CG models of DNA, where generic polymer chain Hamiltonians are used, our model is designed to reproduce complex, anharmonic motions of DNA, which emerge as a result of DNA's chemical fine structural details. Hence, our Hamiltonian represents many important aspects of DNA's stereochemical dynamics, albeit some other chemical details are omitted because of the particular resolution of two beads per base pair.

The quality of the CG model derived with the MRG-CG scheme is determined by the extent of equivalence between AA and CG partition functions, as determined by the matching between various moments of physical observables that enter the CG Hamiltonian (18, 19). In this and prior works, we required similarity of the first moments, which formally corresponds to the optimization problem being addressed on the mean-field

Author contributions: A.S. and G.A.P. designed research; A.S. and G.A.P. performed research; A.S. analyzed data; and A.S. and G.A.P. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: gpapoian@unc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001163107/-/DCSupplemental.

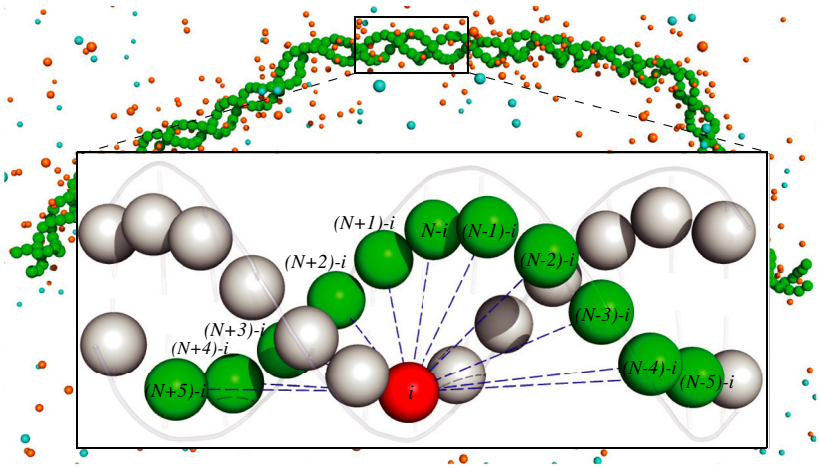

Fig. 1. Our CG model for the DNA chain in the NaCl salt buffer is shown. Each DNA base pair is represented by two beads, each placed in the geometric center of the corresponding atomistic nucleotide. Blue dashed lines indicate the fan interactions, which represent a superposition of stacking and base pairing among two polynucleotides. Labels on beads indicate that fan interactions are between a given bead *i* located on one DNA strand and a number of beads $[(N \pm 0 \ldots 5) - i]$ located on the other strand. *N* is the total number of particles. There are 11 fan interactions imposed on each CG bead.

level. However, as elaborated in the *SI Appendix*, the technique works beyond mean-field for "elementary" order parameters, such as bond lengths and angles, as evidenced by matching even rare fluctuations in these variables (see Fig. 2 below). In addition to evaluating the numerical differences between these averages obtained from AA and CG simulations (see the *SI Appendix*), the quality of the CG model may also be visualized by comparing various structural distribution functions generated for both CG and AA systems. The comparison among some of the DNA structural distributions is illustrated in Fig. 2, demonstrating how iterative adjustment of the Hamiltonian parameters with the MRG-CG technique results in a self-consistent widening of all initial distributions, morphing them into the desired atomistic results. In particular, the final CG distributions accurately follow the asymmetric shapes of the corresponding atomistic curves. Interestingly, structural DNA distributions that were not associated

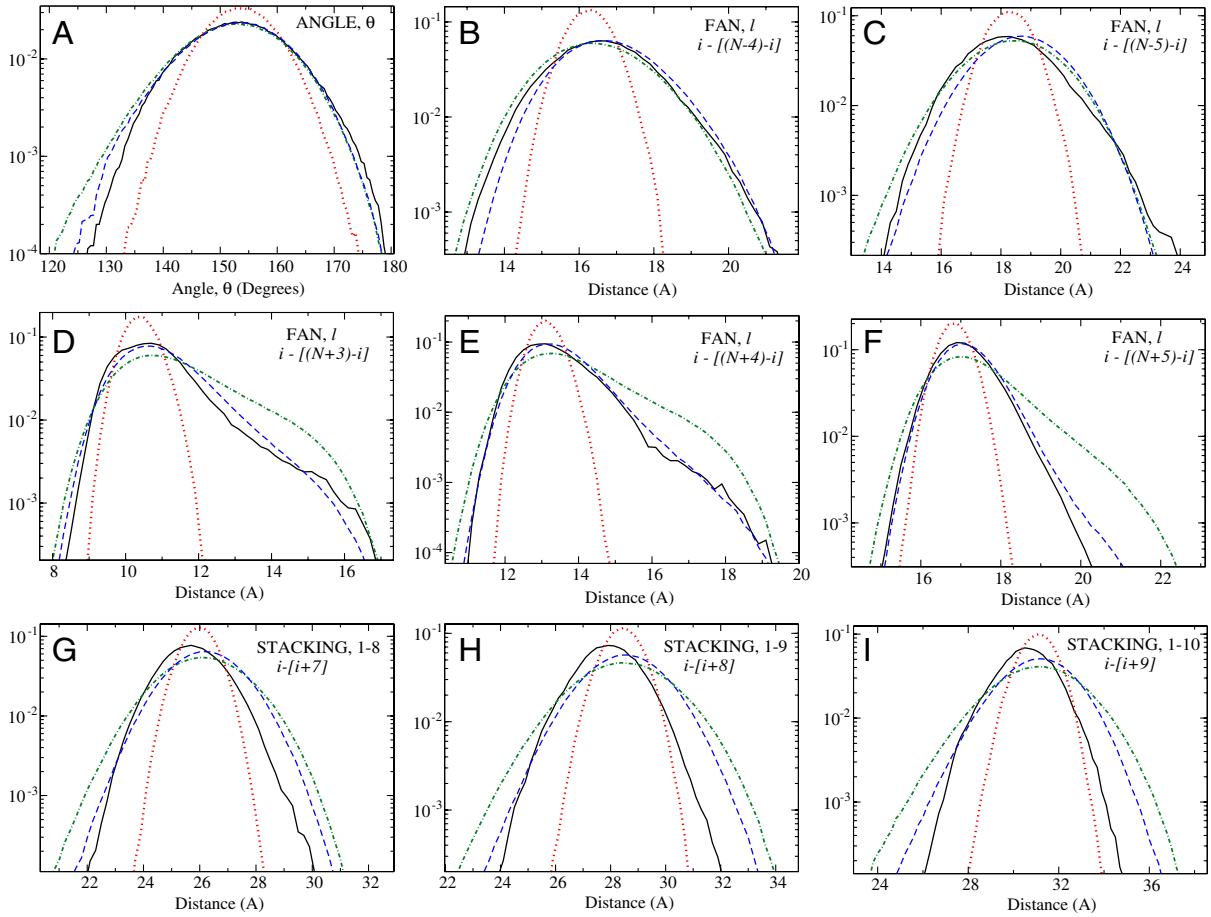

Fig. 2. Semilog plots for some of the DNA structural distributions: (*A*) bending angle; (*B–F*) several of the fan constraints (notations are explained in the caption of Fig. 1); (*G–I*) some of the intrastrand distances (1–8, 1–9, and 1–10 stacking interactions) between DNA particles separated by seven, eight, and nine nucleotides, respectively. Solid black (reference AA), red dotted (initial CG), and blue dashed lines (MRG-CG optimized) structural distributions were obtained from the corresponding simulations. Additionally, wider distributions that have been generated by uniformly rescaling the optimized CG Hamiltonian by 0.7 are shown as green dash-dotted lines (see the text). Note that there are no physical constraints imposed on the DNA chain that correspond to stacking interactions; i.e., they do not enter into the CG force field. However, the last three panels indicate that these (and other) structural distributions are still reasonably well-reproduced.

with any physical observables defining our CG Hamiltonian, thus, not optimized or biased in any way, turned out to be also well-reproduced (Fig. 2 *G–I*). These agreements indicate that our DNA Hamiltonian, elaborated in *Materials and Methods*, is composed of observables that form a nearly complete basis set.

Analogous comparison for the interionic radial distribution functions (RDFs) also revealed a good agreement among Na^+ and Cl^- distributions in AA and CG systems (see Fig. 1 in the *SI Appendix*). The DNA-bead–mobile-ion interactions were derived from a separate series of AA MD simulations of a system composed of a number of *unconnected* DNA backbone "monomers," sodium dimethyl phosphate, and sodium chloride buffers (20). Interestingly, when those trial interaction potentials were used in the current model of DNA chain and NaCl salt, the generated DNA-bead–ion RDFs appeared to be very plausible and close to the atomistic results (see Fig. 3 *A* and *B*). However, subsequent parameter optimization with MRG-CG technique allowed achievement of even higher quality of these distributions.

Two-body distributions shown in Fig. 3 *A* and *B* do not fully represent all aspects of the ionic atmosphere around DNA. Indeed, the tails of distributions shown in Fig. 3 *A* and *B* have contributions from two dissimilar sources: (*i*) the ions located far away from the DNA chain and (*ii*) the ions located in DNA proximity but far away from a particular DNA bead. Because the modulation of DNA dynamics by surrounding mobile ions is important, we explored finer details of ionic atmosphere around DNA by computing three-point correlation functions for ions with respect to two different beads of DNA chain. Such distributions for Na^+ ions are shown in Fig. 3 *C–E*. They demonstrate that spatial correlations of the counterion atmosphere around DNA are well-reproduced in the CG system.

Having found close agreement among a multitude of CG and AA structural distributions, we investigated the dependence of DNA persistence length on the ionic strength of the solution. The central results of this study are presented in Fig. 4. It turned out that DNA persistence lengths generated by optimized Hamiltonian parameters were systematically larger than the corresponding experimental values (see *Inset* in Fig. 4*A*), producing a stiffer DNA at physiological conditions with a persistence length of $l_p \sim 75$ nm. However, the success of the MRG-CG procedure is determined not by directly matching to experimental data but by fiducially reproducing the underlying reference atomistic Hamiltonian. Interestingly, Mazur used an independent technique based on elasticity theory to estimate persistence lengths of various DNA segments with alternating AT and GC sequences, finding values between 75 and 80 nm by using atomistic AMBER simulations (21). Hence, we found not only a close agreement between the CG structural distributions to those measured in the AA system, but also large-scale polymeric properties are accurately reproduced. Furthermore, a uniform rescaling of MD simulation results by a factor of ~0.7 led to nearly exact agreement with the experimental data (see *Inset* in Fig. 4*A*), showing that the experimental trend is also well-reproduced.

However, in practical applications, such as modeling chromatin folding or genome packaging, it is desirable to have a structure-based DNA model that produces persistent lengths in nearly absolute agreement with the experiment. One such possibility is to soften the AMBER Parmbsc0 force field, which might require significant effort, and, subsequently, reoptimize the CG Hamiltonian. It is likely that tenths of k_BT per monomer error in the atomistic force field could propagate in scale, resulting in several tens of nanometers error in persistence length. Up to this step our

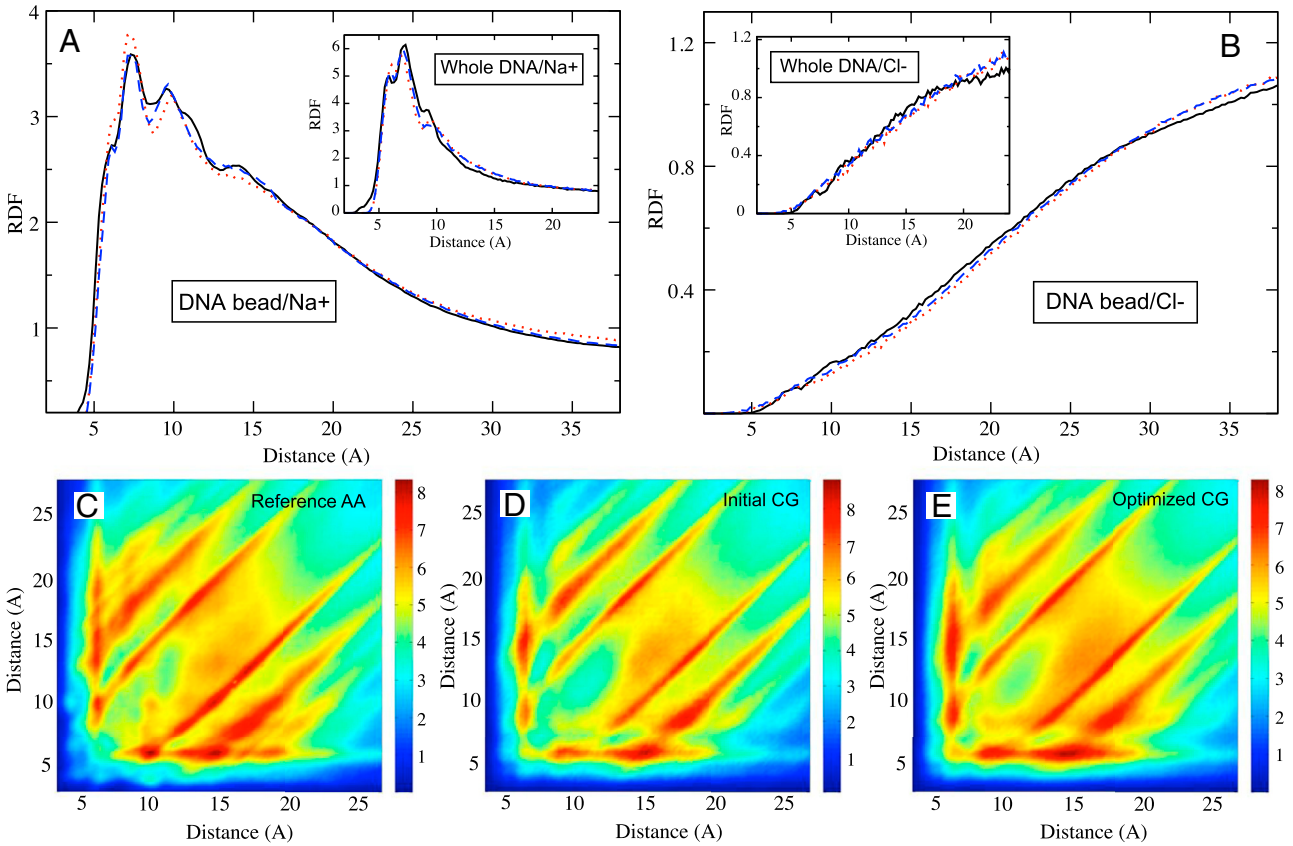

Fig. 3. Distributions of mobile ions around DNA are plotted. The *Upper* panels show one-dimensional RDFs for Na^+ (*A*) and Cl^- (*B*) ions around a single DNA bead (main panels) and the whole DNA chain (*Insets*); line coloring is the same as in Fig. 2. (*C–E*) The *Lower* panels show two-dimensional RDFs for Na^+ ions defined by two distances between the mobile ion and two different beads of DNA; it is seen that agreement with the reference AA distribution is improved after the MRG-CG optimization (see the text).

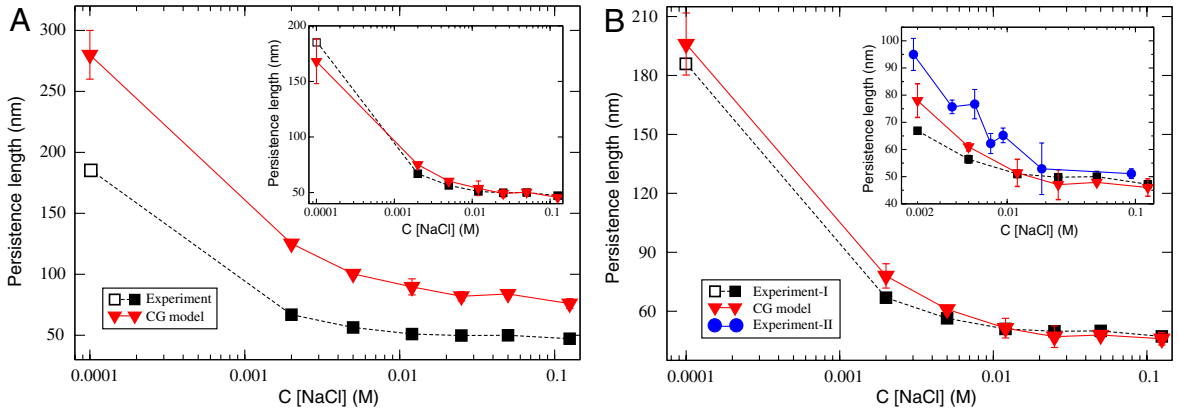

Fig. 4. Dependence of the DNA persistence length on the ionic strength of the solution are shown for various CG models. The main panel (*A*) shows the actual dependence produced by MRG-CG optimization of the AA AMBER Parmbcs0 force-field for the DNA (34) and the latest AMBER10 ionic parameters (35); this plot corresponds to the blue dashed DNA distributions of Fig. 2. The *Inset* of *A* is derived by a uniform rescaling of a red curve in the main panel by a factor of 0.7. (*B*) Response of the DNA persistence length to the change in ionic concentration generated by a CG Hamiltonian, which is uniformly scaled by a factor of 0.7, is shown; this plot corresponds to the green dash-dotted DNA distributions of Fig. 2. Experimental values are taken from ref. 6 (solid squares) and ref. 7 (open square). An additional set of experimental points (blue circles) derived by Baumann et al. (8) for a smaller range of ionic concentrations, $c \sim [2–100]$ mM, is shown in the *Inset*. Persistence lengths were calculated as elaborated in *Materials and Methods*. The error bars represent standard deviations.

CG force field was derived from the atomistic dynamics; however, the previous discussion motivates a physically plausible calibration on the basis of one experimental data point. Hence, we chose a simpler alternative strategy of uniformly rescaling all CG DNA structural parameters by 0.7, leaving the rest of the CG force field unchanged. Fig. 2 shows that such rescaling affected mostly the "fan" DNA distributions, making them slightly wider (i.e., increasing local fluctuations) but preserving the main features of atomistic results, such as asymmetric shapes of the curves. Using this new CG force field, we recomputed DNA persistence lengths for a wide range of NaCl concentrations, $c \sim [0.1–100]$ mM. These results agree well with experimental estimates without any further adjustment, as shown in Fig. 4*B*. It has to be noted that different experimental techniques lead to a noticeable scatter of DNA's persistent length estimates, most likely because of systematic errors and model-dependent interpretation of raw data. Hence, our computational results agree with experiments within the experimental error.

We further tested our CG model on yet another important class of systems, where a delicate balance between elastic and electrostatic interactions determines the chain's structural behavior. Particularly, we have built and simulated a 90-base-pair nanocircular DNA, overtwisted by one helical turn (see Fig. 5*A*), to investigate whether such torsionally stressed DNA chain will undergo a structural transition upon changing the NaCl concentrations in the $c \sim [0.1–500]$ mM range. The importance of studying the topologically constrained DNA chains subject to torsional stress has been recently pointed out in the literature (22, 23). For example, it was experimentally shown that DNA supercoiling may inhibit the process of nucleosomal assembly (22). A recent computational work suggested that supercoiling of torsionally stressed DNA nanocircles depends strongly on the salt concentration, because intrachain electrostatic repulsion opposes elastically induced buckling (23). It should be noted, however, that because of very large system sizes used in explicit solvent atomistic simulations, the systems corresponding to the current work were studied for only 5 ns, which is less than sodium ion equilibration time around DNA, over 50 ns (10, 14). In addition, only a single Na^+ concentration was investigated, corresponding to the DNA neutralization condition, with no additional salt. On the other hand, the DNA model developed in this work is well suited for interrogating the role of explicit mobile ions in modulating the DNA circle's structural behavior as the salt concentration is broadly varied. Sufficiently long equilibration times, around 500 ns for each simulation, allows achievement of equilibration and prediction of the corresponding phase diagram.

Next, we introduce a structural order parameter Q that characterizes DNA circle's supercoiling (see *Materials and Methods* for its definition). The resulting dependence of Q on the ionic strength is shown in Fig. 5*B*. Our CG model predicts that a sharp phase transition occurs from a planar circular DNA to a strongly buckled DNA conformation with the transition midpoint located within a narrow physiological range of salt concentrations, $c \sim [50–200]$ mM (see Fig. 5*B*). Our predictions cannot be directly

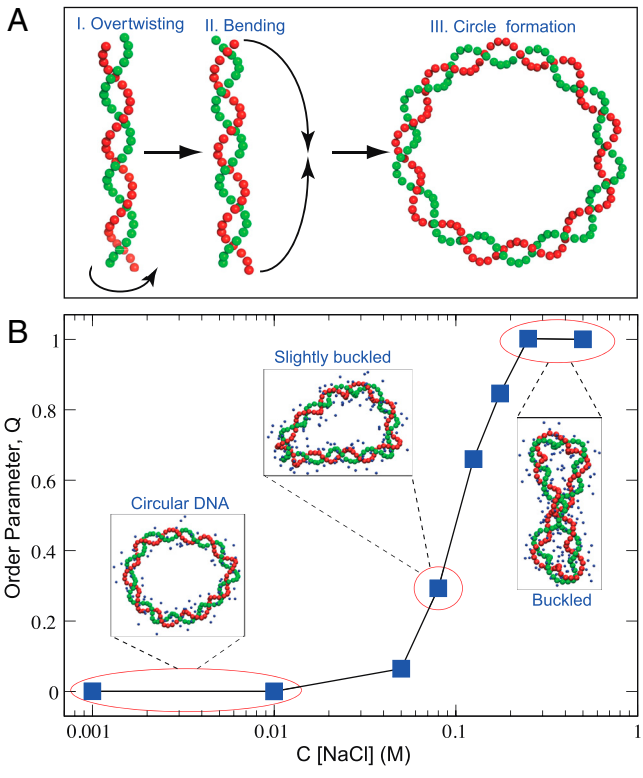

Fig. 5. Phase behavior of the overtwisted 90-base-pair circular DNA chain in explicit NaCl salt. (*A*) Steps of the initial structure preparation are shown. (*B*) The degree of the superhelical twisting at various NaCl concentrations is characterized by the structural order parameter *Q*. See *Materials and Methods* for details. Our CG model predicts a pronounced phase transition at physiological conditions, $c \sim [50–200]$ mM.

compared to the analogous AA simulations of the 90-base-pair DNA nanocircle investigated in ref. 23, because of the reasons outlined above. Nevertheless, Harris et al. found circle bending for $d(GC)_{90}$ and $d(AT)_{90}$, at approximately 140 mM of Na^+ counterion concentration (no Cl^- coions were present). Their results are qualitatively consistent with the behavior of the DNA chain in the very beginning of our simulation at an equivalent salt concentration. However, we find that the DNA nanocircle buckles on longer time scales. On the other hand, it should be noted that the nanocircle becomes nearly planar in our CG simulations already at half of the concentration used in ref. 23, $c \sim 70$ mM (see Fig. 5*B*). In light of significant differences between the corresponding simulation box sizes, ionic conditions, and time scales, these comparisons suggest that our simulation results might be broadly consistent with the fully atomistic models with explicit ions and water.

In summary, we applied the MRG-CG method to develop a two-bead model of DNA with explicit mobile ions. This DNA model, derived by matching molecular correlation functions with atomistic results and uniformly scaled, generates realistic local atomistic motions and, simultaneously, accurately reproduces the large-scale chain dynamics. In particular, quantitative agreement was found with the experimental data on the dependence of DNA persistence length on the solution ionic strength. In addition, we used our model to predict the structural transition of a DNA nanocircle consisting of 90 base pairs as the salt concentration is varied. Overall, the CG DNA model developed in this work may be directly used in various biomolecular simulations, when accessing large length and time scales is desirable, for example, to represent linker DNA in chromatin-folding simulations or to study genome packaging in viruses.

Materials and Methods

Foundation of MRG-CG Method. There are a number of recognized coarse-graining strategies used to simulate many important biological processes (24–29). The optimization scheme used in this work closely follows the RG Monte Carlo method by Swendsen to compute critical exponents in a three-dimensional Ising model (30). This technique was later adapted by Lyubartsev and Laaksonen to simulate several molecular systems with simple pairwise interactions (31). Recently, we generalized the method (18, 19) by extending its applicability to complex systems, such as polymers, which are characterized by more sophisticated interactions (e.g., three-body bending angle interactions). The MRG-CG scheme relies on representing an effective Hamiltonian as a linear combination of N relevant dynamical observables, $\mathcal{H} = \sum_{\alpha=1}^{N} K_\alpha S_\alpha$, whose (various order) correlation functions $\langle S_\alpha \ldots S_\beta \rangle$ need to be reproduced in the CG system. Hence, a "conjugate field" K_α is prescribed to each observable, playing the role of a Hamiltonian force constant, whose numerical value has to be adjusted appropriately to generate the desired system dynamics. Because of Hamiltonian linearity, it is possible to establish a mathematical connection between these conjugate fields and expectation values of dynamical observables in terms of the covariance matrix of *all* observables:

$$\Delta\langle S_\alpha \rangle = -1/(k_B T) \sum_\gamma [\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle] \Delta K_\gamma, \qquad [1]$$

where $\Delta\langle S_\alpha \rangle \equiv \langle S_\alpha \rangle_{CG} - \langle S_\alpha \rangle_{AA}$ is the difference between the expectation values of an observable, S_α, averaged over CG and AA systems and the ΔK_γs are corrections to trial CG Hamiltonian parameters $\{K_\alpha^{(0)}\}$. A similar-in-spirit formalism was used by Matysiak and Clementi to match protein free energy changes upon single point mutations with experimental measurements (32). A set of linear equations (Eq. **1**) is solved at each CG iteration until the convergence is reached for all observables: $\Delta\langle S_\alpha \rangle \approx 0$, $\alpha = 1 \ldots N$. In this way, the process of parameter adjustment explicitly accounts for cross-correlations among various CG degrees of freedom—a key ingredient that is responsible for high fidelity of the local CG dynamics. For example, as we iteratively adjust Hamiltonian parameters for the DNA bending angle potential, we use the information of what impact that adjustment would have on all other CG structural degrees of freedom (for example, bond or stacking dynamical variables).

In a recent work (18), we interpreted the MRG-CG optimization technique in light of the field theory (33). Namely, Hamiltonian linearity allows us to interpret the CG partition function

$$\mathcal{Z}(\{K\}) \propto \sum \exp\left[-1/(k_B T) \sum_{\alpha=1}^{N} K_\alpha S_\alpha\right]$$

as a generating functional, whose differentiation with respect to conjugate fields yields the corresponding auto- and cross-correlation functions of physical observables:

$$\langle S_1 \ldots S_n \rangle \propto \frac{\delta^n \ln \mathcal{Z}}{\delta K_1 \ldots \delta K_n}. \qquad [2]$$

Because the optimization is aimed at matching these various order correlation functions in AA and CG systems, the whole procedure is reminiscent of the central idea of RG theory. Indeed, matching the correlation functions of relevant physical observables ensures a significant equivalence of (restricted) AA and CG partition functions, by matching various order derivatives of the free energy. Additionally, an association with RG theory is strengthened by thinking of the parameter adjustment as a "flow" in space of Hamiltonians, spanned by a set of conjugate fields $\{K_\alpha\}$ coupled to the corresponding observables.

It follows from the last equation that the MRG-CG method can be straightforwardly generalized by demanding to reproduce not only average values but also higher-order correlation functions of observables. Particularly, if AA and CG partition functions generate identical sets of correlation functions of the order n and less, then the value of n may be seen as a quantitative "measure" of similarity between two systems. For example, in the case of $n = 2$ a set of linear equations **1** will be supplemented by $N(N-1)/2$ additional (and still linear) equations aimed at matching second-order correlators $\Delta\langle S_\alpha S_\gamma \rangle \approx 0$.

CG Model. The model for DNA in the current study is structurally analogous to the model developed in our prior related work. The effective Hamiltonian for the double-stranded DNA is the following:

$$\mathcal{H} = \mathcal{U}_{bond} + \mathcal{U}_{ang} + \mathcal{U}_{fan} + \mathcal{U}_{el}, \qquad [3]$$

with the first two terms describing bond and bending angle potential energies (*intrastrand* interactions), and the third and the last terms representing *interstrand* (fan interactions; see Fig. 1) and Coulomb interactions, respectively. Functional forms for individual energetic contributions have been chosen to be quartic polynomials

$$U_{bond,fan} = \sum_{\alpha=2}^{4} K_\alpha (l - l_0)^\alpha, \qquad U_{ang} = \sum_{\alpha=2}^{4} K_\alpha (\theta - \theta_0)^\alpha, \qquad [4]$$

to account for the asymmetric shape of DNA structural fluctuations (see Fig. 2). Here, l_0 and θ_0 are equilibrium interparticle separations for bond and fan interactions and the equilibrium angle for bending angle potential, respectively. These equilibrium interparticle separations, as well as the trial coefficients $\{K_\alpha^{(0)}\}$, were extracted from the fit of the above polynomials to the corresponding potentials of mean forces obtained from AA MD simulations.

Interionic interaction potentials were taken from our prior work on coarse graining a bulk NaCl solution (19). Particularly, this part of CG Hamiltonian has the following functional form:

$$\mathcal{H} = \sum_{i>j} \left[\frac{A}{r_{ij}^{12}} + \sum_{k=1}^{5} B^{(k)} e^{-C^{(k)}[r_{ij} - R^{(k)}]^2} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}\right], \qquad [5]$$

defined by the set of parameters $\{A, B^{(k)}, C^{(k)}\}$ and the positions of Gaussian peaks and minima, $\{R^{(k)}\}$ (five Gaussian functions were introduced to account for short-range hydration effects and to accurately reproduce atomistic behavior of ions).

Finally, functional forms for interaction potentials among beads of DNA and the ions were derived from a separate series of AA MD simulations of a system comprised of *unconnected* DNA backbone monomers (sodium dimethylphosphate) and NaCl salt buffer (20). The latter system was chosen in an attempt to single out a "typical" DNA-bead–ion interaction by suppressing correlation effects caused by DNA connectivity (effects from neighboring DNA beads). These correlation effects were later accounted for by adjusting Hamiltonian parameters with the MRG-CG technique (see Fig. 3 *A* and *B*). The functional form for these types of effective interactions is similar to interionic potentials, however, with softer excluded volume interactions and a lesser number of Gaussian functions to describe hydration effects:

$$\mathcal{H} = \sum_{i>j}\left[\frac{A}{r_{ij}^6} + \sum_{k=1}^{3} B^{(k)} e^{-C^{(k)}[r_{ij}-R^{(k)}]^2} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}\right]. \qquad [6]$$

CG Hamiltonian as a Linear Combination of Dynamical Observables. It follows from the structure of our CG Hamiltonian that the behavior of the system is described by a small number of observables, which may also be seen as structure-based collective order parameters. For example, according to polynomials (Eq. **4**), DNA bond potential energy is described by three collective observables: $S_1^{\text{bond}} = \sum_{\text{all bonds}}(l-l_0)^2$, $S_2^{\text{bond}} = \sum_{\text{all bonds}}(l-l_0)^3$, and $S_3^{\text{bond}} = \sum_{\text{all bonds}}(l-l_0)^4$. Analogously, collective modes characterizing ion–DNA interactions (ionic "shells" around a DNA bead) are $S_\alpha^{\text{Gauss}} = \sum_{\text{all pairs}}[e^{-C_\alpha(r-R_\alpha)^2}]$, $\alpha = 1\ldots3$, whereas the corresponding parameters $\{K_\alpha\}$ are given by the set of constants $\{B^{(k)}\}$; see Eq. **6**. As a result, DNA behavior is associated with $N_{\text{DNA}} = 39$ of structural observables (bond, angle, and fan interactions) coupled to the corresponding conjugate fields $\{K_\alpha\}$, whereas the dynamics of the ionic atmosphere around DNA is described by a total of $N_{\text{ions}} = 2\times4 = 8$ observables (four CG degrees of freedom per interaction of DNA with Na^+ and Cl^-, respectively).

Calculation of the DNA Persistent Length. Persistence length l_p was calculated by using the well-known formula for the semiflexible polymers

$$\langle u_i u_{i+1}\rangle = \exp(-i\cdot\bar{a}/l_p), \qquad [7]$$

defining how a correlation function for tangent vectors decays with the chain length (1). We have chosen the average length of the segment, $\bar{a}$, to be the distance between DNA beads separated by 10 nucleotides.

Preparation and Structural Analysis of the Overtwisted Circular DNA. A 90-base-pair circular DNA chain, overtwisted by one helical turn, was prepared as shown in Fig. 5. The number of base pairs and the degree of the overtwisting were chosen to reproduce one of the DNA nanocircles studied recently with AA MD simulations in ref. 23. First, we imposed a homogeneous stress on the linear and torsionally relaxed DNA chain (Fig. 5*A*), which corresponded to the following excessive twisting angle per base pair:

$$\Delta\phi \equiv \phi - \phi_0 = (360°/N_{\text{bp}})\Delta Lk. \qquad [8]$$

Here, ϕ_0 and ϕ indicate the dihedral angle between consecutive base pairs in the torsionally relaxed DNA chain (in AMBER Parmbcs0 force field $\phi_0 \approx 32°$) and in the overtwisted DNA chain, respectively. This difference can be expressed in terms of the topological quantity known as the linking number, Lk, which is the number of helical repeats within a DNA chain of N_{bp} base pairs. The integer quantity $\Delta Lk \equiv Lk_0 - Lk$ describes the deviation from the torsionally relaxed circular DNA and amounts to the degree of overtwisting (if positive) or undertwisting (if negative). For our model of interest the parameters in Eq. **8** were as follows: $N_{\text{bp}} = 90$ and $\Delta Lk = 1$ (overtwisting by one helical turn), which resulted in $\Delta\phi = 4°$. Subsequent bending of such an overtwisted DNA chain into the closed circle led to the desired initial structure (Fig. 5*A*).

To characterize structural rearrangements of the circular DNA as the concentration of NaCl varied, we introduced the following order parameter:

$$Q = (1/Q_{\max})\sum_{i>j}{}' \exp[-(r_{ij}-d)^2/(2\sigma^2)], \qquad [9]$$

which is reminiscent of the structural order parameters used for description of proteins and spin glasses. The Gaussian functional form is parametrized by the expectation value d for the distance r_{ij} between the DNA beads i and j and the variance σ. The prime on the sum indicates that only pairs of beads i and j separated along the chain by more than a certain number of base pairs are considered. From the structural analysis of the average buckled DNA conformations at high salt concentrations, we estimated these parameters to be $d = 35$ Å and $\sigma = 5$ Å, and the separation between beads along the chain was chosen to be 30 base pairs. This combination of parameters ensured a robust characterization of the structural DNA behavior. For example, it resulted in the zeroth value of Q at small ionic concentrations, as indicated in Fig. 5*B*. Note, finally, that the order parameter was normalized to unity by scaling with the maximum value $Q_{\max}$ attained at the highest salt concentration.

ACKNOWLEDGMENTS. This work was supported by the Beckman Young Investigator Award and Petroleum Research Fund Award 47593-G6. All calculations were carried out by using the University of North Carolina Topsail supercomputer.

1. Doi M, Edwards S (1986) *The Theory of Polymer Dynamics* (Oxford Univ Press, New York).
2. Ma L, Yethiraj A, Chen X, Cui Q (2009) A computational framework for mechanical response of macromolecules: application to the salt concentration dependence of DNA bendability. *Biophys J* 96:3543–3554.
3. Alberts B, et al. (2002) *Molecular Biology of the Cell* (Garland, New York).
4. Knotts TA, Rathore N, Schwartz D, de Pablo JJ (2007) A coarse grain model from DNA. *J Chem Phys* 126:084901.
5. Mielke SP, Gronbech-Jensen N, Benham CJ (2008) Brownian dynamics of double-stranded DNA in periodic systems with discrete salt. *Phys Rev E* 77:031924.
6. Rizzo V, Schellman J (1981) Flow dichroism of t7 DNA as a function of salt concentration. *Biopolymers* 20:2143–2163.
7. Hagerman PJ (1981) Investigation of the flexibility of DNA using transient electric birefringence. *Biopolymers* 20:1503–1535.
8. Baumann CG, Smith SB, Bloomfield VA, Bustamante C (1997) Ionic effects on the elasticity of single DNA molecules. *Proc Natl Acad Sci USA* 94:6185–6190.
9. Shklovskii BI (1999) Wigner crystal model of counterion induced bundle formation of rodlike polyelectrolytes. *Phys Rev Lett* 82:3268–3271.
10. Ponomarev SY, Thayer KM, Beveridge DL (2004) Ion motions in molecular dynamics simulations on DNA. *Proc Natl Acad Sci USA* 101:14771–14775.
11. van der Heyden FHJ, Stein D, Besteman K, Lemay SG, Dekker C (2006) Charge inversion at high ionic strength studied by streaming currents. *Phys Rev Lett* 96:224502.
12. Koculi E, Hyeon C, Thirumalai D, Woodson SA (2007) Charge density of divalent metal cations determines RNA stability. *J Am Chem Soc* 129:2676–2682.
13. Chen Y-G, Weeks JD (2006) Local molecular field theory for effective attractions between like charged objects in systems with strong Coulomb interactions. *Proc Natl Acad Sci USA* 103:7560–7565.
14. Savelyev A, Papoian GA (2006) Electrostatic, steric, and hydration interactions favor Na(+) condensation around DNA compared with K(+). *J Am Chem Soc* 128:14506–14518.
15. Savelyev A, Papoian GA (2007) Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J Am Chem Soc* 129:6060–6061.
16. Langowski J (2006) Polymer chain models of DNA and chromatin. *Eur Phys J E: Soft Matter Biol Phys* 19:241–249.
17. Arya G, Zhang Q, Schlick T (2006) Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys J* 91:133–150.
18. Savelyev A, Papoian GA (2009) Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded DNA. *Biophys J* 96:4044–4052.
19. Savelyev A, Papoian GA (2009) Molecular renormalization group coarse-graining of electrolyte solutions: Application to aqueous NaCl and KCl. *J Phys Chem B* 113:7785–7793.
20. Savelyev A, Papoian GA (2008) Polyionic charge density plays a key role in differential recognition of mobile ions by biopolymers. *J Phys Chem B* 112:9135–9145.
21. Mazur AK (2006) Evaluation of elastic properties of atomistic DNA models. *Biophys J* 91:4507–4518.
22. Gupta P, Zlatanova J, Tomschik M (2009) Nucleosome assembly depends on the torsion in the DNA molecule: A magnetic tweezers study. *Biophys J* 97:3150–3157.
23. Harris SA, Laughton CA, Liverpool TB (2008) Mapping the phase diagram of the writhe of DNA nanocircles using atomistic molecular dynamics simulations. *Nucleic Acids Res* 36:21–29.
24. Whitford PC, et al. (2009) Nonlocal helix formation is key to understanding s-adenosylmethionine-1 riboswitch function. *Biophys J* 96:L7–L9.
25. Thirumalai D, O'Brien EP, Morrison G, Hyeon C (2010) Theoretical perspectives on protein folding. *Annu Rev Biophys* 39:159–183.
26. Hills RD, Brooks CL (2009) Insights from coarse-grained Gō models for protein folding and dynamics. *Int J Mol Sci* 10:889–905.
27. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937–953.
28. Noid WG, et al. (2008) The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J Chem Phys* 128:244114.
29. Nielsen SO, Lopez CF, Srinivas G, Klein M (2004) Coarse grain models and the computer simulation of soft material. *J Phys: Condens Matter* 16:R481–R512.
30. Swendsen RH (1979) Monte Carlo renormalization group. *Phys Rev Lett* 42:859–861.
31. Lyubartsev AP, Laaksonen A (1995) Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys Rev E* 52:3730–3737.
32. Matysiak S, Clementi C (2006) Minimalist protein model as a diagnostic tool for misfolding and aggregation. *J Mol Biol* 363:297–308.
33. Zinn-Justin J (2002) *Quantum Field Theory and Critical Phenomena* (Clarendon, Oxford).
34. Prez A, et al. (2007) Refinement of the amber force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J* 92:3817–3829.
35. Joung IS, Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* 112:9020–9041.