2007-02-23

A coarse grain model for DNA

Thomas A. Knotts
thomas.knotts@gmail.com

Nitin Rathore

See next page for additional authors

Original Publication Citation
Knotts, Thomas A., Iv, Nitin Rathore, David C. Schwartz, and Juan J. de Pablo."A coarse grain model for DNA." The Journal of Chemical Physics 126 (27).

A coarse grain model for DNA

Thomas A. Knotts IV[a)]
Department of Chemical Engineering, Brigham Young University, Provo, Utah 84602

Nitin Rathore[b)]
Amgen Inc., Thousand Oaks, California 91320

David C. Schwartz[c)]
Departments of Genetics and Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706

Juan J. de Pablo[d)]
Department of Chemical and Biological Engineering, University of Wisconsin—Madison, Madison, Wisconsin 53706

(Received 24 July 2006; accepted 13 December 2006; published online 23 February 2007)

Understanding the behavior of DNA at the molecular level is of considerable fundamental and engineering importance. While adequate representations of DNA exist at the atomic and continuum level, there is a relative lack of models capable of describing the behavior of DNA at mesoscopic length scales. We present a mesoscale model of DNA that reduces the complexity of a nucleotide to three interactions sites, one each for the phosphate, sugar, and base, thereby rendering the investigation of DNA up to a few microns in length computationally tractable. The charges on these sites are considered explicitly. The model is parametrized using thermal denaturation experimental data at a fixed salt concentration. The validity of the model is established by its ability to predict several aspects of DNA behavior, including salt-dependent melting, bubble formation and rehybridization, and the mechanical properties of the molecule as a function of salt concentration. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2431804]

I. INTRODUCTION

Over the past two decades, significant experimental advances have increased our ability to control and manipulate individual DNA molecules. This capability has permitted development of high-sensitivity experiments and low-volume, high-throughput assays. Such developments have in turn improved our fundamental understanding of DNA in a wide variety of circumstances. Examples include DNA microarrays,[1] optical mapping,[2,3] and single-molecule force-spectroscopy experiments.[4] Many of these technologies, however, are qualitative in nature; they have yet to realize their full potential and considerable challenges remain.[5] Newly developed devices must be optimized to give consistent, reproducible results, and algorithms must be developed to extract quantitative information. Overcoming these challenges has been hampered by an incomplete, molecular-level understanding of the biophysics involved. Predictive, molecular models of DNA capable of describing length scales ranging from nanometers to microns would be particularly useful in this regard. Such models would not only aid in design and optimization, but also in interpretation of single-molecule experimental data.

Several models of DNA are available in the literature. These range from fully atomistic representations, in which all atoms (including the solvent's) are considered explicitly, to highly coarse grained, where collections of several hundred atoms are represented by a few spherical beads connected by worm-like-chain springs. While a complete description in terms of all atomic coordinates would at first glance appear desirable, as more chemical detail is included in a model the computational requirements associated with its solution increase significantly, thereby restricting severely the length and time scales amenable to study. The challenge is therefore to include just enough detail in a model to capture the physics that are responsible for DNA's relevance in biology.[6]

Atomistic models based on force fields such as CHARMM (Ref. 7) and AMBER (Ref. 8) provide the highest degree of detail; studies employing these representations are generally limited to small oligomers of DNA (usually tens of base pairs in length) or simulation times on the order of tens of nanoseconds. A number of recent reviews have summarized the usefulness of such models.[9–12] Some highlights include a study of the pathways of DNA hybridization for a 3 bp oligonucleotide using transition path sampling,[13] an investigation into the stability of 12 bp tethered DNA on a surface with molecular dynamics,[14] an analysis of the 136 tetranucleotide sequences in the context of 15 bp oligonucleotides,[15] and research into paranemic crossover DNA molecules using oligomers of up to 49 bp in length.[16] The latter study represents one of the largest all-atom simulations of DNA; the molecule measured approximately 17 nm in length, or one third of a persistence length.

The examples above illustrate some of the challenges that are encountered when describing a system with full

[a)]Electronic mail: thomas.knotts@byu.edu
[b)]Electronic mail: nrathore@amgen.com
[c)]Electronic mail: dcschwartz@wisc.edu
[d)]Electronic mail: depablo@engr.wisc.edu

atomic-level resolution. Atomistic models can be used to investigate long molecules for a short amount of time, or short molecules for a longer time, and such calculations generally rely on molecular dynamics techniques. For study of complex phenomena, such as melting, behavior under external fields, bending and stretching, or multiple-molecule interactions, it is advantageous to resort to Monte Carlo sampling techniques. These techniques, which include replica exchange, umbrella sampling, transition path sampling, and various other algorithms, often require extensive amounts of computer time (to simulate multiple copies/trajectories of the system) or a reduced model complexity that facilitates the use of advanced biasing moves. The study of Hagan *et al.* demonstrates this fact; to understand the process of stacking and unstacking along the double helix, only three base pairs were simulated.

At the other end of the spectrum of length scales, e.g., for study of full genomic DNA, several models have been proposed and have shed considerable light onto the dynamical behavior of DNA in various environments.[17] The bead-spring model of Jendrejack *et al.*,[18–20] in particular, gives results in excellent agreement with experimental data for diffusion, structural relaxation, and behavior under different flow fields for bulk and confined DNA.[21] Chopra and Larsen have used a coarse grain model to investigate various aspects of DNA dynamics in confined flows, with specific regard to channels of large dimensions such that the molecule interacts with only one wall at a time.[22] And, more recently, good agreement between single-molecule experimental data for an 84 μm molecule and results from Brownian dynamics simulations of a coarse grain model have further underscored the ability of such representations to describe a variety of rheological properties, including polymer extension, orientation angle, and shear viscosity.[23]

For many applications of interest, however, the approaches mentioned above, either atomistic or bead/spring, are inadequate. For length scales between $\approx$2 nm and $\approx$2 μm, the so-called mesoscale region of multiscale modeling, atomistic models are too computationally demanding and continuum-level models do not provide the resolution or molecular detail required to describe a variety of phenomena, including hybridization, binding of proteins, nanoscale confinement, or melting. In each of these applications, relevant phenomena or processes occur on length and time scales commensurate with the contour length of the molecules and their longest relaxation time, but key effects (e.g., hybridization or melting) occur at a much more localized level. A number of problems of interest in the study of DNA, including microarray design, DNA viral packaging, and single-molecule force spectroscopy, would benefit considerably from a mesoscale representation of DNA.

Such a need has been noted by several authors, and in recent years new mesoscale models for DNA have begun to emerge for use in both theory and simulation. Available mathematical and low-resolution mesoscale models can describe phenomena such as the orientational dependence of successive bases and the elastic properties of the molecule,[24–36] but such formalisms are either not directly applicable to molecular simulation techniques or do not describe melting and hybridization. For example, Bruant and co-workers[37] have investigated several groupings of atoms into rigid bodies and created a model that reproduces bending, torsional, and stretching rigidities. These descriptions, however, do not address thermal denaturation and do not include electrostatic effects. More recently, DNA has been modeled as a complex bead-spring network immersed in a coarse grain solvent,[38] but melting and hybridization are not envisaged in that representation.

Several recent models have been proposed to describe melting and hybridization; however, they could benefit from improvements in other areas. The model of Drukker and Schatz[39] is a bead-spring approach where a nucleotide is represented as two sites—one for the backbone and one for the base. The model allows for hybridization, but it does not include Coulombic interactions and does not describe major and minor grooving. It also does not address the mechanical properties of the molecule. A more recent two-site model adds terms to account for sequence specific base stacking interactions but also neglects Coulombic and mechanical effects and does not possess the correct geometry. Another approach addresses both double and single-chain strands.[40] This bead-pin model reproduces experimental melting curves, but the lattice representation poses some limits on its utility. Moreover, the model does not include electrostatic interactions and does not address the elastic properties of DNA.

This work presents a mesoscale molecular model of DNA that is suitable for study of systems where an understanding of localized phenomena is desirable, and for which long molecules and correspondingly long time scales must be considered. The model permits simulation of DNA from nanometer to micron-length scales. It describes aspects of melting, hybridization, salt effects, Watson-Crick base pairing, the major and minor grooving of DNA, and the mechanical properties of the molecule. Particular emphasis is placed on the predictive capabilities of the model in the context of the thermal and mechanical behaviors of DNA. This article begins with a description of the model. We next outline the protocols and simulation methods used to parametrize and validate the model, and we present the results of the parametrization. The validity of the model is assessed by examining its ability to *predict* several experimentally observed phenomena. These include sequence-, length-, and salt-dependent melting, and the emergence of a large, salt-concentration-dependent persistence length characteristic to double-stranded DNA. Further evidence of the usefulness of the model, with particular regard to melting and hybridization, is provided by simulation of bubble formation and annealing. We conclude our article with a brief summary of our findings and a discussion of current and future applications and extensions of the model.

II. MODEL

The model proposed here was developed to comply with several key tenets. These are

(1) The model should be off-lattice and simple to understand and implement.

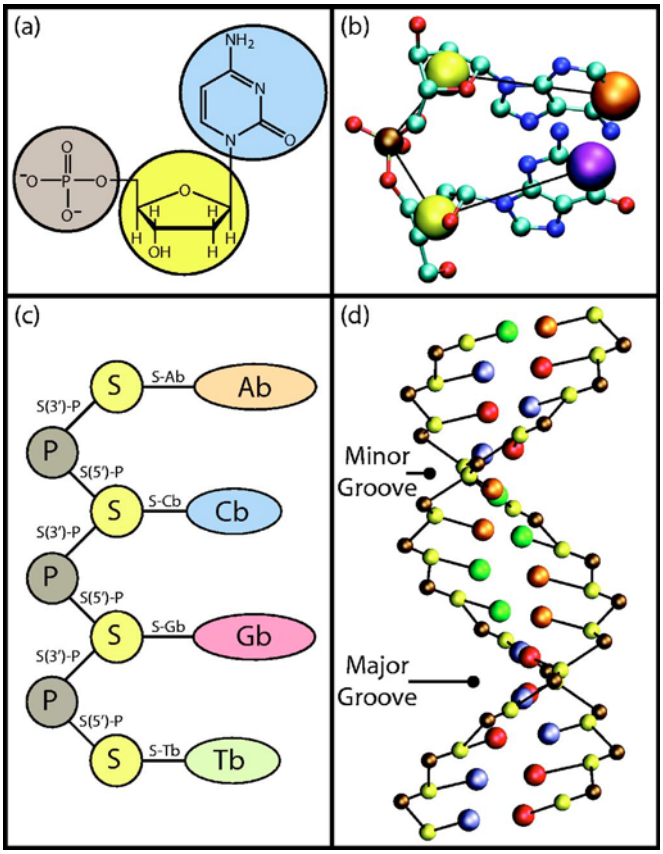

FIG. 1. Schematic representation of the mesoscale model of DNA. Panel (a) Grouping of the atoms for each coarse grain site for a cytosine nucleotide. Panel (b) Atomistic to coarse grain site mapping for the monophosphate dinucleotide 5′-GA-3′. Panel (c) Single-strand topology illustrating the polarity of the strands. Panel (d) Model of a 13 bp oligonucleotide.

(2) The model should reduce the number of sites needed to represent a nucleotide to ease computational demands and permit simulation of long molecules (or concentrated systems) over long time scales.
(3) The model should predict several key physical phenomena, including base specificity, the effects of salt concentration on duplex stability, and the characteristically long persistence length of double-stranded DNA.
(4) The model should permit micron-scale simulations of DNA with nanometer-level resolution, thereby facilitating investigations of DNA in systems such as viral capsids, DNA/histone complexes, and DNA microarrays.

These tenets arise from a desire to have a model that is applicable to a variety of systems and amenable to different simulation techniques, including Monte Carlo (MC) methods, molecular dynamics (MD), and Brownian dynamics.

Our proposed model reduces the complexity of a nucleotide to three interaction sites, one each for the phosphate, sugar, and base. There are four different base sites, one for each type of base in DNA. Panel (a) of Fig. 1 illustrates the groups of atoms represented by each site of the coarse grain model for a cytosine nucleotide. The backbone phosphate and sugar sites are placed at the center of mass of the respective moiety. For purine bases (adenine and guanine), the site is placed at the N1 position. For pyrimidine bases (cytosine and thymine), the site is placed at the N3 position. The coordinates for each of the sites just described were determined from the standard coordinates for the B isoform.[41] The Cartesian and polar coordinates for each of the sites, as well as the protocol to create a DNA molecule, are given in Table I. Also listed are the masses.

TABLE I. Cartesian and polar coordinates and masses for sites in the DNA model. For an atom at (x,y,z), the corresponding atom on the other strand is at $(x,-y,-z)$. Successive residues on each strand are placed using the screw symmetry for the B form of DNA (10 residues per turn and an axial rise per residue of 3.38 Å); thus, an atom at (r,ϕ,z) places the next at $(r,\phi+36°, z+3.38$ Å). This scheme produces a molecule centered upon and aligned with the z axis. See Ref. 41 for more information.

Site (abbreviation)	x (Å)	y (Å)	z (Å)	r (Å)	ϕ (degree)	Mass (amu)
Phosphate (P)	−0.628	8.896	2.186	8.918	94.038	94.97
Sugar (S)	2.365	6.568	1.280	6.981	70.197	83.11
Adenine base (Ab)	0.575	0.516	0.051	0.773	41.905	134.1
Thymine base (Tb)	0.159	2.344	0.191	2.349	86.119	125.1
Cytosine base (Cb)	0.199	2.287	0.187	2.296	85.027	110.1
Guanine base (Gb)	0.628	0.540	0.053	0.828	40.691	150.1

Panel (b) of Fig. 1 depicts the placing of the sites in relation to the atomistic representation for the monophosphate dinucleotide 5′-GA-3′. The phosphate site can bind to a sugar in either a 5′ or a 3′ sense. Panel (c) of Fig. 1, which shows the topology of a single strand, illustrates these bond orientations and introduces the site and bond labeling conventions used in this work. Panel (d) shows a 13 bp oligonucleotide represented with the construction just described and illustrates how the model captures the characteristic major and minor grooves of DNA. Consistent with the atomistic structure, the major groove is approximately twice as wide as the minor groove. Panels (b) and (d) of Fig. 1 were generated using VMD.[42]

The geometrical features of the model, which are imparted by the use of three sites per nucleotide rather than two (as was done by Drukker and Schatz),[39] are important in at least two respects. First, they offer a possibility for "reverse" coarse graining or inverse mapping of the model for coupling between different length scales. The three sites provide the necessary scaffolding to selectively reconstruct an atomistic representation on a local level while keeping the remainder of the molecule coarse grained. Second, such features are necessary for investigations of protein/DNA association. The ability of binding proteins to identify the grooves is a key

TABLE II. Values for energy parameters found in the potential energy function.

Parameter	Value
k_1	ε
k_2	100ε
k_θ	$400\varepsilon/(\text{radian})^2$
k_φ	4ε
ε	0.26 kcal/mol
$\varepsilon_{\text{bp}_{\text{GC}}}$	4ε
$\varepsilon_{\text{bp}_{\text{AT}}}$	$\frac{2}{3}\varepsilon_{\text{bp}_{\text{GC}}}$

TABLE III. Values for geometric parameters found in the potential energy function. A phosphate can bind to a sugar in either a 5′ or a 3′ sense. [See panel (c) of Fig. 1.] Thus, S(5′)-P represents a bond between a phosphate and a sugar belonging to the same nucleotide while S(3′)-P joins together neighboring residues. The bond angle P-(5′)S(3′)-P consists of both types of bonds. [Note: S(5′)−P=P−(5′)S.]

Bond	d_0 (Å)	Bond angle	θ_0 (degree)
S(5′)-P	3.899	S(5′)−P−(3′)S	94.49
S(3′)−P	3.559	P−(5′)S(3′)−P	120.15
S-Ab	6.430	P−(5′)S−Ab	113.13
S-Tb	4.880	P−(3′)S−Ab	108.38
S-Cb	4.921	P−(5′)S−Tb	102.79
S-Gb	6.392	P−(3′)S−Tb	112.72
		P−(5′)S−Cb	103.49
		P−(3′)S−Cb	112.39
		P−(5′)S−Gb	113.52
		P−(3′)S−Gb	108.12

Dihedral angle	ϕ_0 (degree)	Nonbonded	Length (Å)
P−(5′)S(3′)−P−(5′)S	−154.80	σ_{ij}	Interaction specific
S(3′)−P−(5′)S(3′)−P	−179.17	$\sigma_{\mathrm{bp_{AT}}}$	2.9002
Ab−S(3′)−P−(5′)S	−22.60	$\sigma_{\mathrm{bp_{GC}}}$	2.8694
S(3′)−P−(5′)S−Ab	50.69	σ_0 (mismatched bases)	$2^{-1/6}(1.0)$
Tb−S(3′)−P−(5′)S	−33.42	σ_0 (otherwise)	$2^{-1/6}d_{\mathrm{cut}}$
S(3′)−P−(5′)S−Tb	54.69		
Cb−S(3′)−P−(5′)S	−32.72	d_{cut}	$\langle\sigma_{ij}\rangle\approx 6.86$
S(3′)−P−(5′)S−Cb	54.50		
Gb−S(3′)−P−(5′)S	−22.30		

element in their complexation to dsDNA. The proposed model displays these motifs and might therefore facilitate new studies of protein/DNA interactions that cannot be pursued with existing models of DNA.

The potential energy of the system includes seven distinct contributions,

$$V_{\mathrm{total}} = V_{\mathrm{bond}} + V_{\mathrm{angle}} + V_{\mathrm{dihedral}} + V_{\mathrm{stack}} + V_{\mathrm{bp}} + V_{\mathrm{ex}} + V_{\mathrm{qq}}, \tag{1}$$

where

$$V_{\mathrm{bond}} = \sum_{i}^{N_{\mathrm{bond}}} [k_1(d_i - d_{0_i})^2 + k_2(d_i - d_{0_i})^4], \tag{2a}$$

$$V_{\mathrm{angle}} = \sum_{i}^{N_{\mathrm{angle}}} \frac{k_\theta}{2}(\theta_i - \theta_{0_i})^2, \tag{2b}$$

$$V_{\mathrm{dihedral}} = \sum_{i}^{N_{\mathrm{dihedral}}} k_\phi[1 - \cos(\phi_i - \phi_{0_i})], \tag{2c}$$

$$V_{\mathrm{stack}} = \sum_{i<j}^{N_{\mathrm{st}}} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{2d}$$

$$V_{\mathrm{bp}} = \sum_{\text{base pairs}}^{N_{\mathrm{bp}}} 4\varepsilon_{\mathrm{bp}_i}\left[5\left(\frac{\sigma_{\mathrm{bp}_i}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{\mathrm{bp}_i}}{r_{ij}}\right)^{10}\right], \tag{2e}$$

$$V_{\mathrm{ex}} = \sum_{i<j}^{N_{\mathrm{ex}}} \begin{cases} 4\varepsilon\left[\left(\frac{\sigma_0}{r_{ij}}\right)^{12} - \left(\frac{\sigma_0}{r_{ij}}\right)^{6}\right] + \varepsilon & \text{if } r_{ij} < d_{\mathrm{cut}} \\ 0 & \text{if } r_{ij} \geq d_{\mathrm{cut}} \end{cases}, \tag{2f}$$

$$V_{\mathrm{qq}} = \sum_{i<j}^{N} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_k r_{ij}} e^{-r_{ij}/\kappa_D}. \tag{2g}$$

Tables II and III list the values of the energy and geometry parameters that appear in Eq. (2).

The first three terms of Eq. (1) are typical expressions for intramolecular bonds, bond angles, and dihedral angles. The equilibrium distances and angles in these terms are set equal to the values obtained from the atomic coordinates of the standard model of the B form of dsDNA,[41] and are summarized in Table III. We note here that these parameters, in large measure, define the structure of the molecule. Assigning their values relative to the B geometry biases the model to this form of the molecule and renders transitions to other forms of DNA (i.e., A and Z) difficult. A related consequence of this modeling strategy is the implicit assumption that the sugar pucker does not change. As in all coarse-graining approaches, decisions must be made as to which phenomena one wishes to capture in the model. Since the proposed applications of the model involve DNA found in the canonical B form, and removing the B-form bias would require additional angle and dihedral parameters, the bias is deemed acceptable and in line with the key tenets mentioned above.

The remaining terms of Eq. (1) describe various pairwise, nonbonded interactions; the V_{stack} term accounts for the base stacking phenomena and contributes to backbone rigidity. It is an *intra*-strand term and is modeled using the Gō-type, native contact scheme of Hoang and Cieplak[43] with a cutoff radius of 9 Å for the native interaction search. Each σ_{ij} is thus pair dependent. Two sites on the same strand comprise a native contact if the distance between them in the native structure is less than the cutoff distance. The native

structure is determined from the standard coordinates as described in Table I. We note that the 9 Å cutoff scheme creates an interaction not only between bases i and $i+1$, but also between bases i and $i+2$. Smaller cutoff distances were investigated, but these did not adequately maintain the double-helical structure. The term V_{bp} describes hydrogen bonding between any complementary base pair and acts both intra- and interstrand. For this term, $\sigma_{\mathrm{bp_{GC}}}=2.8694$ Å and $\sigma_{\mathrm{bp_{AT}}}=2.9002$ Å. These values are obtained from the standard coordinates[41] and correspond to the respective N1–N3 distances of each complementary pair. The V_{ex} term describes excluded volume interactions. If the pair comprises two mismatched bases (e.g., A-A, A-C, etc.), $\sigma_0=1.0\times 2^{-1/6}$ Å; otherwise, $\sigma_0=6.86\times 2^{-1/6}$ Å.

Coulombic interactions are taken into account using the Debye-Hückel approximation, where κ_D is the Debye length, which is valid for the low-salt, physiological concentrations that are generally encountered in biological systems. Only phosphate sites, which have a −1 charge, contribute to this term. The dielectric constant, ε_k, is set equal to its value for water at room temperature and is 78. We note here that a more desirable representation of DNA would describe counterions in an explicit manner, thereby avoiding the need for an approximate Debye-Hückel treatment. Given the additional complexity and computational demands introduced by counterions, however, we have chosen to limit this first version of our model to the use of screened Coulombic interactions.

Two sites are excluded from all nonbonded interactions (including Coulombic) if they constitute a bond. Any sites forming a bond angle do not experience Coulombic interactions. Moreover, contributions for V_{stack}, V_{bp}, and V_{ex} are mutually exclusive, meaning that a pair of sites contributes to one and only one of these terms. Categorization of each two-body interaction begins by determining which sites belong to V_{stack} by searching for native contacts in the Gō-type sense.[43] Then, those pairs belonging to V_{bp} are identified by searching through the remaining pairwise interactions and selecting the A-T and C-G pairs. All pairs not belonging to the previous two categories are assigned to V_{ex}. This hierarchical approach in assigning nonbonded interactions is necessary to prevent unphysical base pairing between adjacent bases on the same strand. If such an approach is not adopted, two nitrogen base sites could be assigned to both V_{stack} and V_{bp}. For example, if the DNA sequence contains a dinucleotide step of AT, TA, GC, or CG, the coarse grain interaction of the pair of base sites would be both complementary and within the cutoff distance of the stacking term. In reality, such a situation does not occur due to the geometrical constraints of hydrogen bonds. Atomistic models capture this phenomenon because they include all the structural information about the molecule. Since coarse graining removes these finer details, proper categorization and bookkeeping of the interactions is needed to replicate the proper behavior. Stacking is the correct interaction to model in this regard and thus takes precedence over the base pairing term.

This potential energy function contains many adjustable parameters, but only a few affect the behavior of the model to a significant extent. For example, a relatively large change in the value of k_1 causes little change in the value of the melting temperature of a particular oligomer. Those parameters that do play a large role are k_ϕ, ε, and $\varepsilon_{\mathrm{bp_{GC}}}$. The Coulombic term contains no adjustable parameters; the Debye length is related to the ionic strength of the solution through

$$\kappa_D=\left(\frac{\varepsilon_0\varepsilon_k RT}{2N_A^2e_q^2I}\right)^{0.5}, \tag{3}$$

where ε_0 is the vacuum permittivity, N_A is Avogadro's number, e_q is the electronic charge, and I is the ionic strength. Thus, different salt concentrations are taken into account by calculating the appropriate value of κ_D from Eq. (3). For example, to simulate a system with $[\mathrm{Na^+}]=50$ mM, $\kappa_D=13.603$ Å.

III. METHODS

A. Systems

Several DNA sequences of varying lengths and topology were simulated in this study. Each was assembled according to the model described above. Below is a description of each and an appropriate designation.

(1) S1,S2: The model was parametrized with a 14 bp oligomer of DNA for which experimental melting data are available. The sequence, 5′-GCGTCATACAGTGC-3′, and its compliment, 5′-GCACTGTATGACGC-3′, are designated S1 and S2, respectively. The duplex is abbreviated S1·S2. Thermal denaturation of S1·S2, as a function of salt concentration, has been measured experimentally through UV absorbance.[44]

(2) TGGCGAGCAC, CGCCTCATGCTCATC, ATGCAATGCTACATATTCGC: These oligonucleotides were used to validate the model parameters. The melting temperatures of these sequences was measured experimentally by UV absorbance.[45] The GC content and length of each is different from that of S1·S2. The lengths are 10, 15, and 20 bp and the fraction GC contents 70, 60, and 40, respectively.

(3) L60B36: This 60 bp strand of duplex DNA is designed to form a partially melted morphology when the temperature is raised slightly above the melting temperature of the molecule. The conformation is known as a "bubble" and is characterized by single-stranded regions bounded by double-stranded regions. This molecule has been examined experimentally by Zeng *et al.*[46,47] and the sequence is 5′-CCGCCAGCGGCGTTATTACATTTAATTCTTAAGTATTATAAGTAATATGGCCGCTGCGCC-3′.

(4) ssλ[40], dsλ[139], dsλ[281], dsλ[421], dsλ[1489]: Several fragments of DNA from bacteriophage lambda were used to determine both the dsDNA and ssDNA persistence length of the model. The superscript refers to the number of bases/base pairs in each molecule. The topology of each fragment is also denoted as double- or single-stranded. The fragments were obtained by digesting with either the *Taq*I, StyI, or *Hae*III restriction enzymes. Virtual digests were done using Restriction Mapper.[48] The entire genome of bacteriophage lambda

was obtained from Entrez Genome. The fully extended lengths of these fragments are approximately 14, 47, 96, 143, and 500 nm, respectively.

B. Parametrization

Parametrization was carried out in an iterative manner. Initial values for each of the parameters in the potential energy function were selected based upon geometric arguments and following the conventions of Hoang and Cieplak.[43] Replica exchange molecular dynamics simulations were preformed with eight replicas to obtain melting curves. The simulations were performed in the *NVT* ensemble with the temperature maintained using the Nosé-Hoover chain method[49] with four thermostats. The temperature range was 260–400 K with 20 K intervals. The time step was 1 fs; swaps between replicas were attempted every 2000 steps, and nonbonded interactions were cut at $4\kappa_D$. Each replica was equilibrated for 400 ps. A replica was considered to be in an equilibrated state when the potential energy ceased to drift over time. This usually occurred after $\approx$1 ps. After equilibration, each replica was simulated for 10 ns. Thus, the total time of one replica exchange simulation (equilibration and production of all eight replicas) was 83.2 ns. The results of these simulations were analyzed using the weighted histogram analysis method (WHAM).[50] We note here that WHAM allows us to determine the properties of the system as a continuous function of temperature. This is particularly helpful when determining the melting temperature of the DNA molecule of interest.

The parametrization of the model was carried out on S1·S2 at $[Na^+]=50$ mM. The general optimization scheme was to relate all energy values to ε, find the epsilon which reproduced the experimental melting curve, change the other parameters to improve the fit, and then start another cycle. In the final rounds of optimization, $N=20$ independent replica exchange simulations (each with different initial configurations) were performed for each set of parameters in order to achieve good statistical significance. Results reported for an arbitrary property, P, are presented as the average, $\langle P\rangle$, of the N values. Uncertainties were calculated from these N quantities as $\sigma_{\langle P\rangle}/\sqrt{N-1}$, where $\sigma_{\langle P\rangle}$ is the standard deviation of the N averaged property values.

C. Characterization and validation

One of the advantages of a coarse grain model is that it can eliminate high-frequency modes, thereby permitting use of longer time steps than a fully atomistic model. When performing simulations using molecular dynamics, it is advantageous to use the highest time step (Δt) that maintains an accurate integration. One quantity that is useful in assessing an appropriate time step is the conservation of the extended Hamiltonian in an *NVT* simulation using Nosé-Hoover dynamics.[51] If the quantity is not conserved, the time step is too high. The criterion used in this work to monitor the fluctuation in the conserved quantity of the Nosé-Hoover Hamiltonian[51] is the average deviation given by

$$|\Delta E| = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{E_k - E_0}{E_0}\right|, \tag{4}$$

where N is the total number of steps, E_k is the value of the conserved quantity at step k, and E_0 is the initial value of the conserved quantity. For stable integration, $\log|\Delta E| \leq -2.5$. Optimization of the time step consisted of determining $|\Delta E|$ for $\Delta t = 1, 3, 5, \ldots, 21$ fs. For each value of Δt, three independent simulations, with different random number seeds, were performed to estimate the errors in the reported values. The system was S1·S2 at $[Na^+]=50$ mM, the temperature was 300 K, the simulation time of each replicate was 20 ns, and snapshots were saved every 50 ps.

In order to demonstrate the computational requirements of the model, the processor time needed to simulate S1·S2 at $[Na^+]=50$ mM and 300 K was also determined. The length of the simulation was 100 ns. Two time steps were used, namely 1 and 10 fs. Each simulation was performed on a single Intel Xeon processor with a clock speed of 3.0 GHz.

The predictive capabilities of the model were tested using the parameters obtained by optimization with S1·S2 at $[Na^+]=50$ mM. Both thermal and mechanical properties were investigated. For salt-dependent, thermal melting, a replica exchange molecular dynamics scheme, similar to that described for parameterization, was used on S1·S2 at $[Na^+]=20$ and 120 mM and on TGGCGAGCAC, CGCCTCATGCTCATC, and ATGCAATGCTACATATTCGC at $[Na^+]=69$ mM.

For other properties of the system, such as bubble stability and persistence length, traditional molecular dynamic simulations were performed in the *NVT* ensemble with Nosé-Hoover chain dynamics. (See the Parametrization section above for the simulation details.) L60B36 was simulated at 360 K to examine bubble formation. Annealing of the bubble to induce hybridization was done at 300 K.

To characterize the mechanical properties of the system, the persistence length[52] was determined for several fragments of λ-DNA. For double-stranded DNA, simulations were performed on $ds\lambda^{139}$, $ds\lambda^{281}$, $ds\lambda^{421}$, and $ds\lambda^{1489}$. Different fragments were used to verify that length and sequence effects were not present. The simulations were performed with *NVT* molecular dynamics at $[Na^+]=150$ mM and 300 K. To study the dependence of salt on the persistence length, simulations were performed on $ds\lambda^{421}$ at $[Na^+]=8$, 13, 25, 50, 70, 100, and 150 mM at 300 K. To determine the persistence length of single-stranded DNA, $ss\lambda^{40}$ was simulated at $[Na^+]=150$ mM and 300 K. A smaller fragment is used for the single-stranded case to prevent hairpin formation. Five to ten independent simulations were performed for each fragment and condition, and data for analysis were collected only after each molecule was appropriately relaxed.

The persistence length, l_p, can be extracted from the decay of the correlation of unit vectors tangent to a chain according to

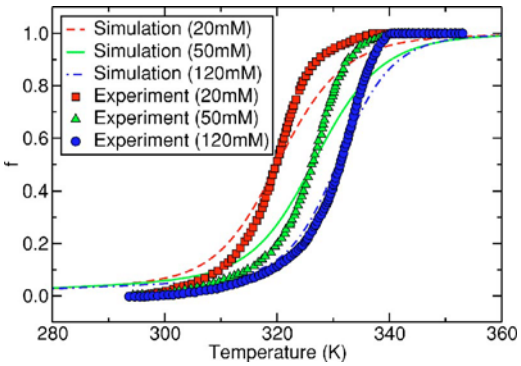

FIG. 2. Agreement between simulated and experimental thermal melting curves for S1·S2 at $[Na^+]=20$, 50, and 120 mM. f is the fraction of denatured bases.

$$\langle \hat{\boldsymbol{u}}(s) \cdot \hat{\boldsymbol{u}}(0) \rangle = \exp\left(-\frac{s}{l_p}\right), \tag{5}$$

where $\hat{\boldsymbol{u}}(s)$ is the tangent unit vector at position s along the length of the chain.[53] For systems in which atoms are represented by discrete sites, the tangent vectors can be defined explicitly by suitably chosen bond vectors along the chain. Since B-DNA has 10 bases per turn, this bond vector, $\boldsymbol{r}_i$ (where i is the contour length index), is chosen as the connection between every tenth sugar site. For example, $\boldsymbol{r}_0 = \boldsymbol{R}_{S10} - \boldsymbol{R}_{S0}$ and $\boldsymbol{r}_1 = \boldsymbol{R}_{S20} - \boldsymbol{R}_{S10}$, where $\boldsymbol{R}_{S10}$ is the position of the sugar site in residue 10. Using this approach, Eq. (5) can be rewritten as

$$\langle \hat{\boldsymbol{r}}_i \cdot \hat{\boldsymbol{r}}_0 \rangle = \exp\left(-\frac{ia}{l_p}\right), \tag{6}$$

where $\hat{\boldsymbol{r}}_i = \boldsymbol{r}_i / \|\boldsymbol{r}\|_i$, $\|\boldsymbol{r}\|_i$ is the norm or magnitude of $\boldsymbol{r}_i$, and a is the average length of the bond vectors given by $a = \langle \|\boldsymbol{r}_i\| \rangle$. The persistence length, l_p, is found by fitting the results of simulations to Eq. (6).

IV. RESULTS AND DISCUSSION

A. Parametrization and salt-dependent melting

The parameters in the model were fit to experimental data for S1·S2, whose melting behavior has been studied experimentally in several salt solutions.[44] Figure 2 shows representative experimental data and simulated results for the fraction of denatured base pairs, f, as a function of temperature for S1·S2 at three salt concentrations. Parameters were fit only at $[Na^+]=50$ mM; the triangles represent that particular concentration. The curves at other salt concentrations, as well as all other simulation results presented hereafter, represent *predictions* of the model; no additional adjustment of parameters was necessary. As the model was fit to reproduce data at $[Na^+]=50$ mM, the melting temperature obtained from simulation at this concentration is, by construction, in agreement with experiment. However, the results also demonstrate that the model is in close agreement with the *temperature range* over which melting occurs. Though the simulated melting transition is slightly broader than its experimental counterpart, particularly in the shoulder areas of the curve, this phenomenon is common in coarse grain approaches, and the discrepancy in this case is minor compared to that encountered in previously available models. The likely cause resides in the actual definition of a denatured base pair. Experimentally, it is known that the absorbance in the UV range of the electromagnetic spectrum is directly proportional to the fraction of denatured base pairs. In simulation, there is some arbitrariness in the molecular-level definition of base pairing; in fact, by simply adjusting the criterion for base pairing, the shoulder regions of the curve can be tuned to provide better agreement with experiment, but this action does not affect the behavior of the system. More important than these finer details in the shape of the curve is the fact that the onset and completion of melting occur at the same temperatures in both experiment and simulation.

TABLE IV. Melting temperatures of S1·S2 at different salt concentrations from simulation and experiment.

$[Na^+]$ (mM)	Melting temperature (K)	
	Simulation	Experiment
20	321.8±2.3	321.0±0.2
50	327.9±2.2	328.5±0.3
120	335.9±1.8	333.2±0.5

The effect of salt on the behavior of DNA is of central importance to its function in numerous applications. Increasing the salt concentration of a solution of DNA causes its melting temperature to increase because the negatively charged backbone of the molecule experiences increased screening. Figure 2 demonstrates that, over the range of concentrations considered in the experiments of Holbrook *et al.*,[44] the model can predict the effects of salt on the melting temperature of DNA. Table IV provides precise figures and error bars for the conditions depicted in Fig. 2.

A more stringent test of the predictive capabilities of the model is provided by simulations of the melting temperature of DNA fragments that differ from S1·S2. To this end, we considered three fragments of complimentary dsDNA for which experimental data are available. The length and GC content of each of these oligonucleotides differ from those of S1·S2. The salt concentration is also different, namely $[Na^+]=69$ mM. Table V includes the sequence of the molecules and the results. The agreement between simulation and experiment is reasonable and serves to demonstrate the general validity of the model.

B. MD time step and "long" simulations

One advantage of a coarse-grained model is that it permits increasing the time step, Δt, required for integration of the equations of motion of the system. An appropriate time

TABLE V. Predicted and experimental (Ref. 45) melting temperatures of DNA duplex oligomers ($[Na^+]=69$ mM).

DNA sequence (5′ to 3′)	Melting temperature (K)		% Error
	Simulation	Experiment	
TCCGCAGCAC	328.7±2.5	317.7	3.5
CGCCTCATGCTCATC	322.8±2.1	326.0	1.0
ATGCAATGCTACATATTCGC	321.9±2.2	328.4	2.0

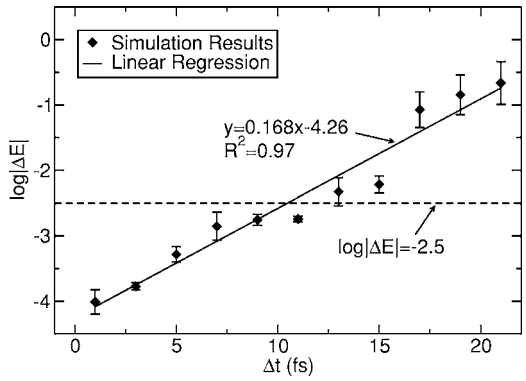

FIG. 3. Average conserved quantity, ΔE, as a function of time step, Δt, for 20 ns molecular dynamics simulation of S1:S2.

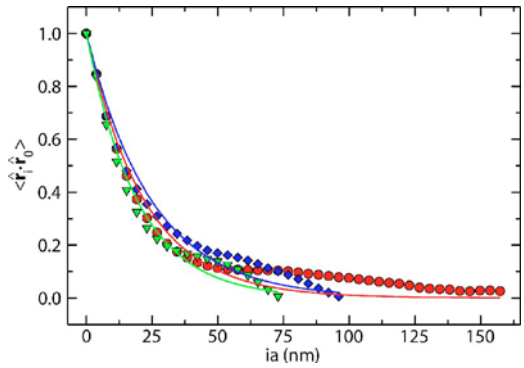

FIG. 4. The average scalar product of successive backbone bond vectors as a function of the distance between those vectors for calculation of the persistence length of dsλ[421]. The symbols represent the simulation data while the lines represent the fit to Eq. (6). Colors represent results from different independent simulations.

step, however, must be small enough to resolve the high-frequency motion of the model of interest. For atomistic models, this usually corresponds to the vibrational motion of bonds containing hydrogen, and a time step of $\approx 0.5-2$ fs (depending on whether constraints are placed on the bonds in question) is the largest that can be used and still lead to proper energy conservation. As no explicit bonds involving hydrogen are present in our coarse grain model of DNA, it is possible to increase considerably the time step for molecular dynamics simulations.

Figure 3 shows the average deviation of the conserved quantity of the Nosé-Hoover Hamiltonian, $\log|\Delta E|$, as a function of the time step. The points represent the results of simulations and associated errors, and the solid line is a fit to a simple linear model. The dashed horizontal line is located at $\log|\Delta E|=-2.5$ to facilitate analysis of the results. The energy conservation as a function of time step exhibits a linear trend fairly well, with a Pearson correlation coefficient, R^2, of 0.97. From the linear analysis, it appears that $\Delta t \approx 10$ fs is the largest time step that can be used with our model and still maintain good energy conservation. This increase is significant (compared to the time step of atomistic simulations), and immediately raises the simulation times that are amenable to study by an order of magnitude.

As discussed above, coarse grain models reduce the computational demands needed to simulate a certain molecule for two reasons: (1) a reduction of the number of sites needed to represent the system, and (2) an increase in the time step of integration (if MD techniques are used). A natural question is how much computer time is required to achieve a certain level of simulation time. To demonstrate the ability of the model in this regard, the computer time needed to simulate 100 ns of S1·S2 (14 bp) on one processor was determined for two different time steps, 1 and 10 fs. (See the Methods section for simulation details.) Though it is a shifting target, 100 ns is considered a "long" simulation time by current standards. The simulation using $\Delta t=1$ fs required 13 h and 45 min of CPU time, while that for $\Delta t=10$ fs required only 1 h and 29 min. Such simulations with an atomistic model with explicit solvent would have required approximately 24 000 sites. Computer times for this system would require tens of days to accomplish, even using the most sophisticated algorithms and software. For example, GROMACS (Refs. 54 and 55) is widely considered to be an efficient MD program. Its website[56] posts a benchmark calculation of a system of 23 207 particles. The hydrogens of the system were treated as dummy particles to remove the bond and angle vibrations and simulate with a time step of 4 fs. A multiple-time-step algorithm was also employed. Results are reported for calculation of Coulombic interactions using both particle mesh Ewald (PME) summation and a cutoff scheme. For PME, the program can produce 1.34 ps of simulation time per day on a 2.8 GHz Intel Xeon processor. If the Coulombic interactions are cut off, the rate increases to 2.512 ps/day. Thus, 100 ns would take approximately 40 and 75 days of cpu time for cutoff Coulombics and PME, respectively. These capabilities are slower than those produced by our model by factors of 640 and 1267, respectively. While S1·S2 is not a particularly long molecule, these numbers serve as a reference point for future investigations and systems.

C. Persistence length and "long" molecules

Over the past several years, single-molecule force-spectroscopy measurements have provided important insights into the mechanical stability of DNA. One key finding has been the fact that dsDNA exhibits worm-like-chain behavior, with a characteristic persistence length, $l_{p_{ds}}$, of 45 -50 nm.[4] An adequate description of this property is important in such applications as viral packaging and optical mapping in micro/nanofluidic devices. We now consider the predictive capabilities of the model in this regard.[57]

The model persistence length for dsDNA was determined from duplex λ-DNA fragments of four different lengths: 139, 281, 421, and 1489 bp. The persistence length of each fragment was determined by fitting simulation results to Eq. (6). Figure 4 shows representative fitting results for dsλ[421]; other fragments show similar behavior. The average scalar product of bond vectors, the left-hand side of Eq. (6), is shown as a function of the distance between these vectors, *ia*. Note the different rates with which the individual replicates decay; this diverse behavior is inherent to persistence length calculations and is also seen experimentally,[4] where different measurements give a broad range of values for l_p.

TABLE VI. Persistence length of fragments of double-stranded, λ-DNA.

Fragment	$l_{p_{ds}}$ (nm)
$ds\lambda^{139}$	21.2±4.1
$ds\lambda^{281}$	16.2±0.8
$ds\lambda^{421}$	19.6±1.1
$ds\lambda^{1489}$	22.0±1.6

Table VI summarizes the outcome of the analysis for each fragment. The dsDNA persistence length for the model is 20±1 nm. The experimentally determined value at the same salt concentration is approximately 45–50 nm. The model therefore reproduces the mechanical properties of dsDNA within a factor of only ≈2.3±0.1. While at first glance such a discrepancy might appear large, we note that previous coarse grain models capable of describing the melting and hybridization of DNA have persistence lengths on the order of tens of angstroms. The present model gives a value of tens of nanometers. Given that the persistence length did not enter our parametrization of the model, we view the factor of 2 as reasonable but note the room for improvement. The value for the persistence length is largely dependent upon k_ϕ and ε of Eqs. (2c) and (2d), respectively. A future parametrization of the model will address this issue.

Further validation of the model is possible by investigating the mechanical properties of ssDNA. Experimental reports for the persistence length of single-stranded DNA, $l_{p_{ss}}$, range between 0.75 and 3 nm.[58,59] Simulations of a single-stranded, 40 base fragment of λ-DNA (abbreviated $ss\lambda^{40}$) at $[Na^+]=150$ mM gives $l_{p_{ss}}=1.9\pm0.1$ nm, which is in the middle of the experimental range. The error in $l_{p_{ss}}$ is determined from results of ten independent simulations.

It has been experimentally observed that the persistence length of dsDNA depends on salt concentration.[4] As salt concentration decreases, the screening of the charges along the backbone of the DNA molecule decreases and the phosphate groups seek to separate themselves from their neighbors. The result is a more elongated molecule at lower salt concentrations than in more screened environments. Baumann *et al.* showed that the behavior is captured by the nonlinear Poisson-Boltzmann theory for uniformly charged cylinders;[4] these authors showed that

$$l_p = l_{p_0} + l_{el} = l_{p_0} + \frac{1}{4\kappa^2 l_B} = l_{p_0} + 0.324I^{-1} \text{ nm}, \quad (7)$$

where l_B is the Bjerrum length (7.14 Å for water at 25 °C) and l_{p_0} and l_{el} are the nonelectrostatic and electrostatic contributions to the persistence length, respectively.

Our proposed model captures this effect of salt concentration on persistence length. Figure 5 shows the dependence of the persistence length on salt concentration for $ds\lambda^{421}$. The symbols are the simulation results and the line is the best-fit line to Eq. (7) with $l_{p_0}=17$ nm. Baumann *et al.*[4] report that l_{p_0} values of 45 and 50 nm fit the experimental data equally well. As discussed above, the persistence length of the model is off by a factor of ≈2.3, so the discrepancy in l_{p_0} is expected; the proposed model, however, reproduces the relative effect of salt concentration on persistence length quantitatively.

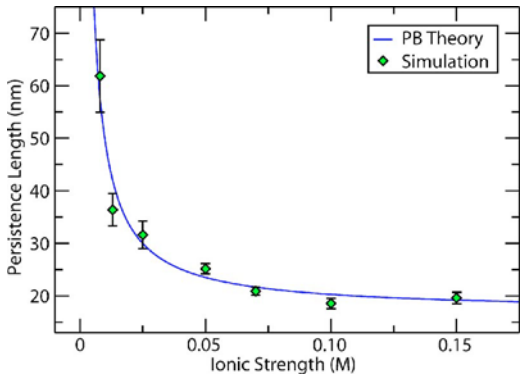

FIG. 5. Dependence of dsDNA persistence length, $l_{p_{ds}}$, on salt concentration.

These calculations serve to underscore an important attribute of the model. The length of the longest molecule considered in this work ($ds\lambda^{1489}$) is about 0.5 μm. Simulations of 0.5 μm dsDNA with an atomistic representation would involve ≈10^9 sites; the demands of microsecond calculations for a system of that size are well beyond current computational capabilities. The model proposed in this work is able to do so with only ≈1/10 000 of the sites (a 10^5-fold reduction), while still giving results in agreement with experiments; it permits simulation of micron-length DNA with molecular resolution and opens up opportunities for investigation that were previously not accessible.

D. Bubble dynamics

One of the key tenets in the design of this model was the need to describe sequence-dependent melting and hybridization. DNA bubble formation and annealing provide a suitable test of the model in this regard. A "bubble" is a structure containing a single-stranded, melted region (the bubble) bounded by a double-stranded region on either side. It is produced because the melting and hybridization of dsDNA is accomplished through intermediates states. Since understanding and characterizing these states is important in understanding the cooperativity of melting and hybridization and how DNA is manipulated in biological processes, studying bubbles not only serves to validate the model but also provides an interesting molecular-level view into an important aspect of DNA biophysics. In nature, regions of dsDNA are "opened" and "closed" repeatedly in processes such as replication and translation; the study of bubble morphology is therefore an active area of research.[47] In particular, Zeng *et al.*[46,47] have studied the formation and hybridization of bubbles using a 60 bp sequence of dsDNA (termed L60B36 hereafter). This molecule consists of two GC-rich regions at its ends, and an AT-rich region in its middle. It is experimentally observed that, upon heating in a solution of 50 mM ionic strength, the middle of the sequence melts to produce a bubble. Molecular dynamics simulations indicate that our model also produces a stable bubble for $[Na^+]=50$ mM when the temperature is raised above the melting temperature and held at 360 K.

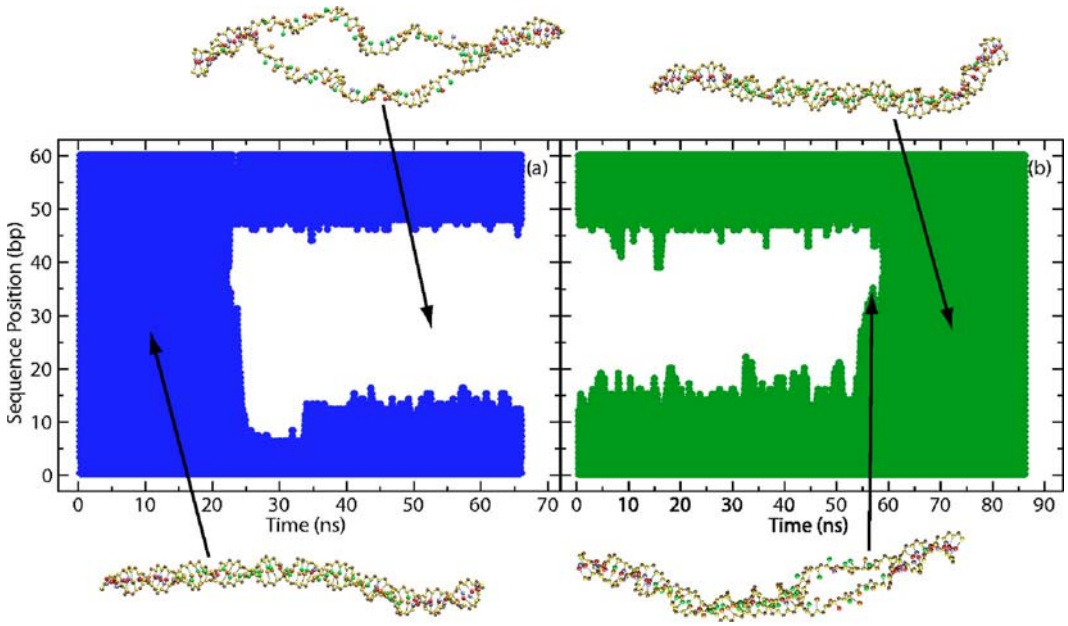

FIG. 6. Processes of bubble formation (a) and rehybridization (b) of L60B36.

Figure 6 describes this process of bubble formation. It also shows the dynamics of bubble annealing. Annealing is accomplished by dropping the temperature of the DNA bubble from 360 K back to room temperature (300 K). The figure depicts the occurrence of each natively hydrogen-bonded base pair, labeled 1 through 60 (according to the position in the molecule), as a function of time. Panel (a) shows the bubble formation at 360 K and panel (b) the hybridization or closing of the bubble after quenching to 300 K. Also shown in the figure are representative snapshots of the molecule, which clearly reveal the formation and disappearance of the bubble.

A comparison of panels (a) and (b) suggests that bubble formation occurs on a faster time scale than rehybridization. For this molecule, bubble formation is complete in 15 ± 1 ns, while rehybridization is complete in 44 ± 13 ns. (These results represent an average from five independent simulations.) The rehybridization process is approximately three times slower than the actual formation of the bubble. The spread in the times of both processes, however, is considerable. The ranges are $3.6-51$ and $15-82$ ns for bubble formation and rehybridization, respectively. While the actual characteristic times that emerge from simple molecular dynamics of our model should be viewed with caution (the model does not include specific solvent effects, friction losses, or hydrodynamic interactions), the results paint a useful picture of the mechanism by which annealing might occur in the laboratory. Hybridization does not take place in one continuous motion. Rather, short "zipping" events take place where several base pairs on one end of the bubble re-pair with their complements but melt soon after. The longer the zipped region, the longer the re-pairing persists. The critical length needed for complete hybridization appears to be ≈12 bp. Zipping and unzipping continues until this critical length is attained (at ≈15 ns with base pairs 15–27). This re-paired stretch then expands to bp 35 (Fig. 6 shows this state) and the hybridization is completed from the other side soon after.

V. CONCLUSION

A. Summary

A predictive, coarse grain model for DNA has been proposed. The validity of the model has been established by comparing its predictions to available experimental thermal and mechanical data. The model preserves considerable molecular-level detail by reducing the complexity of a nucleotide to three interaction sites, while permitting simulation of micron-long DNA. It successfully reproduces salt-dependent thermal melting and the dynamics of bubble formation. It exhibits a persistence length for dsDNA consistent with experiment (although lower by a factor of 2), and reproduces quantitatively the effect of salt concentration on that persistence length. It also correctly describes the persistence length of ssDNA. The model is relatively simple, and should facilitate investigations of a variety of systems that would benefit considerably from simulations, but that have not been amenable to numerical studies as a result of computational limitations. Such systems include DNA microarrays, viral DNA packaging, and high-throughput DNA microfluidic devices for optical mapping.

B. Future work

While the model, in its present form, could be useful in a variety of settings, opportunities exist to expand its applicability. For example, one benefit of designing the model with three sites per residue rather than a two-site construction is the presence of a true major and minor groove. Such structure is important for binding proteins. These proteins recognize certain sequences within the grooves when forming complexes with the double helix. Since the model contains both sequence information and grooving, DNA/protein studies can be performed by using a protein model that is aware of the base types of the DNA. Modeling RNA represents another natural extension of the model. This can be

accomplished by simply defining a new base type, uracil, that pairs with the cytosine site. The standard coordinates of RNA are available[41] from which the bond, angle, and dihedral parameters can be obtained. With a RNA model that is compatible with that of DNA, new insights might be gained into the binding competition that the two nucleic acids experience.

Another opportunity concerns the joining of models of different resolution in one simulation environment. This was alluded to previously. The idea is to represent critical regions of the molecule atomistically, such as the protein binding site of the duplex, while modeling the remainder of the system with more coarse grain approaches. The model presented here is useful in this regard; it maintains the geometric shape seen in atomistic models, thereby making it feasible to link the two scales. Since it exhibits the properties of DNA over long length scales, it is congruent with more coarse grain approaches. The difficulty lies in appropriately linking the two length scales. This is an area of active research in the community and encompasses a variety of classes of compounds.

One area where improvements are necessary is in the persistence length. As previously discussed, the persistence length is off by a factor of 2. Current work includes efforts to bring the persistence length given by the model into closer agreement with experiment. We also note that the use of Debye-Hückel screening to account for salt effects poses a number of limitations. Our results indicate that this approach works well for low-salt conditions that are generally encountered in biological systems. For conditions of high salt, a more involved approach that includes explicit counterions must be considered. The addition of counterions poses a number of challenges; an appropriate scheme must be selected to calculate the Coulombic energy (e.g., shifting, switching, Ewald summation), the energetic balance between the DNA and the counterions must be deduced, and the size of the problem must be contended with. Some of our current efforts are aimed at including counterions to not only give a more complete picture of the behavior at low salt concentrations under confinement, but also to enable simulation of concentrations above the 150 mM limit of Debye-Hückel.

ACKNOWLEDGMENTS

The authors are grateful to Manan Chopra for useful discussions. This work is supported by the National Science Foundation through the Nanoscale Science and Engineering Center (NSEC) at the University of Wisconsin (NSEC DMR-0425880). T. Knotts is grateful for support from the NIH (NHGRI 5T32HG002760). D. Schwartz is also funded by NHGRI 5R01HG000225.

[1] Nat. Genet. **21**, (1s): Entire volume (1999).
[2] C. Aston, B. Mishra, and D. C. Schwartz, Trends Biotechnol. **17**, 297 (1999).
[3] S. Zhou, J. Herschleb, and D. C. Schwartz, *A Single Molecule System for Whole Genome Analysis. New Methods in DNA Sequencing* (Elsevier, Amsterdam, in press), p. 16.
[4] C. G. Baumann, S. B. Smith, V. A. Boomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).
[5] M. A. Branca, Nat. Biotechnol. **23**, 769 (2005).
[6] R. Phillips, The Bridge **34**, 22 (2004).
[7] A. D. MacKerell, Jr., J. Wiórkiewicz-Kuczera, and M. Karplus, J. Am. Chem. Soc. **117**, 11946 (1995).
[8] T. E. Cheatham III, P. Cieplak, and P. A. Kollman, J. Biomol. Struct. Dyn. **16**, 845 (1999).
[9] D. L. Beveridge and K. J. McConnell, Curr. Opin. Struct. Biol. **10**, 182 (2000).
[10] T. E. Cheatham III and P. A. Kollman, Annu. Rev. Phys. Chem. **51**, 435 (2000).
[11] T. E. Cheatham III, Curr. Opin. Struct. Biol. **14**, 360 (2004).
[12] T. E. Cheatham III, *Molecular Modeling and Atomistic Simulation of Nucleic Acids* (Elsevier, Amsterdam, 2005), Vol. 1, Chap. 6, pp. 75–89.
[13] M. F. Hagan, A. R. Dinner, D. Chandler, and A. K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. **100**, 13922 (2003).
[14] K.-Y. Wong and B. Montgomery Pettitt, Biopolymers **73**, 570 (2004).
[15] S. B. Dixit and L. David, Biophys. J. **89**, 3721 (2005).
[16] P. K. Maiti, T. A. Pascal, N. Vaidehi, J. Heo, and W. A. Goddard III, Biophys. J. **90**, 1463 (2005).
[17] E. S. G. Shaqfeh, J. Non-Newtonian Fluid Mech. **130**, 1 (2005).
[18] R. M. Jendrejack, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **116**, 7752 (2002).
[19] R. M. Jendrejack, D. C. Schwartz, M. D. Graham, and J. J. de Pablo, J. Chem. Phys. **119**, 1165 (2003).
[20] R. M. Jendrejack, D. C. Schwartz, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **120**, 2513 (2004).
[21] Y.-L. Chen, M. D. Graham, and J. J. de Pablo, Macromolecules **38**, 6680 (2005).
[22] M. Chopra and G. Ronald, J. Rheol. **46**, 831 (2002).
[23] C. M. Schroeder, R. E. Teixeira, E. S. G. Shaqfeh, and S. Chu, Macromolecules **38**, 1967 (2005).
[24] R. C. Maroun and W. K. Olson, Biopolymers **27**, 561 (1988).
[25] R. C. Maroun and W. K. Olson, Biopolymers **27**, 585 (1988).
[26] M.-H. Hao and W. K. Olson, Biopolymers **28**, 873 (1989).
[27] R. K. Z. Tan and S. C. Harvey, J. Mol. Biol. **205**, 573 (1989).
[28] R. K. Z. Tan, D. Sprous, and S. C. Harvey, Biopolymers **39**, 259 (1996).
[29] D. Sprous, R. K. Z. Tan, and S. C. Harvey, Biopolymers **39**, 243 (1996).
[30] D. Sprous and S. C. Harvey, Biophys. J. **70**, 1893 (1996).
[31] A. Matsumoto and W. K. Olson, Biophys. J. **83**, 22 (2002).
[32] B. D. Coleman, W. K. Olson, and D. Swigon, J. Chem. Phys. **118**, 7127 (2003).
[33] M. Peyrard, Nonlinearity **17**, R1 (2004).
[34] J. C. LaMarque, Biopolymers **73**, 348 (2004).
[35] A. Flammini, A. Maritan, and A. Stasiak, Biophys. J. **87**, 2968 (2004).
[36] A. Vologodskii, Biophys. J. **90**, 1594 (2005).
[37] N. Bruant, D. Flatters, R. Lavery, and D. Genest, Biophys. J. **77**, 2366 (1999).
[38] H. L. Tepper and G. A. Voth, J. Chem. Phys. **122**, 124906 (2005).
[39] K. Drukker and G. C. Schatz, J. Phys. Chem. B **104**, 6108 (2000).
[40] M. Sales-Pardo, R. Guimerà, A. A. Moreira, J. Widom, and L. A. N. Amaral, Phys. Rev. E **71**, 051902 (2005).
[41] P. J. Struther Arnott, C. Smith, and R. Chandrasekaran. *Atomic Coordinates and Molecular Conformations for DNA-DNA, RNA-RNA, and DNA-RNA Helices*, Vol. 2 of CRC Handbook of Biochemistry and Molecular Biology, 3rd ed. (CRC Press, Cleveland, 1976), pp. 411–422.
[42] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).
[43] T. X. Hoang and M. Cieplak, J. Chem. Phys. **112**, 6851 (2000).
[44] J. A. Holbrook, M. W. Capp, R. M. Saecker, and M. Thomas Record, Jr., Biochemistry **38**, 8409 (1999).
[45] R. Owczarzy, Y. You, B. G. Moreira, J. A. Manthey, L. Huang, M. A. Behlke, and J. A. Walder, Biochemistry **43**, 3537 (2004).
[46] Y. Zeng, A. Montrichok, and G. Zocchi, Phys. Rev. Lett. **91**, 148101 (2003).
[47] Y. Zeng, A. Montrichok, and G. Zocchi, J. Mol. Biol. **339**, 67 (2004).
[48] http://www.restrictionmapper.org
[49] G. J. Martyna, M. L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).
[50] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, J. Comput. Chem. **13**, 1011 (1992).
[51] M. Watanabe and M. Karplus, J. Chem. Phys. **99**, 8063 (1993).
[52] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[53] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1988), pp. 316–317.

[54] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[55] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[56] http://www.gromacs.org
[57] K. Jo, D. M. Dinghra, T. Odijk, J. J. de Pablo, M. D. Graham, R. Runnheim, D. Forrest, and D. C. Schwartz (unpublished).
[58] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[59] M. C. Murphy, Biophys. J. **86**, 2530 (2004).