Advanced Review

Recent successes in coarse-grained modeling of DNA

Davit A. Potoyan,[1,2] Alexey Savelyev[4] and Garegin A. Papoian[1–3]*

The growing interest in the DNA-based mesoscale systems of biological and nonbiological nature has encouraged the computational molecular science community to develop coarse-grained (CG) representationsof the DNA that will be simple enough to permit exhaustive simulations in a reasonable amount of time, yet complex enough to capture the essential physics at play. In the recent years, there have been some major developments in the DNA coarse-graining area and several fairly sophisticated models are now available that faithfully reproduce key mechanical and chemical properties of the double- and single-stranded DNA. However, there are still many challenges, which limit the applicability of the present models, and much has to be done yet to develop more reliable schemes which would have a predictive power beyond the target domain of the intrinsic parametrization. A development of robust, controllable, and transferrable CG DNA force fields will provide an invaluable tool for gaining physical insights into the molecular nature of complex DNA-based nanoscale entities such as the chromatin, virus capsids, and DNA nanocomposites. In the present contribution, we provide an overview of the recent developments in the DNA coarse-graining field. Our aim is to review the existing CG models of the double-stranded DNA, where a small selection of models, which we believe provide avenues for promising future development, are discussed in some detail.

How to cite this article:
WIREs Comput Mol Sci 2013, 3: 69–83 doi: 10.1002/wcms.1114

INTRODUCTION

Although DNA may seem as a relatively simple polymer, fundamental aspects of DNA physics, including the origin of DNA's elasticity,[1–6] dominant modes in the conformational dynamics,[7,8] and the DNA topology,[9–11] are not well understood. Furthermore, even less is understood about the structure, dynamics, and functioning mechanisms of DNA chains *in vivo*, where many interesting problems have recently been identified.[12,13] In particular, the total contour length of DNA in a micrometer-sized nucleus of a eukaryotic cell is of the order of 1 m.[14] The physical basis and the biological implications of the nearly million-fold compression of the highly charged and relatively stiff macromolecule are currently actively investigated. *In vivo*, DNA is packed into chromatin fibers in complexation with positively charged proteins, called histones. Despite structural condensation and packing, the chromatin organization allows retrieval of the desired portions of DNA in a timely manner for processing and manipulation.[11,15] Developing a more complete physical picture of DNA in isolation as well when complexed with various proteins is needed for making further progress in uncovering the principles of genetic regulation in biology.

Recently, a great interest has also emerged in the nonbiological applications of the DNA, namely, the design of the DNA-based nanoscale materials.[16] Through the pioneering works of Seeman and coworkers on DNA nanoassembly, the programmable design of the DNA nanomaterials resulted

*Correspondence to: gpapoian@umd.edu

[1]Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA

[2]Chemical Physics Program, University of Maryland, College Park, MD, USA

[3]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, USA

[4]Institute of Applied Physics, National Academy of Sciences, Sumy, Ukraine

DOI: 10.1002/wcms.1114

in innovative structures, such as the DNA origami,[17–19] where arbitrarily long strands of the DNA can be sealed together to create two- and three-dimensional objects. In these type of applications, the DNA is simply used as a template material for constructing intricately shaped nano-objects. The knowledge of the mechanics that operates at these lengthscales will undoubtedly be of much help in achieving a better control over functional properties and in guiding the knowledge-driven development of the novel materials. The typical sizes of these objects pose a challenge for developing computational models that will be simple enough to permit fast calculation on a medium-sized workstation, yet complex enough to capture the essential physics at play. For more details and an overview of computational efforts in the field of the DNA nanocomposites, we refer the interested reader to a recent review by de Pablo.[20]

In the view of the complexity of problems related to the DNA physics, computer modeling has always played an important role in bridging the gap between the bulk experiments and molecular-level understanding. The currently available all atom (AA) force fields have significantly evolved since their inception in the 1980s[21,22] and are now able to reproduce a diverse set of subtle structural fluctuations that take place from pico- to microsecond timescales.[23–27]

Development of more refined atomistic force fields for the DNA is an area of active research.[25,28,29] For example, recently, the collaborative effort of several research groups termed as Anacosia B-DNA consortium (ABC) carried out a large set of extensive AA simulations on short pieces of DNA strands,[27] compiling a library of all possible DNA nearest neighbor fragments. It was found that the nearest neighbor effects are quite substantial in defining the local chain flexibility, suggesting that they need to be taken into account to reproduce sequence-dependent conformational properties.[27]

It should also be noted that the currently available parameterization for the AA DNA is biased by the narrow subset of known crystallographic structures,[25] leaving behind many less common DNA isoforms such as the protein–DNA complexes, the highly kinked forms of the DNA, tetra-loops, and G-quadruplexes. Furthermore, the longest AA simulations on the DNA reported to date has been a few microseconds long study of Drew–Dickerson dodeckamer,[24,25] whereas the nucleosome, containing approximately 150 base pairs of DNA, was simulated for only a hundreds of nanoseconds.[30] However, for most interesting biological and material science problems, the need to access long timescales and large lengthscales rules out direct use of atomistic simulations. For instance, to study many interesting rare dynamical events like bubble formation, contour fluctuations, breathing motions and large conformational transitions, one needs to run long-time simulations to obtain reliable statistics. This can be computationally demanding for the AA models and moreover some of the microscopic fluctuations might be irrelevant, as they average out during long timescales. Hence, coarse graining can be viewed as a clever way of separating and disregarding irrelevant atomistic noise from the subset of degrees of freedom which facilitate the sampling of the more interesting long timescale behavior. Thus, it is expected that the progress in the fields like chromatin folding and dynamics, viral genome self-assembly or DNA nanotechnology will likely be mainly driven by simulations based on coarse-grained (CG) approaches.

For a long time, development of CG molecular models was more an art than science. Recently, more rigorous approaches based on statistical mechanics have allowed to think about coarse graining in a more systematic way, as either a formal integration of faster degrees of freedom[31,32] or as a one-step renormalization.[33] These conceptual advances, however, do not yet provide a foolproof recipe for constructing comprehensive CG force fields for such a complex molecule as DNA. Hence, the art of coarse graining lives on, albeit aided by more systematic theoretical approaches. While for smaller molecules, such as lipids, a number of CG force fields are available,[34] the field of DNA coarse graining is unfortunately among the less mature ones, as currently there are only a handful of realistic CG models available.[2,35–40] All of these models are incomplete in some important ways, addressing either only certain facets of DNA physics or being rather low resolution. One of the key difficulties specific to the DNA coarse graining is the correct handling of electrostatics forces, which are rather complex near highly charged DNA, and where many-body effects of the ionic environment need to be properly taken into account.[2] Another intriguing possibility for accurately capturing the long timescale dynamical properties is the hybrid CG/AA simulations which treat the critical chemical fragments with a high-resolution atomic details, leaving the rest to a lower resolution CG treatment.[41]

On a more technical note, code sharing may greatly facilitate the rapid development of the CG methodologies. For instance, the impressive evolution of atomistic force fields has been largely driven by the community of users (see, for instance, Refs 42 and 43) who over many years reported the critical performance issues, which triggered the need for further optimizations.[28] One popular venue for

sharing of CG force fields and codes is the large-scale atomic/molecular massively parallel simulator (LAMMPS) open source MD engine.[44]

The present review on DNA coarse graining reflects the personal interests and biases of the authors, and covers only several models which we thought are representative of the state-of-art in the field, and demonstrate strengths and weaknesses of the current crop of CG force fields. We aim to give a broader sense of the field, meanwhile emphasizing the challenges which may lay ahead. The review is organized in the following way: first, we introduce the basic concepts and the technical language; then we discuss in detail a small selection of CG models of DNA, mention a number of additional CG models, and conclude with the a brief discussion of some nascent efforts in modeling chromatin using CG approaches.

BASIC PHYSICS OF DNA AS A SEMIFLEXIBLE CHAIN

Before discussing the variety of sophisticated molecular schemes that are nowadays available for simulating the DNA, it is worthwhile to go over some of the classical theoretical models that establish a conceptual framework for discussing structural and statistical aspects of the DNA's conformational behavior. These analytical, 'toy-model' descriptions still play an important role in guiding the development of more refined models and helping to rationalize a large body of experimental data. Therefore, we would like to provide the reader with a brief overview of the key theoretical concepts that we will use throughout the text.

The most important analytical model for understanding the mechanical properties of the DNA is the celebrated worm-like chain (WLC) model, also known as the Kratky–Porod model.[45,46] The WLC is a mathematical description of a thin inextensible elastic filament with a single parameter to account for the thermal bending susceptibility, called the persistence length. In a nutshell, the WLC is a small-angle limit of an ideal freely rotating chain, where the latter has fixed bond lengths and angles with all the other degrees of freedom being random. Hence, the orientational memory of the chain comes solely from correlations between adjoining segments. Using a parametric description of the chain given by the contour position vector $\vec{\mathbf{r}}(s)$ one can show[47,48] that the correlation of the tangent vectors $(\vec{\mathbf{t}}(s) = \mathrm{d}\vec{\mathbf{r}}(s)/\mathrm{d}s)$ along the chain decays with a 'rate' set by the persistence length (see also Figure 1):

$$\langle \vec{\mathbf{t}}(s) \cdot \vec{\mathbf{t}}(0) \rangle = \mathrm{e}^{-s/l_p}. \tag{1}$$

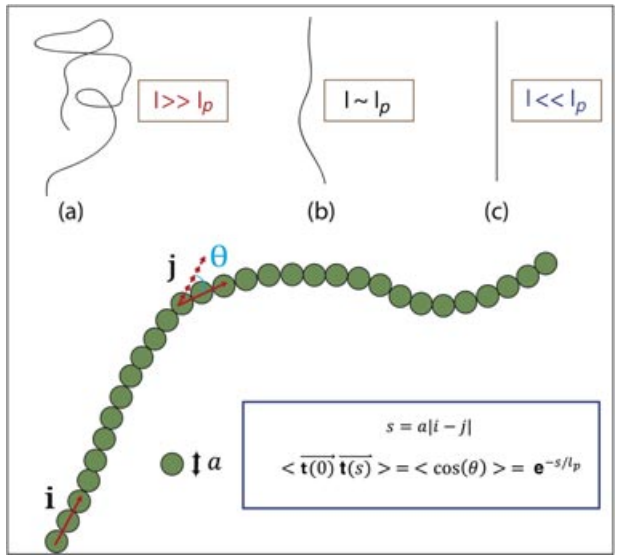

FIGURE 1 | A discrete model of a hypothetical polymer drawn to illustrate the worm-like chain (WLC) model. The *i* and *j* denote the indices for beads and *a* is the diameter of the bead. The upper part of the figure shows three regimes predicted by the WLC model. (a) The coil regime where the chain is governed by the conformational entropy and assumes random-like configurations; (b) semiflexible regime, where chain has occasional small kinks along the contour but is overall aligned in one direction; (c) rod limit, where the chain can be effectively regarded as a straight line.

The Hamiltonian of the WLC consists of one term, accounting for the only degree freedom in the system at each point along the chain, namely, the bending fluctuations:

$$\frac{\mathcal{H}}{k_{\mathrm{B}}T} = \frac{\kappa_b}{2} \int_0^L \mathrm{d}\vec{\mathbf{s}} \left(\frac{\mathrm{d}^2\vec{\mathbf{r}}(s)}{\mathrm{d}s^2} \right)^2. \tag{2}$$

The bending constant κ_b quantifies the susceptibility of the chain to bending from the straight configuration and thus contains the same information as the persistence length. In fact, one can show that in the WLC model the two are related by the simple expression: $l_p = \kappa_b / k_{\mathrm{B}}T$.[47,48] Despite the crude nature of the model, which ignores the molecular features of a chain, the WLC Hamiltonian provides a surprisingly good description for relatively long strands of the double-stranded DNA, whose mode of flexibility is dominated by contour fluctuations (see Figure 1), rather than dihedral angle rotations as is the case with proteins. In particular, the WLC model provides highly accurate description of extensions of single ds-DNA molecules due to forces up to 10 pN.[49] Recently, there has been much controversy about the validity of WLC model for describing the bending

statistics of the short chains of DNA on the subpersistent lengthscales.[4,50] However, recent experimental results by Alivisatos and coworkers[51] and also the Brownian dynamics simulations performed by Mazur[52] seem to support the applicability of WLC even for the subpersistence length DNA. The currently accepted value for the persistence length of dsDNA in the physiological salt conditions is estimated to be around 45–50 nm, which comprises approximately 150 base pairs.[4]

The second potentially relevant analytical model, from the theory of polyelectrolytes, is the Odijk–Skolnick–Fixman (OSF) theory which was proposed as a generalization of the WLC model for stiff polyelectrolytes.[53,54] The OSF model ignores reorganization of charges due to bending and thermal fluctuations and treats the chain as a uniformly charged continuous filament, whose different segments interact via screened electrostatic potentials according to the Debye–Hückel (DH) theory (DH). A key prediction of the theory is the electrostatic component (l_p^{el}) of the net persistent length:[53,54]

$$l_p = l_p^0 + l_p^{\mathrm{el}}. \tag{3}$$

The expression for the electrostatic persistent length within the framework of the OSF is the following:

$$l_p^{\mathrm{el}} = \frac{\sigma_{\mathrm{eff}}^2 l_{\mathrm{B}}}{4\kappa^2}, \tag{4}$$

where σ_{eff} is the linear charge density of the chain, l_{B} is the Bjerrum length, and κ^{-1} is the Debye length of the solution.[55] Although, the OSF model seems to provide a reasonable fit to experimental data on the salt dependence of DNA's persistent length,[56] its estimate of l_p^{el} is of the order of 5 nm at physiological conditions, which seems to be several fold underestimated based on recent atomistic and CG computer simulations[1] as well as on conclusions from various experiments that chemically neutralize specific DNA charges.[4] Based on these recent results, it seems that some of the assumptions underlying the OSF model may be flawed when specifically applied to DNA, and more complete analytical models need to be developed to explain the origin of DNA's flexibility. Furthermore, there has been a recent discussion in the literature, whether l_p^{el} depends quadratically on κ^{-1}, as the OSF theory predicts [see Eq. (4)], or whether the dependence is in fact linear.[57] In particular, recent CG MD simulations demonstrated a continued decrease of the DNA persistence length by approximately 25% when concentrations of monovalent mobile ions increased 10-fold in a range of approximately $[0.1 - 1]$ M, thus, also questioning the OSF theory predictions given by Eq. (4).[58]

RECENT COARSE-GRAINED MODELS OF THE DOUBLE-STRANDED DNA

Although all coarse-graining efforts have the same aim, namely, to allow computationally efficient, yet accurate description of biomolecular structure and dynamics, many philosophies exist on how to do this in practice. In particular, it may be useful to classify the way the microscopic details are coarsened, which can be based either on top-down or bottom-up approaches. In the top-down approach, the force field is chosen based on either a structural intuition or trial and error simulations. Afterward, the undetermined parameters are fitted using the available experimental data. In the bottom-up approach, the CG Hamiltonian is chosen and further parameterized using the AA simulations as a reference, based on matching CG and AA partition functions. At least at the present time, both approaches to coarse graining are rather defensible, where the top-down approach relies on the reliability of the experimental data, but which are usually rather scarce and of low structural resolution, whereas the bottom-up approach's both strength and weakness is the quality of the underlying AA model. In the following few sections, we discuss coarse-graining strategies for DNA in the context of three specific examples, concluding by a brief overview of some additional CG models of DNA and nascent attempts to model chromatin assembly and dynamics.

Three Spherical Beads Representing a Single Base

De Pablo and coworkers[35,59] have developed a CG representation of a DNA chain, which provides a mesoscale-level description of DNA's conformational and thermal properties. The main focus of the model is on reproducing the thermodynamics of melting, bubble formation, hybridization, and salt dependence of the persistence length. Inspired by the off lattice Go-like potentials from the protein folding studies,[60,61] the model contains a special biasing potential that penalizes the large-scale structural fluctuations from the crystallographic B-form of the DNA. The parameters in the model are tuned to reproduce the experimental curves of melting for a variety of sequences. As a result, reversible denaturation processes and complex bubble formation dynamics may be studied (see Figure 2b).

In the CG scheme, each nucleotide is mapped into three beads corresponding to phosphate, sugar

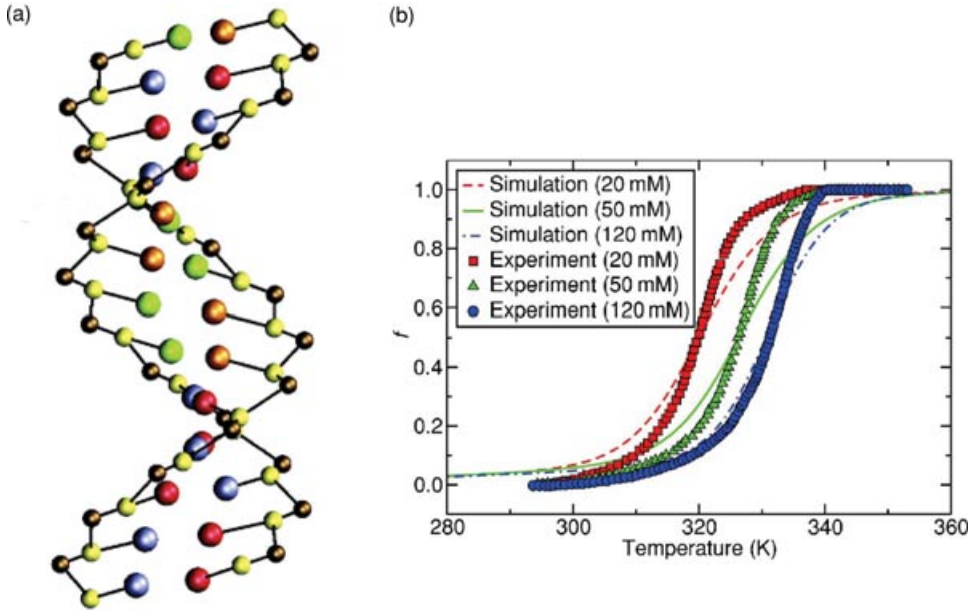

FIGURE 2 | (a) Schematic representation of the mesoscale model of DNA (b) Comparison of thermal melting curves from the coarse-grained simulations with the experiments. (Reprinted with permission from Ref 59. Copyright 2007 AIP.)

and base functional groups (see Figure 2a). The corresponding Hamiltonian consists of the following terms,

$$V_{\text{total}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihed}} + V_{\text{stack}} + V_{\text{bp}} + V_{\text{excl}} + V_{\text{qq}} + V_{\text{solv}}, \quad (5)$$

where the individual contributions are

$$V_{\text{bond}}(r_i) = k_{b1}\left(r_i - r_i^0\right)^2 + k_{b2}\left(r_i - r_i^0\right)^4, \quad (6)$$

$$V_{\text{angle}}(\theta_i) = k_a\left(\theta_i - \theta_i^0\right)^2, \quad (7)$$

$$V_{\text{dihed}}(\phi_i) = k_\phi\left[1 - \cos\left(\phi_i - \phi_i^0\right)\right], \quad (8)$$

$$V_{\text{stack}}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \quad (9)$$

$$V_{\text{bp}}(r_{ij}) = 4\epsilon_i^{\text{bp}}\left[5\left(\frac{\sigma_i^{\text{bp}}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_i^{\text{bp}}}{r_{ij}}\right)^{6}\right], \quad (10)$$

$$V_{\text{excl}}(r_{ij}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6}\right] + \epsilon, & \text{if} r_{ij} < d_{\text{cut}}, \\ 0, & \text{if} r_{ij} \geq d_{\text{cut}}, \end{cases} \quad (11)$$

$$V_{\text{qq}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon(T, C)r_{ij}} e^{-r_{ij}/\kappa_D}, \quad (12)$$

$$V_{\text{solv}}(r_{ij}) = \epsilon_s[1 - e^{-\alpha(r_{ij} - r_s)}]^2 - \epsilon_s. \quad (13)$$

The first three terms represent expressions for treating dynamics of bonds, angles, and dihedral angles, which are commonly used in many molecular mechanics force fields. All the equilibrium values were taken from the crystallographic structures of the B-DNA. The V_{stack} term is responsible for maintaining the hydrophobically driven intrastrand base stacking interactions and backbone rigidity. The cutoff for V_{stack} is set to 9 Å which extends the range of nonbonded interactions up to the next neighbor bases . The V_{excl} term accounts for excluded volume interactions. The electrostatic interactions are included in the V_{qq} term, based on a simplified DH approximation for the ionic environment, with a temperature and concentration dependent dielectric constant. The temperature dependence of the dielectric constant ϵ is accounted for by using some phenomenological expressions for the saline water.[62] Sites that are part of a bond are excluded from all nonbonded interactions, which eliminates the potential structural instabilities. The V_{stack}, V_{bp}, and V_{excl} potentials are mutually exclusive, namely, a pair of sites can have only one of these terms, which is done to eliminate unphysical geometries. The V_{solvent} term is introduced to mimic the many body effects due to hydration water layers, which plays crucial role in the thermodynamics of the melting transition. A set of extensive

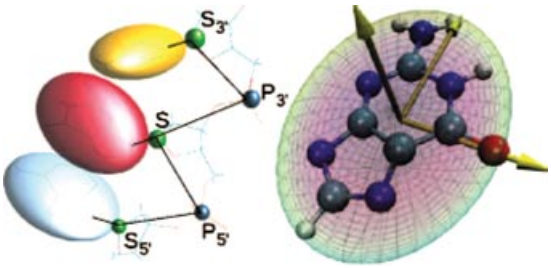

FIGURE 3 | (a) Schematic representation of three nucleotides, with ellipsoidal beads corresponding to bases and the beads labeled as S and P to sugars and phosphate groups, respectively. (b) The all atom framework of nucleic base, showing the principal axes of the ellipsoid, which uniquely determine the orientation of the base. (Reprinted with permission from Ref 37. Copyright 2010 AIP.)

simulations on short DNA chains were carried out to find the optimal parameters for the solvation potential to reproduce the thermodynamics of melting. Using replica exchange molecular dynamics simulations conducted under the low salt concentrations (~0.069 M), authors were able to reproduce a number of thermal properties such as the melting curve and the temperature dependence of heat capacity. Parameter fitting for the thermal melting behavior was carried out at one salt concentration, which then reproduced the salt dependence of persistence length. Persistence length of the single-stranded DNA comes out significantly lower than that for the dsDNA which is qualitatively consistent with the experiments. In sum, the CG model lives up to its claims by reproducing melting thermodynamics and showing sequence-dependent bubbling phenomena, which opens a way for studying the equilibrium conformational behavior of long DNA strands. However, the usage of Go-like potentials somewhat restricts the application of model to systems where the structure of DNA may deviate from canonical B form. Also, the validity of the DH treatment of electrostatics is rather questionable given very large surface charge density of DNA, as evident when comparing fully atomistic and continuous electrostatics treatments of DNA and nucleosomes.[63,30] Moreover, it is expected that the quality of the DH approximation will further deteriorate when DNA needs to be studied at high salt concentrations.

Accounting for Anisotropic Shapes of Coarse-Grained Functional Groups

Plotkin and coworkers[37] have developed a CG DNA model that stands out from others by the usage of nonisotropic potentials which are more accurate in handling the geometry of the nucleotide bases (see Figure 3). Their CG Hamiltonian contains purely physicochemical interactions, which are mainly parameterized by relying on corresponding AA MD simulations, with an aim to reproduce the key mechanical properties of the DNA. Each nucleotide is represented by three beads: two spherical ones representing the sugar and phosphate groups and one rigid ellipsoid that captures the 'pancake' shape of nucleotide bases. Despite the fact that anisotropic potentials are typically more computationally time consuming and require extra optimization parameters, there are a number of advantages for using them. For one thing, they allow straightforward calibration of stacking interactions and by varying their strength one can gain insight into how they influence various structural and mechanical properties of DNA, such as the helical twisting of double-stranded DNA.[37] In addition, when using only isotropic potentials, it is difficult to account for structural information about base tilting, twisting and proper stacking, where more explicit control over these interactions should allow for finer regulation of DNA chain geometry and local dynamics. Also, as it is known that hydrogen bonds are directional, the nonisotropic potentials in principle do a much better job of representing that aspect of the chemical reality.

The configurational part of the Hamiltonian for the model of Plotkin and coworkers takes the following form:

$$\begin{aligned} V_{\text{total}} = {} & V_{\text{bond}}(r) + V_{\text{angle}}(\theta) + V_{\text{dihed}}(\phi) \\ & + V_{\text{RE}^2}(\mathbf{B_1}, \mathbf{B_2}) + V_{\text{RE}^2}(\mathbf{B_1}, \text{res}_2) \\ & + V_{LJ}(r_{\text{ss}}, r_{\text{sp}}) + V_{\text{C}}(r_{\text{pp}}). \end{aligned} \tag{14}$$

The fist three expressions describe the bond, angle, and dihedral angle fluctuations; however, no explicit forms are assumed for any of them. Instead the functional forms along with parameters are determined by either fitting the AA potentials of mean force (PMF) plots (e.g., for angles, $V(\theta) = -k_{\text{B}}T\ln p(\theta)$) to some either custom or common functional forms. This fitting procedure results in widely varying values for spring constants for the different bonds, angles, and dihedrals. The use of one-dimensional PMFs to fit individual terms in the Hamiltonian neglects cross-correlations between degrees of freedom, which under certain conditions can be a reasonable approximation but also sometimes can lead to significant artifacts.[64,33] The potential for bonds was all fit with a simple quadratic function:

$$V_{\text{bond}}(r) = \frac{1}{2}k_b(r - r^0)^2. \tag{15}$$

Several different expressions were used to describe angular interactions, depending on their specific type [Eqs (16) and (17)], to achieve an optimal fit to one-dimensional PMFs:

$$V_{\text{angle}}(\theta) = \frac{1}{2}k_\theta(\theta - \theta^0)^2 \tag{16}$$

$$V'_{\text{angle}}(\theta) = -k_{\text{B}}T\ln\left(e^{-k_1\frac{(\theta-\theta_1)^2}{2k_{\text{B}}T}} + Ae^{-k_2\frac{(\theta-\theta_2)^2}{2k_{\text{B}}T}}\right) \tag{17}$$

The Gay–Berne[65] style potential, proposed by Babadi et al.[66] was used to describe the nonisotropic component of potentials, which are denoted with RE^2 in the argument of the respective potentials in Eq. (14). These potentials have a rather complicated form, therefore we refer the interested reader to an appendix of the original paper.[37] Interaction between two ellipsoids is fully defined by the shape tensors of the interacting ellipsoids given by $S^{(i)} = \text{diag}(\sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)})$ and their energy well depths in the respective directions: $E^i = \text{diag}(\epsilon_x^{(i)}, \epsilon_y^{(i)}, \epsilon_z^{(i)})$. Similar to other nonbonded potential there is also a distance and energy scales and a cutoff distance to be specified for every unique pair. The base sugar interactions are accounted via $V_{\text{RE}^2}(\mathbf{B_1}, \text{res}_2)$ potential which is simply the limiting case of the base–base potential $V_{\text{RE}^2}(\mathbf{B_1}, \mathbf{B_2})$ when one ellipsoid is replaced by a sphere, e.g., $(\sigma_x = \sigma_y = \sigma_z)$. Sugar–sugar and sugar–phosphate interactions are modeled by the LJ potential V_{LJ} which only acts between residues which are more than two neighbors apart $(i \geq j + 3)$. Hydrogen bonding between complementary bases is modeled by a separate HB potential of the form:

$$V_{\text{HB}}(r_{ij}) = 4\epsilon_{\text{HB}}\left[5\left(\frac{\sigma_N}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_N}{r_{ij}}\right)^{6}\right] \times (\cos^4(3\theta_i)\cos^4(3\phi_i) + \cos^4(3\theta_j)\cos^4(\phi_j)). \tag{18}$$

The angular part of the hydrogen bonding potential is meant to penalize the orientations, which deviate from the ideal planar geometry. The cosines in the last expression are raised to the fourth power to magnify the energetic cost. Ions are treated in an implicit manner, hence the electrostatic potential is the same as in the de Pablo's model, except that the dielectric constant is temperature independent. In the model, each base (consisting of the base, sugar, and phosphate moieties) has a total of nine degrees of freedom (coming from three constrained rigid bodies), in contrast to the order of hundred degrees of freedom for the fully atomistic model. This, in turn, results in a significant reduction of computer time for simulating the CG system. The performance of the model was tested against several nontrivial mechanical effects, such as stacking patterns, salt dependence of the persistent length for the ds and ss DNA, and thermally driven collapse–expansion transition of the ssDNA. In particular, the trend of DNA rigidifying at low ionic strength was well captured, although the absolute value of DNA's persistent length turned out to be approximately twice as small compared with the experimental estimates.[37] The results demonstrated the robustness of the model by qualitatively capturing the trends for several key mechanical properties. In addition, the authors provided systematic study of coupling between several potentials, suggesting avenues for further improvement of the model. The model also shows stable major and minor grooves and has a correct bias toward the right helical B-DNA conformation, which is something that has not been explicitly accounted for neither by designing potentials nor by placing geometric constrains in prior physics-based CG models of DNA, where no Go-like potentials are used.

MRG-CG: A Bottom-Up, Statistical Mechanics Approach to Coarse-Graining DNA

Savelyev and Papoian[2] systematically derived a two-bead per base pair model of double-stranded DNA with explicit mobile ions by matching CG and detailed atomistic partition functions. The coarse-graining approach used is called molecular renormalization coarse-graining (MRG-CG),[33,67] which was introduced by Savelyev and Papoian as a generalization of pairwise coarse-graining technique of Swendsen[68] and Laaksonen–Lyubartsev[69] to many-body molecular interactions. This, in turn, provides the needed groundwork for developing CG models for complex macromolecules, such as DNA. The uniqueness of the MRG-CG approach consists of thinking of a CG Hamiltonian as being spanned by a compact basis set of functions, in the same spirit as basis set expansions in quantum chemistry, and then deriving the corresponding parameters by using the MRG-CG procedure.[33,67] The technique takes into account the complicated correlations between various degrees of freedom, hence avoiding the problem of simply relying on one-dimensional potentials of mean force, as discussed earlier. The resulting CG model of a double-stranded DNA chain generated a fluctuation spectrum of complex local motions that was shown to be highly similar to what was observed in atomistic simulations with explicit solvent, ions, and detailed representation

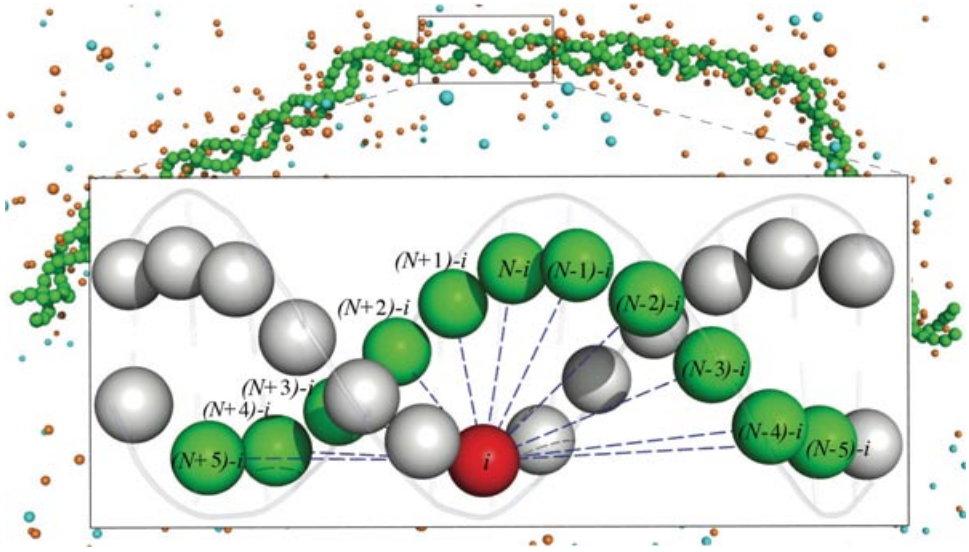

FIGURE 4 | A recently developed chemically accurate coarse-grained model of the double-stranded DNA with explicit mobile ions by Savelyev and Papoian.[2] Each DNA base-pair is represented by two beads, each placed in the geometric center of the corresponding atomistic nucleotide. Blue dashed lines indicate effective interactions which represent a superposition of stacking and base pairing among two polynucleotides.

of DNA.[2] Furthermore, propagation of these local motions also allows for a correct description of large-scale DNA behavior.

Each DNA base pair in a model is represented by two beads of the same type, where each bead is placed in the geometric center of the corresponding base pair nucleotide (see Figure 4). Although currently the model is averaged over DNA sequence, it should be straightforward to develop sequence-specific extensions. Overall, because of approximately 30-fold reduction of DNA degrees of freedom and absence of explicit solvent, a dramatic saving of computer time results, allowing to simulate hundreds of DNA base pairs on the microsecond timescale using relatively modest computational resources. Despite the simplicity of the structural representation, the major and minor groove structural patterns are preserved. The effective Hamiltonian of the Savelyev and Papoian double-stranded DNA model has the following form:

$$\mathcal{H} = \mathcal{U}_{\text{bond}} + \mathcal{U}_{\text{ang}} + \mathcal{U}_{\text{fan}} + \mathcal{U}_{\text{el}}, \tag{19}$$

with the first two terms describe bond and bending angle potential energies (*intra*-strand interactions), and the third and the last terms represent *inter*-strand ('fan' interactions represented by blue dashed lines in Figure 4) and electrostatic interactions, respectively. Functional forms for individual energetic contributions have been chosen to be quartic polynomials,

$$U_{\substack{\text{bond,}\\ \text{fan}}} = \sum_{\alpha=2}^{4} K_\alpha (l - l_0)^\alpha,$$

$$U_{\text{ang}} = \sum_{\alpha=2}^{4} K_\alpha (\theta - \theta_0)^\alpha, \tag{20}$$

to account for asymmetric shape of atomistic DNA structural fluctuations, hence going well beyond the WLC and taking into account the significant anharmonicities of DNA chain motions. The parameters l_0 and θ_0 are equilibrium interparticle separations for bond and fan interactions, and the equilibrium angle for bending angle potential, respectively. As customary, equilibrium values l_0 and θ_0, as well as the trial set of coefficients $\{K_\alpha^{(0)}\}$ were obtained by fitting these polynomials to the corresponding PMF, extracted from AA MD simulations. However, this is only the first step, followed subsequently by optimization of all trial interaction parameters using the iterative scheme of MRG-CG.

In the first version of the model,[33] only a number of key polymeric interactions, such as bond, bending angle, and fan potentials, were optimized with MRG-CG technique to reproduce the local motions of the underlying atomistic DNA molecule with high fidelity, while electrostatic interactions were treated implicitly, by using simplified mean-field theory, as discussed above for the CG DNA models developed by de Pablo and Plotkin groups. However, the discrete nature of mobile ions and spatial correlations among

them may significantly affect structure, dynamics, and the electrostatic atmosphere of DNA,[30,63,70] and explicit inclusion of mobile ions into the CG model of DNA is desirable from the physical standpoint. Therefore, in a subsequent model[2] the local dynamics of both the DNA chain and explicit mobile ions, including coupling between DNA motions and the ionic environment fluctuations, have been accurately captured. Specifically, the following interionic interaction potentials and the potentials among beads of DNA and the ions,

$$\mathcal{H}_{\text{ion-ion}} = \sum_{i>j}\left[\frac{A}{r_{ij}^{12}} + \sum_{k=1}^{5} B^{(k)} \mathrm{e}^{-C^{(k)}\left[r_{ij}-R^{(k)}\right]^2} + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}}\right], \quad (21)$$

$$\mathcal{H}_{\text{DNA-ion}} = \sum_{i>j}\left[\frac{A}{r_{ij}^{6}} + \sum_{k=1}^{3} B^{(k)} \mathrm{e}^{-C^{(k)}\left[r_{ij}-R^{(k)}\right]^2} + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon r_{ij}}\right], \quad (22)$$

defined by the set of parameters, $\{A, B^{(k)}, C^{(k)}\}$, and the positions of Gaussian peaks and minima, $\{R^{(k)}\}$, were used in the latest CG DNA model. The functional forms of these potentials were derived in prior works of Savelyev and Papoian[71,67] on the basis of a structural analysis of the corresponding atomistic radial distribution functions (RDF). In these expressions, the first terms indicate the energy due to excluded volume interactions, the second terms are responsible for ionic hydration effects (Gaussian functions were introduced to capture structural ionic peaks and minima in atomistic RDF), and the last terms represent the sum of electrostatic (Coulombic) interactions. Similarly to potentials describing bonded interactions, initial values for parameters entering Hamiltonians (21) and (22) were obtained by fitting the above functional forms into corresponding atomistic PMFs.

The MRG-CG technique developed by Savelyev and Papoian to optimize parameters in effective CG Hamiltonian[33] closely follows the RG Monte Carlo method by Swendsen[68] to compute critical exponents in a three-dimensional Ising model, and extends its applicability to complex molecular systems, such as polymers, which are characterized by significant many-body effects that should not be ignored (e.g., three-body bending angle interactions and dihedral angles along the polymer chain). 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 CG system so to match partition functions of CG and atomistic systems, as explained below. Hence, a 'conjugate field', K_α, is prescribed to each observable, playing a role of a Hamiltonian force constant, whose numerical value has to be adjusted appropriately to generate the desired system dynamics. Because of Hamiltonian's linearity with respect to K_αs, 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,[68]

$$\Delta\langle S_\alpha\rangle = -1/(k_\mathrm{B}T)\sum_{\gamma}[\langle S_\alpha S_\gamma\rangle - \langle S_\alpha\rangle\langle S_\gamma\rangle]\Delta K_\gamma, \quad (23)$$

where $\Delta\langle S_\alpha\rangle \equiv \langle S_\alpha\rangle_{\mathrm{CG}} - \langle S_\alpha\rangle_{\mathrm{AA}}$ is the difference between the expectation values of an observable, S_α, averaged over CG and AA systems, and the ΔK_γ's are corrections to trial CG Hamiltonian parameters, $\{K_\alpha^{(0)}\}$. A set of linear equations, Eq. (23), 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 (absent in numerous other CG techniques) which is responsible for high fidelity of the local CG dynamics. For example, as the Hamiltonian parameters for DNA bending angle potential are iteratively adjusted, the information of what impact of that adjustment would have on all other CG structural degrees of freedom (e.g., bond or stacking dynamical variables) is taken into account. In contrast, it has been well recognized that the commonly used formula for PMF of one specific degree of freedom, $V(\zeta) = -k_\mathrm{B}T\ln p(\zeta)$, neglects cross-correlations among various degrees of freedom, and may potentially lead to significant artifacts, especially for polymeric systems and macromolecules at ambient conditions.[64]

The CG models of DNA derived by Savelyev and Papoian was investigated in several biophysical applications. For example, the model generated values for DNA persistence length in a wide range of concentrations of NaCl salt buffer, approximately [10^{-4}–0.1] M, in near quantitative agreement with several experimental measurements[2] (see Figure 5a). Importantly, the correct large-scale DNA behavior was naturally generated by propagating in scale of local motions optimized with MRG-CG technique. In addition, the model was recently used by by Cao et al.[72] to study

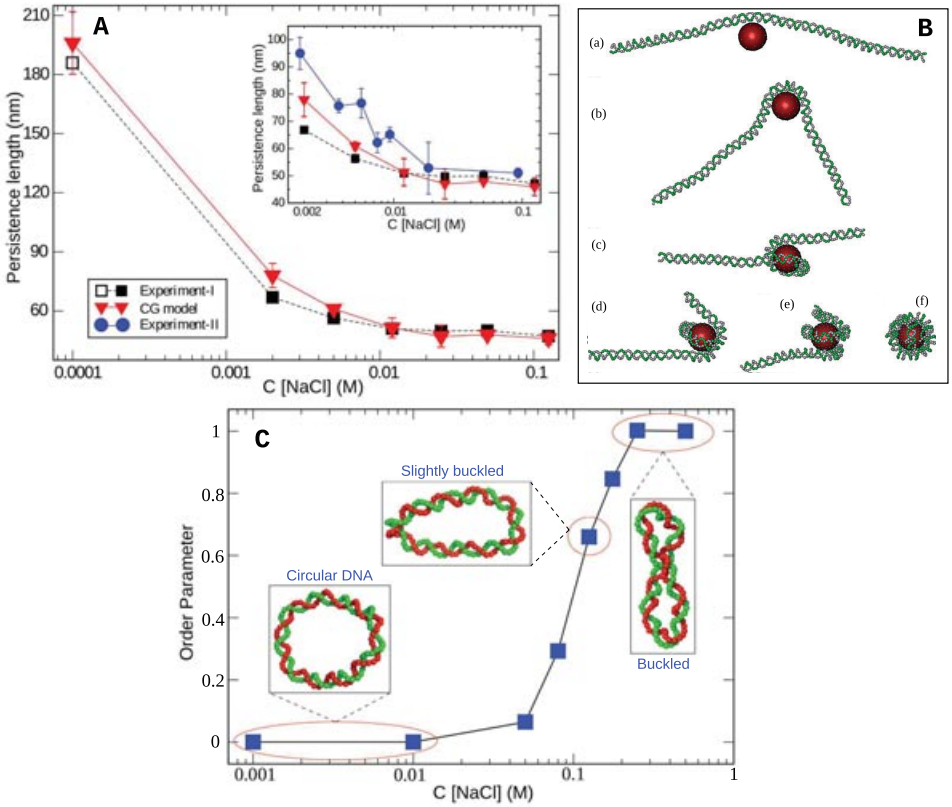

FIGURE 5 | Recent applications of the CG DNA model by Savelyev and Papoian. (a) The model was used to measure DNA persistence length in a wide range of NaCl concentrations, covering three order of magnitudes; computational results appeared to be in quantitative agreement with experimental data. (Graphs are taken from Ref 2.) (b) The model was employed in a study of the behavior of DNA–nanosphere complex mimicking nucleosome core particle[72]; typical MD simulation snapshots demonstrate nucleosomal wrapping (taken from Ref 72). (c) The model predicts a pronounced phase transition to a buckled state in the overtwisted 90-base-pair DNA nanocircle at physiological concentrations of NaCl salt buffer (taken from Ref 2).

the conformational behavior of DNA interacting with a charged sphere, subject to an external stretching force exerted on the ends of DNA chain, mimicking the unwrapping of the nucleosome (see Figure 5b). Another application of this model was the prediction of structural phase transitions in torsionally stressed DNA nanocircles caused by variation of the salt concentration[2] (see Figure 5c). This computational study has an important biological implication in nucleosomal assembly, where torsional stress may be present, and a delicate balance between DNA elastic and electrostatic effects may regulate nucleosomal stability.[73] Finally, recently the same CG model of DNA has been used to address one of the long-standing questions in the field of DNA biophysics, namely, whether DNA's rigidity is dominated by elastic or electrostatic interactions, where it was found that the OSF theory seems to significantly underestimate the electrostatic contribution to DNA's persistent length at physiological conditions.[1] Additional CG simulations, where the monovalent salt concentrations were varied in a range of ~[0.1 – 1] M, also indicate some softening of DNA at high salt concentrations.[58] It turns out that the experimental results are also contradictory with each other, where many experiments ruling out such softening, while other experiments support this possibility (see the corresponding discussion in Ref 1). As to other numerous biological implications involving sequence-dependent effects, such as predictions of sequence-dependent DNA structure and curvature or studying a sequence-dependent melting and hybridization, the homopolymeric CG DNA model of Savelyev and Papoian needs to be extended by introducing all four types of DNA nucleotides. This may be achieved with the MRG-CG technique in a straightforward way by introducing additional interactions that depend on sequence. To further improve the accuracy of CG potentials and to account for finer details of the underlying atomistic system on top of the mean field

picture, one may choose not only to reproduce the expectation values ($\langle S_\alpha \rangle$'s), but also higher-order correlation functions ($\langle S_\alpha \ldots S_\beta \rangle$'s) of various structural observables associated with different monomeric types. This possibility was elaborated in prior works,[2,33]

Finally, CG model of double-stranded DNA by Savelyev and Papoian[2,33] does not allow separation of strands, hence, melting and bubble formation cannot be studied. Hence, it would be interesting to explore new 'hybrid' CG models of DNA, for example, by combining the anisotropic, elastic Hamiltonian of Plotkin and coworkers[37] with the explicit-ion electrostatics of the Savelyev–Papoian DNA model,[2] and use the MRG-CG approach[2,33] to parameterize the resulting overall CG Hamiltonian. As discussed below, such an approach may be one of the possibilities that could lead to the next generation of high-resolution CG models of DNA.

Other CG Descriptions of DNA Chains

During the last decade, a significant number of additional CG models have been put forward for modeling the conformational dynamics and thermodynamics of single- and double-stranded DNA chains.[36,38–40,74,75] A particularly promising one is the CG model developed in the group of Pantano,[36] which provides finer description of the DNA at the basal level by mapping nucleotides into six beads, thus maintaining the chemical skeleton of the AA nucleotide. When combining the DNA model with the CG description of ions and water model WAT4[74] developed by the same group, a fair agreement between CG and AA systems was obtained, producing fine structural details such as the preferential ionic binding to minor grooves, the A −> B structural transformations, the sequence-dependent thermal melting and the DNA breathing dynamics.

Another finer description of the DNA has been provided by DeMille et al.[40] who incorporated ions and water at a CG level by combining the de Pablo's three-site CG model with a water-ion potential developed by the group of Molinero.[76] Calculations based on this model successfully replicate the hydration shell structures around grooves and ionic radial distribution functions. They also mimic the qualitative features of solvation dynamics by giving a good agreement on residence times of water and ions, upon normalizing the inherently fast CG timescale using the CG water diffusion constant. Detailed treatment of solvation in this model will be important for studying various ligand binding phenomena, where solvent is known to play a thermodynamically important role. The key weaknesses of the model are the absence of electrostatic interactions and the intrinsic limitations mentioned in the context of de Pablo's original model, where the former may hinder the understanding of electrostatically driven self assembly phenomena, such as chromatin organization and DNA nanocomposites.

At the opposite end of the spectrum of coarse graining is a single-bead CG model of the DNA by Doi et al.[38] The CG Hamiltonian of the model is minimalistic, including a Morse potential for hydrogen bonding interactions and simple harmonic potentials for maintaining the backbone. Through judicious parameterization the model is capable of reproducing salt dependence of persistence length and shows good agreement of melting behavior with the experiments. Even though there are now many CG models that can reproduce various aspects of DNA elasticity and its salt dependence, the one bead per base resolution presents a significant simplification which can lead to drastic savings of computer time of the simulation. Therefore, if the problem calls for a mesoscale description of the DNA, the model of Doi et al. might be helpful mainly because of its good balance between reasonable physical description and computational feasibility.

At last, there is a work done by Ouldridge et al.[39,75] who took quite an interesting top-down approach to coarse graining the DNA. The authors developed three-bead CG representation of the DNA chain and used it to simulate the DNA nanotwizers.[75] Their model is specifically targeted to reproduce the thermodynamics of DNA melting, which includes hairpin formation, duplex hybridization and various effects associated with stacking. However, the model does not explicitly account for electrostatics and sequence specificity, which significantly limits the range of problems to which it can be applied.

TOWARD MESOSCALE MODELING OF THE CHROMATIN

At last, we would like to provide a brief overview of the coarse-graining efforts in the chromatin modeling area. The computational modeling of chromatin is still a field in its infancy and many challenges have to be addressed to pave the way to the generation of robust and predictive models, which could illuminate experiments and provide novel physical insights. Nevertheless, there seems to be a steadily increasing research activity in the computational molecular science community.[77–85] In particular, the group of Schlick has developed a mesoscale CG representation

of polynucleosome arrays[78,79] based on a discrete surface charge optimization (DiSCO) algorithm.[86] The later approach treats the nucleosome core as a uniformly charged, fine discretized surface with the charges optimized to match the solution of the non linear Poisson–Boltzmann equation for an AA nucleosome. The electrostatic interactions and the salt dependence in chromatin is accounted by the DH approximation. The model was employed for studying the impact of histone tails on the organization of polynucleosomal arrays,[80] architecture of the 30 nm chromatin fiber[81] and the counterion condensation patterns as a function of the geometry of the underlying chromatin fiber.[87] However the mean field treatment of electrostatics at the level of the DH approximation raise serious concerns about the credibility of the obtained physical picture, since the ion–ion correlations are known to be an important factor for inducing polyelectrolyte condensation at such high charge densities. Also the ions with higher valency pose serious challenges of representing them within the framework of DH theory. Therefore, more refined approximations for accounting the electrostatics should be developed to help uncover the driving forces behind chromatin fiber self assembly, its salt dependence and structural changes induced by covalent modifications. Similarly, treating histone tails as a coarse WLC is difficult to justify since it has been recently shown that histone tails are characterized by significant conformational organization and are poorly described as a WLC.[88]

Yang et al.[82] and Korolev et al.[83,84] took a different path for modeling the polynucleosomal arrays by representing the NCP as a linker-free negatively charged sphere connected to charged primitive tail domains and surrounded with the explicit ions. The explicit treatment of electrostatics lead to a good agreement with the experiments on salt induced aggregation, reproducing the sharp rise of the second virial coefficient at the transition point. Also, the model demonstrated well pronounced tail bridging effect and showed reduction in the condensing propensity upon acetylation. Despite the initial promise, however, at the present stage the model is too crude to address the myriad of questions related to the structural organization of the chromatin fiber. In particular, because it does not capture the geometrical shape of NCP and histone tails and models the NCP as a uniformly charged sphere, it is not clear whether it can shed light on structurally specific questions concerning chromatin folding and dynamics. For example, the specific patches on the histone cores are purported to play a key role in mediating the nucleosomal assembly,[89,30] which may not be adequately captured by treating the nucleosome as a sphere. Thus, the further refinement of the model will be necessary to study the structure, dynamics, and thermodynamics of the chromatin fibers.

CONCLUSIONS

With the DNA-based mesoscale systems at the center of the contemporary biophysical and material science research, never before has the need been greater for the development of the simplified CG molecular models of the DNA. At this stage, it has become clear that many long-standing questions in the chromatin science and DNA–nanotechnology could only be pursued by coarse graining the atomic level representation of the DNA and thus paving the way for simulations on scales that are beyond the reach for present models. Several potentially promising CG representations have been developed in the recent years that dramatically reduce the degrees of freedom of the fully atomic models, but yet show the remarkable ability of preserving some key mechanical and chemical properties. The currently available CG representations of the DNA are now at the stage of successfully reproducing phenomena that are sometimes outside the immediate range of their parameterization, and can predict among others effects as diverse as thermal melting curves, salt dependence of persistence length, bubbling dynamics, reversible denaturation. In the lieu of the freshness of proposed models, more efforts are necessary to thoroughly test the performance of the CG models against several more complex systems and in particular against tasks that are relatively uncoupled to the initial design or parameterization of the models. Such investigations will reveal the intrinsic limitations, pinpointing to the places for further improvement and might potentially lead to new insights into the link between the molecular level description and mechanochemical couplings within the DNA. The need for further refinement of CG representations of the DNA is also motivated by the desire to move from the the simple replication of experiments to the prediction of new phenomena, allowing for credible numerical explorations. Thus, we are hoping that the next generation of CG models will to not only help in interpreting the bulk experimental results but also will serve as a guiding light for the new and exciting discoveries.

REFERENCE

1. Savelyev A, Materese CK, Papoian GA. Is DNA's rigidity dominated by electrostatic or nonelectrostatic interactions? *J Am Chem Soc* 2011, 133: 19290–19293.
2. Savelyev A, Papoian GA. Chemically accurate coarse graining of double-stranded DNA. *Proc Natl Acad Sci USA* 2010, 107: 20340–20345.
3. Kornyshev AA. Physics of DNA: unravelling hidden abilities encoded in the structure of 'the most important molecule'. *Phys Chem Chem Phys* 2010, 12: 12352–12378.
4. Peters JP, Maher LJ. DNA curvature and flexibility in vitro and in vivo. *Q Rev Biophys* 2010, 43: 23–63.
5. Ortiz V, de Pablo JJ. Molecular origins of DNA flexibility: sequence effects onconformational and mechanical properties. *Phys Rev Lett* 2011, 106: 238107–238111.
6. Linna RP, Kaski K. Analysis of DNA elasticity. *Phys Rev Lett* 2008, 100: 168104–168108.
7. Marko A, Denysenkov V, Margraf D, Cekan P, Schiemann O, Sigurdsson ST, Prisner TF. Conformational flexibility of DNA. *J Am Chem Soc.* 2011, 133: 13375–13379.
8. Bae S, Kim D, Kim KK, Kim Y-G, Hohng S. Intrinsic Z-DNA is stabilized by the conformational selection mechanism of Z-DNA-binding proteins. *J Am Chem Soc* 2011, 133: 668–671.
9. Bates AD, Maxwell A. *DNATopology*. New York: Oxford University Press; 2005.
10. Cherstvy AG. DNA cyclization: suppression or enhancement by electrostatic repulsions? *J Phys Chem B* 2011, 115: 4286–4294.
11. Schiessel H. The physics of chromatin. *J Phys Condens Matter* 2003, 15: 699–744.
12. Hübner MR, Spector DL. Chromatin dynamics. *Annu Rev Biophys* 2010, 39: 471–489.
13. Levens D, Benham CJ. DNA stress and strain, in silico, in vitro and in vivo. *Phys Biol* 2011, 8: 035011–035020.
14. Bloomfield V. DNA condensation. *Curr Opin Struc Biol* 1996, 6: 334–341.
15. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. New York: Garland Science; 2007.
16. Seeman N. Nanomaterials based on DNA. *Ann Rev Biochem* 2010, 79: 65–87.
17. Rothemund P. Folding DNA to create nanoscale shapes and patterns. *Nature* 2006, 440: 297–302.
18. Tørring T, Voigt NV, Nangreave J, Yan H, Gothelf KV. DNA origami: a quantum leap for self-assembly of complex structures. *Chem Soc Rev* 2011, 40: 5636–5646.
19. Han D, Pal S, Nangreave J, Deng Z, Liu Y, Yan H. DNA origami with complex curvatures in three-dimensional space. *Science* 2011, 332: 342–346.
20. de Pablo JJ. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annu Rev Phys Chem* 2011, 62: 555–574.
21. Tidor B, Irikura KK, Brooks BR, Karplus M. Dynamics of DNA oligomers. *J Biomol Struct Dyn* 1983, 1: 231–252.
22. Levitt M. Computer simulation of DNA double-helix dynamics. *Cold Spring Harb Symp Quant Biol* 1983, 47:251–262.
23. Laughton C, Harris SA. The atomistic simulation of DNA. *WIREs Comput Mol Sci* 2011, 1: 590–600.
24. Pérez A, Luque FJ, Orozco M. Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc* 2007, 129: 14739–14745.
25. Pérez A, Luque FJ, Orozco M. Frontiers in molecular dynamics simulations of DNA. *Acc Chem Res* 2012, 45: 196–205.
26. Perez A, Orozco M. Real-time atomistic description of DNA unfolding. *Angew Chem Int Ed Engl* 2010, 49: 4805–4808.
27. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T 3rd, Dixit S, Jayaram B, Lankas F, Laughton C. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res* 2010, 38: 299–313.
28. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* 2007, 92: 3817–3829.
29. Zgarbová M, Otyepka M, Sponer J, Mládek A, Banáš P, Cheatham TE, Jurečka P. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 2011, 7: 2886–2902.
30. Materese CK, Savelyev A, Papoian GA. Counterion atmosphere and hydration patterns near a nucleosome core particle. *J Am Chem Soc* 2009, 131: 15005–15013.
31. Rudzinski JF, Noid WG. Coarse-graining entropy, forces, and structures. *J Chem Phys* 2011, 135: 214101–214116.
32. Noid WG, Chu J-W, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC. The multiscale coarse-graining method. I. A rigorous bridge between

atomistic and coarse-grained models. *J Chem Phys* 2008, 128: 244114–244123.

33. Savelyev A, Papoian GA. Molecular renormalization group coarse-graining of polymer chains: application to double-stranded DNA. *Biophys J* 2009, 96: 4044.
34. Voth GA, ed. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. New York: CRC Press; 2009.
35. Sambriski EJ, Schwartz DC, de Pablo JJ. A mesoscale model of DNA and its renaturation. *Biophys J* 2009, 96: 1675–1690.
36. Dans P, Zeida A, Machado M, Pantano S. A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. *J Chem Theory Comput* 2010, 6: 1711–1725.
37. Morriss-Andrews A, Rottler J, Plotkin S. A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J Chem Phys* 2010, 132: 035105–035122.
38. Doi K, Haga T, Shintaku H, Kawano S. Development of coarse-graining DNA models for single-nucleotide resolution analysis. *Phil Trans R Soc A* 2010, 368: 2615–2628.
39. Ouldridge TE, Louis AA, Doye JPK. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J Chem Phys* 2011, 134: 085101–085121.
40. DeMille RC, Cheatham TE, Molinero V. A coarse-grained model of DNA with explicit solvation by water and ions. *J Phys Chem B* 2011, 115: 132–142.
41. Machado MR, Dans PD, Pantano S. A hybrid all-atom/coarse grained model for multiscale simulations of DNA. *Phys Chem Chem Phys* 2011, 13: 18134–18144.
42. Beveridge DL, Barreiro G, Byun KS, Case DA, Cheatham TE, Dixit SB, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Seibert E, Sklenar H, Stoll G, Thayer KM, Varnai P, Young MA. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* 2004, 87: 3799–3813.
43. Dixit SB, Beveridge DL, Case DA, Cheatham TE, Giudice E, Lankas F, Lavery R, Maddocks JH, Osman R, Sklenar H, Thayer KM, Varnai P. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II. Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophy J* 2005, 89: 3721–3740.
44. Plimpton S. Fast parallel algorithms for short-range molecular dynamics. *J Comput Phys* 1995, 117: 1–19.
45. Doi M, Edwards SF. *The Theory of Polymer Dynamics*. New York: Oxford University Press; 1988.
46. Kratky O, Porod G. Diffuse small-angle scattering of X-rays in colloid systems. *J Colloid Sci* 1949, 4: 35–70.
47. de Gennes P-G. *Introduction to Polymer Dynamics*. Ithaca: Cornell University Press; 1990.
48. Fredrickson G. *The Equilibrium Theory of Inhomogeneous Polymers*. New York: Oxford University Press; 2007.
49. Bustamante C, Smith SB, Liphardt J, Smith D. Single-molecule studies of DNA mechanics. *Curr Opin Struc Biol* 2000, 10: 279–285.
50. Yuan C, Chen H, Lou XW, Archer LA. DNA bending stiffness on small length scales. *Phys Rev Lett* 2008, 100: 018102–018106.
51. Mastroianni AJ, Sivak DA, Geissler PL, Alivisatos AP. Probing the conformational distributions of subpersistence length DNA. *Biophys J* 2009, 97: 1408–1417.
52. Mazur A. Wormlike chain theory and bending of short DNA. *PhysRev Lett* 2007, 98: 218102–218106.
53. Odijk T. Polyelectrolytes near the rod limit. *J Pol Sci* 1977, 15: 477–483.
54. Skolnick J, Fixman M. Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules* 1977, 10: 944–948.
55. Robinson RA, Stokes RH. *Electrolyte Solutions*. New York: Dover Publications; 1965.
56. Baumann CG, Smith SB, Bloomfield VA, Bustamante C. Ionic effects on the elasticity of single DNA molecules. *Proc Natl Acad Sci USA* 1997, 94: 6185–6190.
57. Dobrynin A. Electrostatic persistence length of semiflexible and flexible polyelectrolytes. *Macromolecules* 2005, 38: 9304–9314.
58. Savelyev A. Do monovalent mobile ions affect DNA's flexibility at high salt content? *Phys Chem Chem Phys* 2012, 14: 2250–2254.
59. Knotts TA IV, Rathore N, Schwartz DC, de Pablo JJ. A coarse grain model for DNA. *J Chem Phys* 2007, 126: 084901–084913.
60. Cheung MS, Finke JM, Callahan B, Onuchic JN. Exploring the interplay between topology and secondary structural formation in the protein folding problem. *J Phys Chem B* 2003, 107: 11193–11200.
61. Hoang T, Cieplak M. Molecular dynamics of folding of secondary structures in Go-type models of proteins. *J Chem Phys* 2000, 112: 6851–6860.
62. Stogryn A. Equations for calculating the dielectric constant of saline water. *IEEE Trans Microw Theory Tech* 1971, 19: 733–736.
63. Savelyev A, Papoian GA. Inter-DNA electrostatics from explicit solvent molecular dynamics simulations. *J Am Chem Soc* 2007, 129: 6060–6061.
64. Fukunaga H, ichi Takimoto J, Doi M. A coarse-graining procedure for flexible polymer chains with

bonded and nonbonded interactions. *J Chem Phys* 2002, 116: 8183–8190.

65. Gay J, Berne B. Modification of the overlap potential to mimic a linear site–site potential. *J Chem Phys* 1981, 74: 3316–3319.
66. Babadi M, Everaers R, Ejtehadi MR. Coarse-grained interaction potentials for anisotropic molecules. *J Chem Phys* 2006, 124: 174708–174718.
67. Savelyev A, Papoian GA. Molecular renormalization group coarse-graining of electrolyte solutions: application to aqueous NaCl and KCl. *J Phys Chem B* 2009, 113: 7785–7793.
68. Swendsen RH. Monte Carlo renormalization group. *Phys Rev Lett* 1979, 42: 859–861.
69. Lyubartsev A, Laaksonen A. Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach. *Phys Rev E* 1995, 52: 3730–3737.
70. Savelyev A, Papoian GA. Electrostatic, steric, and hydration interactions favor Na(+) condensation around DNA compared with K(+). *J Am Chem Soc* 2006, 128: 14506–14518.
71. Savelyev A, Papoian GA. Polyionic charge density plays a key role in differential recognition of mobile ions by biopolymers. *J Phys Chem B* 2008, 112: 9135–9145.
72. Cao Q, Zuo C, Ma Y, Li L, Zhang Z. Interaction of double-stranded DNA with a nanosphere: a coarse-grained molecular dynamics study. *Soft Matter* 2011, 7: 506–514.
73. Gupta P, Zlatanova J, Tomschik M. Nucleosome assembly depends on the torsion in the dna molecule: a magnetic tweezers study. *Biophys J* 2009, 97: 3150–3157.
74. Dans P, Herrera F, Pantano S. Another coarse grain model for aqueous solvation: WAT FOUR? *J Chem Theory Comput* 2010, 115: 3793–3807.
75. Ouldridge T, Louis A, Doyle JPK. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys Rev Lett* 2010, 104: 178101–178105.
76. DeMille RC, Molinero V. Coarse-grained ions without charges: reproducing the solvation structure of NaCl in water using short-ranged potentials. *J Chem Phys* 2009, 131: 034107–034123.
77. Wedemann G, Langowski J. Computer simulation of the 30-nanometer chromatin fiber. *Biophys J* 2002, 82: 2847–2859.
78. Arya G, Zhang Q, Schlick T. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys J* 2006, 91: 133–150.
79. Arya G, Schlick T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci USA* 2006, 103: 16236–16241.
80. Arya G, Schlick T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J Phys Chem A* 2009, 113: 4045–4059.
81. Grigoryev SA, Arya G, Correll S, Woodcock CL, Schlick T. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc Natl Acad Sci USA* 2009, 106: 13317–13322.
82. Yang Y, Lyubartsev AP, Korolev N, Nordenskiöld L. Computer modeling reveals that modifications of the histone tail charges define salt-dependent interaction of the nucleosome core particles. *Biophys J* 2009, 96: 2082–2094.
83. Korolev N, Allahverdi A, Yang Y, Fan Y, Lyubartsev AP, Nordenskiöld L. Electrostatic origin of salt-induced nucleosome array compaction. *Biophys J* 2010, 99: 1896–1905.
84. Korolev N, Lyubartsev AP, Nordenskiïld L. Computer modeling demonstrates that electrostatic attraction of nucleosomal DNA is mediated by histone tails. *Biophys J* 2006, 90: 4305–4316.
85. Perišić O, Collepardo Guevara R, Schlick T. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J Mol Biol* 2010, 403: 777–802.
86. Zhang Q, Beard DA, Schlick T. Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (DISCO) algorithm. *J Comput Chem* 2003, 24: 2063–2074.
87. Gan HH, Schlick T. Chromatin ionic atmosphere analyzed by a mesoscale electrostatic approach. *Biophys J* 2010, 99: 2587–2596.
88. Potoyan DA, Papoian GA. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J Am Chem Soc* 2011, 133: 7405–7415.
89. Sharma S, Ding F, Dokholyan NV. Multiscale modeling of nucleosome dynamics. *Biophys J* 2007, 92: 1457–1470.

