A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality

Alex Morriss-Andrews, Joerg Rottler, and Steven S. Plotkin[a)]
Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T1Z1, Canada

We introduce a coarse-grained model of DNA with bases modeled as rigid-body ellipsoids to capture their anisotropic stereochemistry. Interaction potentials are all physicochemical and generated from all-atom simulation/parameterization with minimal phenomenology. Persistence length, degree of stacking, and twist are studied by molecular dynamics simulation as functions of temperature, salt concentration, sequence, interaction potential strength, and local position along the chain for both single- and double-stranded DNA where appropriate. The model of DNA shows several phase transitions and crossover regimes in addition to dehybridization, including unstacking, untwisting, and collapse, which affect mechanical properties such as rigidity and persistence length. The model also exhibits chirality with a stable right-handed and metastable left-handed helix. © *2010 American Institute of Physics*. [doi:10.1063/1.3269994]

I. INTRODUCTION

DNA is likely the most well-studied biomolecule, with structural, energetic, and kinetic characterizations spanning over 6 decades of research since its isolation as the carrier of genetic information by Avery and co-workers.[1] Elucidating the varied behaviors of DNA has been primarily an experimental endeavor due in large part to the difficulties in capturing the molecule's complex motion and function either computationally or theoretically. The computational difficulties are primarily due to the fact that much of the interesting behavior takes place on time scales three to six orders of magnitude longer than the longest all-atom simulations of a system comparable to the typical size of a gene (currently nanoseconds[2]).

For example, RNA polymerase transcribes DNA at a rate of about 14 ms/nucleotide in eukaryotes[3] during elongation with comparable rates in *E. Coli.*[4] These rates are further slowed by transcriptional pausing to regulate arrest and termination[5–7] by ~seconds per pause. Even "fast" processes such as bacteriophage DNA ejection have translocation times greater than 10 μs/bp.[8,9] Time scales for DNA packaging into the viral capsid are ~10 ms/bp.[10] Nucleosome condensation time scales at *in vivo* histone concentrations are ~10 ms.[11] To address any of these biologically relevant phenomena computationally currently requires the introduction of coarse-grained models. The choice of coarse-grained DNA model reflects the empirical phenomena the model intends to capture. For example, a description of zinc-finger protein binding to DNA would require an accurate representation of major and minor grooves, whereas a description of the sequence dependence of nanopore translocation would require an accurate description of the stereochemistry of bases. As in the above examples, these phenomena also exhibit slow kinetics compared to time scales accessible with

atomistic simulations. Binding rates for transcription factors[12] are of the order of 1/ms per promoter at a cellular concentration of ~10^6/nucleus, and translocation times for single-stranded DNA at typical experimental voltages are of the order of 100 μs for a 100 bp sequence.[13]

On the coarsest scale, a piece of double-stranded DNA may be described as a semiflexible polymer using the worm-like chain model. On this level, the only parameters characterizing the molecule are its total length and its bending rigidity κ, which determines the persistence length $\ell_p = \kappa/k_\mathrm{B}T$. This model is very successful in capturing chain properties on scales larger than ℓ_p, for instance, the force-extension relation,[14] but carries no information about the internal structure of the double strand. Slightly more refined models approximate single-stranded DNA as a semiflexible chain of sterically repulsive spheres,[15,16] which may carry charge and thus interact with an external field, although Coulombic monomer-monomer interactions were neglected in these models. One bead per nucleotide representations of DNA have also been used to describe supercoiling and local denaturation in plasmids.[17] Bead-spring models of double-stranded DNA with nonlinear interchain coupling through hydrogen bonds have been used to study vibrational energy transport and localization.[18] An extension of the bead-spring idea, where base pairs were represented by a planar collection of 14 harmonically coupled beads, allowed for an investigation of spontaneous helix formation from initial ladder-like conformations.[19] More detailed representations have been proposed that describe DNA on the level of spherically symmetric base monomers attached to a chain of similar monomers representing the backbone.[20–22] In these approaches base molecules and repeat units on the backbone are modeled as spheres that interact with other bases through phenomenological potentials mimicking van der Waals and hydrogen bonds. On this level of resolution, it is, for example, possible to study thermal denaturation of DNA, as

---
a)Electronic addresses: steve@phas.ubc.ca and steve@physics.ubc.ca.

well as mechanical properties such as bending rigidity. Zhang and Collins[20] treated base-ribose moieties as a single rigid body with residues at the positions of hydrogen bonding heavy atoms. While the rigid bodies were constrained to undergo two-dimensional motion, base pairing through a modified Morse potential allowed for a systematic study of the structural changes during thermal denaturation. More recently, Knotts *et al.*[22] introduced a refined bead-spring model that distinguishes between sugar and phosphate groups on the backbone and individual base molecules. After parameter adjustment, which included introducing a Gō-like potential[23] to bias the system to the crystal structure, the model was able to describe several features of DNA physics such as the dependence of ℓ_p and duplex stability on ionic concentration, as well as the dynamics of thermally activated bubble formation in hybridization.

For phenomena such as protein-DNA binding and DNA translocation, the base stereochemistry characterizing the DNA sequence is of particular importance. To this end, we introduce a model of DNA where phosphate and sugar groups are represented by one coarse-grained (spherical) residue and bases by a rigid-body ellipsoid. Sterically, bases of DNA more accurately resemble a flat plate than a spherical object. Energetically, base-stacking interactions, predominantly governed by electron correlation (van der Waals) interactions,[24] play a significant if not dominant role in the stability of the double helix.[25] We adopt a systematic coarse-graining approach, where we parameterize the effective interactions through all-atom simulations wherever possible. The interactions describing the stacking of bases are optimized against a fully atomistic representation of the base residues. Interactions describing covalent bonds, angles, and dihedrals along the DNA strand are obtained from equilibrium simulations of a short all-atom strand of DNA. One central difference in the present model from previous models is therefore the absence of any structure-based potentials, i.e., we do not use effective Gō-like potentials to bias the system toward an experimentally determined structure. All behavior in the present model is a direct consequence of physicochemical interactions and the systematic parameterization procedure.

The definition of the model and details of our coarse-graining scheme are described in the subsequent Model section and in the supplementary material (SM).[26] We then investigate the predictions of the model for several important properties of DNA, in particular persistence length ℓ_p and radius of gyration R_g of both double- and single-stranded DNA. We introduce intuitive and quantitative methods to calculate both helical twist and the stacking of bases as order parameters characterizing the degree of native DNA structure. The behavior of these quantities is quantified as several internal and environmental factors are varied, including sequence, interaction strength, strand length, temperature, and ionic concentration. Finally, we investigate the chiral bias of right-handed DNA over left-handed DNA exhibited by the model.
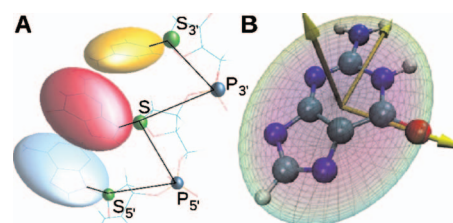

FIG. 1. Our coarse-graining scheme overlaid onto an all-atom visualization. Panel (a) shows a section of three nucleotides: spherical sugar (s) and phosphate (p) residues together with ellipsoidal bases. Panel (b) shows a close up of a base with ellipsoid overlay. The axes indicate the principal axes of the base and are aligned with the principal axes of the ellipsoid, which determine the base's initial orientation.

II. MODEL

A. Parameterization of the model

Figure 1 depicts a piece of DNA in our coarse-grained model superimposed on the all-atom representation. Alternating sugar (S) and phosphate (P) groups are replaced by spherically symmetric residues. The sugar residue is located at the position of the $C1'$ atom in standard crystallographic notation, which is connected to the base by a single covalent bond. The phosphate residue is located at the center of mass of the phosphate group (PO_4). The bases adenine, cytosine, guanine, and thymine are represented as ellipsoidal objects whose structure and interaction potential are described below.

An immediate problem in coarse-grained systems is the loss of knowledge regarding the Hamiltonian or energy function governing dynamics in the system. In all-atom classical molecular dynamics simulations, this problem is resolved by characterizing quantities such as partial charges and interaction potentials with empirical parameters derived from best fit to combinations of experimental and *ab initio* quantum mechanical target data.[27] We wish to adopt a similar methodology with respect to the extraction of phenomenological parameters here by characterizing our coarse-grained system in terms of effective parameters derived from all-atom simulations. The potential energy function for our system has the following form:

$$V = V_{RE^2}(\mathbf{B}_1, \mathbf{B}_2) + V_{RE^2}(\mathbf{B}_1, \text{res}_2) + V_{HB}(\mathbf{B}_1, \mathbf{B}_2)$$
$$+ V_{LJ}(r_{ss}, r_{sp}) + V_C(r_{pp}) + V_{bond}(r) + V_{angle}(\theta)$$
$$+ V_{dihedral}(\phi). \tag{1}$$

The individual terms in Eq. (1) are described below.

B. Base-base van der Waals interactions

The base-base interaction potentials $V_{RE^2}(\mathbf{B}_1, \mathbf{B}_2)$ must account for the shape and relative orientation of the anisotropic ellipsoids representing the bases. We adopt the functional form derived by Babadi *et al.*,[28,29] which is a modification of Gay–Berne potentials[30] to better capture the long-distance convergence to all-atom force fields. This so-called RE^2 potential has a somewhat complicated form and is summarized in the Appendix. Describing the effective ellipsoids characterizing the bases involves several geometrical and energetic parameters. These are the three "half-diameters" for

each base along principal axes ($\sigma_x, \sigma_y, \sigma_z$), three corresponding "reciprocal well depths" for each base ($\epsilon_x, \epsilon_y, \epsilon_z$), a parameter σ_c characterizing the interaction range between atoms in the all-atom potential, and the Hamaker constant $A_{12} = 4\pi\epsilon_{LJ}\rho^2\sigma^6$ in Lennard-Jones (LJ) units [given by Eq. (2)], where ρ is the number density or reciprocal of the effective volume of each ellipsoid. The above 14 parameters are found for the ten possible base-base interactions (A-A, A-C, etc.) by best fit between the RE^2 functional form and an all-atom molecular mechanics (MM3) force field[31] with Buckingham exponential-6 potentials, following the parameterization procedure in Ref. 29. The resulting values are used in $V_{RE2}(\mathbf{B}_1, \mathbf{B}_2)$ in Eq. (1) and are summarized in Table SI in SM.[26]

Though the values in Table SI are essential to describe the potential, it is more useful intuitively to determine the half-diameters of the bases by taking two identical bases and aligning them along the three principal axes. Two identical bases are brought together along a principal axis until their RE^2 potential is equal to $1k_BT$ and $10k_BT$. The distances that result are measures of the energetically determined effective diameters of the ellipsoid representing the base and are tabulated for all residues in Tables SII and SIII in SM.[26] The distances determined in this manner from the RE^2 equipotentials should correlate with the size of the bases as determined by other independent measures such as the effective hydrodynamic radii of the bases, which can be extracted from all-atom simulation studies of the diffusion of a base in water. This comparison is described below and in the Appendix.

C. Base-sugar and base-phosphate van der Waals interactions

Bases may interact as well with phosphate or sugar residues according to so-called sphere-asphere potentials $V_{RE2}(\mathbf{B}_1, \text{res}_2)$, which are a limiting case of RE^2 interactions when one entity is spherical. The form of this potential is shown in Eqs. (A5). Computation of sphere-asphere potentials is more efficient than the full RE^2 potential. We take RE^2 parameters between the base in question and a sphere of the LJ radius σ. These parameters are given in Tables SIV and SV for the interactions of bases with phosphates and sugars, respectively. Base-phosphate potentials are derived from fits to the MM3 potential in the same manner as base-base interactions and result in well depths between about $0.2k_BT$ and $0.7k_BT$. Base-sugar interaction parameters are selected so that their corresponding potentials are nearly purely repulsive. This results in small values of the energy parameters and an interaction radius of 2 Å (see Tables SIV and SV).

D. Sugar-sugar and sugar-phosphate van der Waals interactions

Sugars may interact with nonlocal sugar and phosphate residues that happen to be in spatial proximity. We model their interaction with a LJ potential
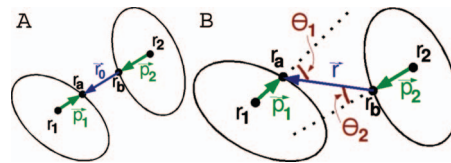

FIG. 2. Explanation of the hydrogen bond potential Eq. (3). (A) shows two bases aligned in the minimum energy configuration where $\mathbf{r} = \mathbf{r}_N$, and (B) shows the two bases in a higher energy configuration where $\mathbf{r} \neq \mathbf{r}_N$ and $\theta_1, \theta_2 \neq 0$.

$$V_{LJ}(r) = 4\epsilon_{LJ}[(\sigma/r)^{12} - (\sigma/r)^6] \qquad (2)$$

between residues i and j such that $j > i + 2$. We use a well depth of $\epsilon_{LJ} = 0.25$ kcal/mol and a distance of $\sigma = 2$ Å. This potential mainly serves to prevent steric overlap of backbone residues.

E. Base-base hydrogen bonds

Bases in DNA may pair by hydrogen bonding, an interaction not accounted for by the RE^2 potential which is based on atomic van der Waals interactions. We add this explicitly using a phenomenological potential with a form generalized from that used in all-atom studies,[32]

$$V_{HB}^{(ij)} = \epsilon_{HB}\left[5\left(\frac{r_N}{r}\right)^{12} - 6\left(\frac{r_N}{r}\right)^{10}\right](\cos^4(3\theta_i)\cos^4(3\phi_i)$$
$$+ \cos^4(3\theta_j)\cos^4(3\phi_j)). \qquad (3)$$

The hydrogen bond potential for a Watson–Crick pair can be interpreted as the sum of two potentials $V_{HB}^{(i)} + V_{HB}^{(j)}$, representing the hydrogen bonds from base $i \rightarrow j$ and base $j \rightarrow i$, respectively. The potential can be best understood from Fig. 2. $V_{HB}^{(i)}$ is a three-body potential between the center of mass of base i ($\mathbf{r}_i$), a point on the exterior of the ellipsoid considered to be the origin of the hydrogen bond ($\mathbf{r}_A$), and a point on the other base ($\mathbf{r}_B$) corresponding to either donor or acceptor. The hydrogen bonding in the other direction is similarly defined. We take $\mathbf{r}_A$ and $\mathbf{r}_B$ to be the unique points along the line segments joining $\mathbf{r}_1$ and $\mathbf{r}_2$, which intersect the surfaces of the two ellipsoids when the bases have the positions determined by the standard coordinates of the B isoform structure of DNA.[33] In Eq. (3), $r = |\mathbf{r}_A - \mathbf{r}_B|$ and $r_N = |\mathbf{r}_A^N - \mathbf{r}_B^N|$ in the B isoform structure [see Fig. 2(a)].

The angle θ in Eq. (3) is defined in Fig. 2(b). The angle ϕ in Eq. (3) captures the empirical fact that multiple hydrogen bonds between bases results in a rigidity to fluctuations that would prevent base 2 from rotating about $\mathbf{p}_2$, for example. Let $\mathbf{n}_1$ and $\mathbf{n}_2$ be the normals of bases i and j, respectively. The ϕ_i-dependent terms break the symmetry of rotations of base i about $\mathbf{p}_1 = \mathbf{r}_A - \mathbf{r}_1$ and rotations of j about $\mathbf{p}_2$. In Fig. 2(b), these normals are all pointing out of the page for simplicity. For the case of ϕ_i we project the normal $\mathbf{n}_2$ onto the plane perpendicular to $\mathbf{p}_1$, and ϕ will be the angle between this projected vector and $\mathbf{n}_1$: $\cos(\phi_i) = \mathbf{n}_1 \cdot (\mathbf{n}_2 - \mathbf{n}_2 \cdot \mathbf{p}_1)/|\mathbf{n}_2 - \mathbf{n}_2 \cdot \mathbf{p}_1|$. We perform the symmetric calculation ($i \leftrightarrow j$, $1 \leftrightarrow 2$) to get ϕ_j. Since the θ-dependent terms in Eq. (3) already provide rigidity against fluctuations such as those shown in Fig. 2(b) (right) as well as to buckling out of plane,

we found the above formulation superior to simply using $\mathbf{n}_1 \cdot \mathbf{n}_2$, which would "overcount" the potential cost of those fluctuations.

As in all-atom hydrogen bond potentials, raising the cosines in Eq. (3) to the power of 4 ensures that the potential is strongly orientation dependent. The factor of 3 inside the cosine is a result of the fact that we apply an angle cutoff of $\pi/6$ to both θ and ϕ. Therefore, when either angle reaches $\pi/6$, its contribution to the hydrogen bond potential vanishes. This also ensures a well depth ϵ_{HB} and avoids multiple minima in the cosine terms. The equilibrium separation r_N is chosen so that the minimum of the potential matches the correct equilibrium separation in the crystal structure.

Energy well depths must account for the fact that when a base pair hydrogen bond is broken, the two bases can form hydrogen bonds with the solvent. In implicit water, the well depth of the hydrogen bond represents the free energy difference between bonded and unbonded base pairs in solvent, which has been determined experimentally to be 1.2 ± 0.4 kcal/mol for each hydrogen bond.[34] Additionally, there is cooperativity in the energies of three hydrogen bonds, which makes G-C bonds more than 1.5 times stronger than A-T bonds. The stability of a base pair is also reduced by the presence of neighbors in the crystal structure, which induces strain on the base pair,[35] and reduces the value from that of an isolated base pair,[36] to stabilities of 10.2 and 17.2 kcal/mol for A-T and G-C pairs, respectively, *in vacuo*. The ratio between these two hydrogen bond well depths is 1.686, which when applied to the reasonable estimate of 2.4 kcal/mol for the hydrogen bond strength of A-T pairs in water gives an estimate of 4.0 kcal/mol for the hydrogen bond strength of G-C pairs. We therefore take 2.4 and 4.0 kcal/mol as our well depths ϵ_{HB} for A-T and G-C hydrogen pairings, respectively.

We allow hydrogen bonds between any pair of bases that satisfies the Watson–Crick pairing requirement. Due to the fact that our hydrogen bond potential is strongly orientation and position dependent, we found that it was very rare that a base shared a bond with more than one base at a time: on average, about 95% of the hydrogen bond energy was concentrated between the putatively bonded base pairs. The average base-base-stacking energy at typical temperatures and ionic concentrations is approximately -1.6 kcal/mol per pair of stacked bases. For the same conditions we find an average hydrogen bond energy per hybridized base pair of -3.2 kcal/mol. Because a double helix of length N has N base pairs and $2N-2$ stacking interactions, the total hydrogen bond energy and stacking are comparable in strength in our model.

F. Phosphate-phosphate electrostatic interactions

Phosphate atoms in DNA atoms carry a partial charge close to $-e$ in solution; in our coarse-grained model we thus assign a charge $-e$ on each P residue. These charged residues interact by a screened Coulomb potential in the presence of ionic solution, which is approximated here by Debye–Hückel theory. Ions are treated implicitly, and the potential between charged sites is given by
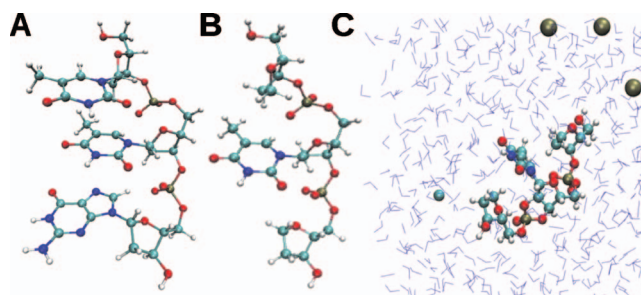

FIG. 3. Recipe for the determination of bond, angle, and dihedral potentials from all-atom simulations (see text).

$$V_{qq}(r) = \frac{q_i q_j e^{-r/\ell_D}}{4\pi\epsilon_o \epsilon r_{ij}}.$$ (4)

Here $\epsilon = 78$ is the dielectric constant of water, and $\ell_D = (\epsilon_o \epsilon k_B T / 2e^2 c)^{1/2}$ is the Debye length. ϵ_o is the permittivity of free space, and c is the ionic concentration. In water with 200 mM ionic concentration of KCl, the Debye length $\ell_D \approx 6.8$ Å.

G. Determination of local potentials by all-atom simulation

For all local potentials: bond, angle, and dihedral in Eq. (1), no functional form is assumed *a priori*. Instead the functional form and the potential parameters are extracted from thermal sampling of equilibrium configurations in all-atom simulations. For example, bond angle potentials up to an unimportant constant are given by $V_{eff}(\theta) = -k_B T \ln p(\theta)$, where $p(\theta)$ is the probability to observe a bond angle of θ. To obtain statistics that are undistorted by the presence of potential terms already accounted for in Eq. (1), we designed a modified system that had no base-base interactions and minimal Coulombic interactions between phosphates. Three consecutive bases are taken with center base A, T, G, or C as in Fig. 3(a). The end bases are then removed and replaced with a hydrogen atom with the correct bond length to each $C1'$ atom [Fig. 3(b)]. This removes any bias in the effective potentials that would have been due to base-base interactions, which are already included in Eq. (1). The molecule is then solvated in a box of water such that the water extends 8 Å beyond the boundary of the simulated molecule on all 6 sides, and K^+ and Cl^- ions are added at an average concentration of 200 mM [Fig. 3(c)]. In practice this involves adding of order 1 ions to the system. Charge neutrality biases the number of ions of each type: there are more positive charges present to balance the negative phosphate backbone. The resulting system is then simulated for $\sim$250 ns for each base using the NAMD simulation package[37] with CHARMM27 potential parameter set.[27] This long simulation time is required for the statistics gathered to converge within reasonable limits. Bond potentials converge to within 1% of their asymptotic value in 0.02 ns, single well angle potentials converge in 0.5 ns, double well angle potentials converge in 2.5 ns, and the dihedral potentials converge in 5 ns. The longer simulation time was to be certain of the convergence and was not prohibitive due to the very small size of the strand we were simulating. Dihedral potentials converge slower: their poten-

tials are shallower resulting in longer relaxation times, and a larger phase space (larger number of atoms) is involved.

We found that the potentials extracted from all-atom simulations often differed markedly from the commonly assumed phenomenological forms for such potentials between the same sets of atoms. In the supplemental materials section C, we show representative plots of the statistics-derived potentials for the case of bonds lengths (Fig. S1), bond angles (Fig. S2), and dihedral angles (Fig. S3). We use A, T, G, and C to represent the location of the center of mass of each base. The parameters for all these potentials are given in Tables SVII–SX in SM.[26]

We also found that correlations between the coordinates used for Boltzmann inversion were generally small but non-zero, with a rms value for cross-correlations of ≈ 0.19. Larger cross-correlations tended to be between overlapping degrees of freedom such as bond $P_{3'}S$ and angle $P_{3'}SP_{5'}$. We plot the correlation matrix in Fig. S4.

Almost all bonds show Gaussian statistics consistent with harmonic potentials: $V_{\text{bond}}(r) = \frac{1}{2}k_r(r - r_o)^2$; however the spring constants k_r vary considerably. The stiffest bonds were between the center of mass of the bases and the $C1'$ sugar residue. For this bond, pyrimidines, being the smaller bases, were stiffer than purines. The effective bond spring constant correlates well with that predicted by the phonon dispersion relation for a one-dimensional (1D) chain of coupled oscillators (Fig. 18). The bond potentials between base center of mass and $C1'$ are quite stiff, having natural frequencies of order 0.1 fs^{-1}. Explicitly including these would necessitate a prohibitively short time step in coarse-grained simulations. For this reason we treat the base-$C1'$ system as a single entity, i.e., the base ellipsoid plus $C1'$ residue are hard-coded as a rigid object with fixed internal geometry. Because the position of the neighboring sugar residue is always the same relative to the base, a force on the sugar induces the same force as well as a torque on the base ellipsoid.

There are 11 different bond angles in the coarse-grained model. Five of these have harmonic potentials, $V_{\text{angle}}(\theta) = \frac{1}{2}k_\theta(\theta - \theta_o)^2$, with stiffness coefficients k_θ ranging from about 3 to 19 kcal/mol·rad^2. The stiffness of $S_{5'}P_{5'}S$ and $SP_{3'}S_{3'}$ are equal within error bars and might differ only through end effects; hence we take the average. Six of the bond angles were much more readily fit by double well potentials,

$$V_{\text{angle}}(\theta) = -k_{\text{B}}T \ln\left(e^{-k_1(\theta - \theta_1)^2/2k_{\text{B}}T} + A e^{-k_2(\theta - \theta_2)^2/2k_{\text{B}}T}\right),$$
(5)

with minimum angles θ_1 and θ_2, effective stiffness constants k_1 and k_2, and dimensionless weighting factor A. While typical barrier heights for interconversion were about $1k_{\text{B}}T$, the largest barrier for interconversion between wells was about $4k_{\text{B}}T$ for $\theta_{TSP_{3'}}$. Similar to the bond potentials, the stiffness coefficients of the harmonic bond angles show a decreasing trend with mean distance between the participating atoms (Fig. 18).

The minima of the parameterized angle potentials compare well with the angles extracted from the atomic coordi-nates of the standard model of $(B, 10_1, 0.338)$-DNA derived from the crystal structure[33] (Fig. S5). In cases with double-welled angle potentials, the B-form standard model angles lie between the two parameterized wells, closer to the deeper one. Therefore, the parameterization taken from a piece of DNA too small to form anything like a double helix actually biases the angle potentials to a double helix quite well. An exception was backbone angles, which showed a significant discrepancy with those of the B-DNA structure. Other potentials such as stacking may strongly influence the equilibrium angle observed for larger strands of DNA, as noted by Tepper and Voth,[19] who set the equilibrium backbone angle to π in their model.

There are four types of dihedral potential in the model: using the notation X for any of the four bases, these are $S_{5'}P_{5'}SX$, $S_{5'}P_{5'}SP_{3'}$, $P_{5'}SP_{3'}S_{3'}$, and $XSP_{3'}S_{3'}$ [these coordinates can be seen in Fig. 1(a)]. The dihedral potentials for different bases generally have different parameters, which split the above four types into ten distinct dihedral potentials. The dihedral potentials are well fit by the functional form (see Table SIX in supporting materials[26])

$$V_{\text{dihedral}}(\phi) = \sum_{n=1}^{3} K_n(1 - \cos(n\phi - \phi_n)).$$
(6)

Barriers between local minima on the dihedral potential were all less than about $1.5k_{\text{B}}T$, indicating rotations with respect to these degrees of freedom are all quite facile. A comparison of the minima of the resulting dihedral angles with those extracted from the crystal structure of B-DNA can be seen in Fig. S6.

There is an improper angle between the normals to the planes defined by $P_{3'}SX$ and $P_{5'}SX$, where X represents the base. This angle is taken to be zero when the $P_{3'}$ residue coincides with $P_{5'}$, and π when $P_{3'}SXP_{5'}$ all lie in the same plane in the shape of a "Y." Generally the parameters entering into this potential were base-dependent (see Table SX) and well fit by Eq. (6). There was generally a preferred angle around 0.8π for the purines and $\sim 0.5\pi$ for the pyrimidines and a large barrier ($\sim 5k_{\text{B}}T$ for purines and $\sim 8k_{\text{B}}T$ for the pyrimidines) that inhibited full rotation.

We found that in the all-atom simulations bases were not free to rotate: interactions with the rest of the molecule biased the orientation of the effective ellipsoid to have a preferred angle. However these interactions are already counted in base-P or P-P interactions in Eq. (1), so to explicitly include such effects in an angle potential representing the orientation of the base would be redundant. Moreover, we observed that explicitly adding such a base-orientation potential derived from the all-atom statistics for three consecutive bases decorrelates the normal vectors of each base from its neighbors and thus competes with stacking order. We thus allowed bases to freely rotate about the $C1'$-base bond.

H. Langevin thermostat

We have adapted the molecular dynamics package LAM-MPS (Ref. 38) to simulate the coarse-grained DNA model. The equations of motion are integrated using a conventional velocity Verlet method for translational and rotational de-

grees of freedom with a timestep $\Delta t = 2-10$ fs depending on the degree of collapse in the system. To avoid the well-known singularities associated with Euler angles, a quaternion representation is used to describe the orientation of the ellipsoids.[39] Since we represent the DNA in implicit solvent, we employ a Langevin thermostat to maintain constant temperature. The thermostat adds forces $\mathbf{F}^{(L)}$ and torques $\boldsymbol{\tau}^{(L)}$ to the deterministic forces and torques arising from the interaction potentials. In the principal (body-centered) basis, these stochastic forces and torques are given by

$$F_i^{(L)} = -\zeta_i^{\mathrm{tr}} v_i + \xi_i^{\mathrm{tr}}, \tag{7a}$$

$$\tau_i^{(L)} = -\zeta_i^{\mathrm{rot}} \omega_i + \xi_i^{\mathrm{rot}}. \tag{7b}$$

Here v_i and ω_i denote the Cartesian components of translational and angular velocities, ζ_i^{tr} and ζ_i^{rot} are the eigenvalue coefficients of the translational and rotational friction tensors and ξ_i^{tr} and ξ_i^{rot} are components of white noise whose amplitudes are related to the friction coefficients through the fluctuation dissipation theorem.

In the body frame of the ellipsoid, the friction tensor ζ^{rot} is diagonal so that Eqs. (7a) and (7b) for the torques can be evaluated directly. The Langevin forces on the translational degrees of freedom in the fixed laboratory frame, however, depend on the orientation of the ellipsoid relative to the direction of motion. On time scales shorter than a typical rotation time, the diffusion of an ellipsoid is anisotropic but crosses over to an effective isotropic diffusion at longer times.[40] We found it easiest in calculating these forces to first project the velocity vector into the body frame of the ellipsoid (where the friction tensor is diagonal), evaluate Eqs. (7a) and (7b) in the body frame, and lastly rotate the resulting force vector $\mathbf{F}^{\mathrm{lang}}$ back into the frame of the simulation cell. The same basis transformation was performed in calculating torques in Eqs. (7a) and (7b).

The values of the translational and rotational friction coefficients in Eqs. (7a) and (7b) are taken directly from all-atom simulations of the diffusion of a single base solvated in a $(20\ \text{Å})^3$ box of water with 200 mM KCl. Separate 10 ns simulations are performed for each base. From these results, the effective radii of the ellipsoid representing the base can be found by comparing the observed friction coefficients with that predicted from continuum hydrodynamic theory. It is encouraging that these radii are comparable to the energetic equipotential radii at $10 k_{\mathrm{B}} T$ (Fig. 17). However it is worth noting that the radii extracted from continuum hydrodynamics tend to be smaller than the energetic radii most likely due to a breakdown of the no-slip condition at the interface between bases and water.

I. Energy scale for the effective simulation temperature

We express temperature in units of a natural energy scale ϵ in the system. We construct this energy scale by taking the minima of the RE^2 potential between identical bases given as V_{min} in Table SII and averaging this value over all bases A, C, G, and T and orientations x, y, and z (an average of the 12 energetic values in Table SII). This results in a value of $\epsilon = 1.45$ kcal/mol.

J. Sequence definitions

Sequences used in our simulations will be referred to with the following convention. We will use C_N-G_N to mean a strand of poly(C) of length N, which is hydrogen bonded to its complementary poly(G) strand of the same length. A_N-T_N is similarly defined. HET denotes several heterogeneous sequences defined in the $5' \rightarrow 3'$ direction as

$\mathrm{HET}_{\mathrm{SS},25} = \mathrm{CAGGATTAATGGCGCCTACCTTACC},$

$\mathrm{HET}_{\mathrm{SS},30} = \mathrm{CATCCTCGACAATCGGAACCAGGAAGCGCC},$

$\mathrm{HET}_{\mathrm{SS},60} = \mathrm{HET}_{\mathrm{SS},30}\text{-}\mathrm{CCGCAACTCTGCCGCGATCGGTGTTCGCCT}.$

Finally, $\mathrm{HET}_{\mathrm{DS},N}$ is a strand of $\mathrm{HET}_{\mathrm{SS},N}$ that is hydrogen bonded to its complementary strand.

III. RESULTS

A. Persistence length of ssDNA and dsDNA

The persistence length ℓ_p is a measure of the rigidity of a polymer and is given by the decay constant of the backbone unit tangent vector $\hat{\mathbf{t}}(s)$ as a function of base index or position s along the strand: $\langle \hat{\mathbf{t}}(s_0) \cdot \hat{\mathbf{t}}(s) \rangle = e^{-s/\ell_p}$, where $\langle \cdots \rangle$ is a thermal average. For single-stranded DNA (ssDNA), the tangent $\mathbf{t}$ was calculated by taking the vector from the sugar ($C1'$) residue on base i to the sugar on base $i+1$ and nor-malizing to unity. For double-stranded DNA (dsDNA), the tangent $\mathbf{t}$ was calculated by taking the vector from the midpoint of the sugar residues of hydrogen bonded base pairs at i to the midpoint of sugar residues of base pairs at $i+5$ and then normalizing to unity. The persistence length ℓ_p was then obtained by exponential fits of the data to the above correlation function. We found that for dsDNA, if the local principal axis tangent to the contour length of DNA (see Sec. III B) is used instead of the above recipe, the same persistence length is obtained to within 2%.

Without the stabilizing structure of the double helix, single-stranded DNA at ionic concentrations of $0.04M$ has a persistence length on the order 1 nm, which corresponds to
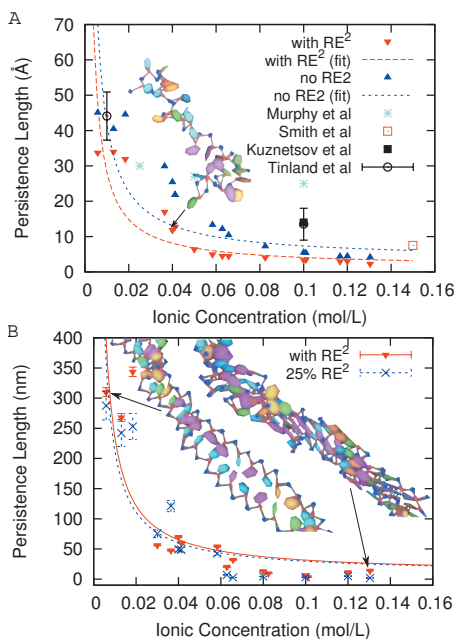
FIG. 4. (a) Persistence length vs ionic concentration of a random 25bp ssDNA sequence ($HET_{SS,25}$) at a temperature of 0.42ϵ simulated for 300 ns. Also shown are simulation results for ssDNA with the RE^2 potential turned off (blue triangles). The remaining data sets are from various experiments: Murphy *et al.* (Ref. 41), Smith *et al.* (Ref. 42), Kuznetsov *et al.* (Ref. 43), and Tinland *et al.* (Ref. 44). (b) Persistence length vs ionic concentration of a 60 bp dsDNA sequence ($HET_{DS,60}$) at a temperature of 0.10ϵ simulated for 600 ns. Results for the RE^2 potential scaled to 1/4 strength are also shown (blue X's). Dashed lines in both panels show the theoretical model of Nguyen *et al.* (Ref. 45) for the persistence length of a polyelectrolyte, whose functional form consists of an intrinsic persistence length plus a term inversely proportional to ionic concentration. Insets show representative snapshots taken from the simulations [rendering with BIOVEC (Ref. 46)].

two to three bases (the distance between successive base pairs is approximately 0.4 nm). ℓ_p is chiefly governed by the repulsive Coulomb interaction, which tends to straighten out the strand to maximize the distance between phosphate residues. As the interaction becomes more screened by the presence of ions in solution, ℓ_p will tend to decrease, as shown in Fig. 4. The functional dependence on concentration is captured by a constant bare persistence length added to a term inversely proportional to concentration.[45]

The values for ℓ_p found in our simulations are generally less than those observed experimentally, but it is worth noting that the experimental measurements themselves are highly variable. This can be understood from the persistence length being highly sensitive to other factors such as base sequence[47] and the experimental setup used to measure the persistence length, for example, fluorescence spectroscopy,[41] laser tweezers,[42] hairpin loops,[43] and gel electrophoresis.[44]

Figure 4(a) also shows that without the RE^2 potential, the stiffness of the single-stranded DNA *increases*. This occurs because the RE^2 potential is an attractive interaction, which tends to collapse the strand. At the temperatures in our simulation for Fig. 4(a), the stacking and hydrogen bonds are of the same order as k_BT. This temperature is above the dehybridization temperature in our model and also above the unstacking temperature for ssDNA described in more detail below. The consequence of this here is that base-base interactions tend to be nonlocal, involving nonconsecutive bases
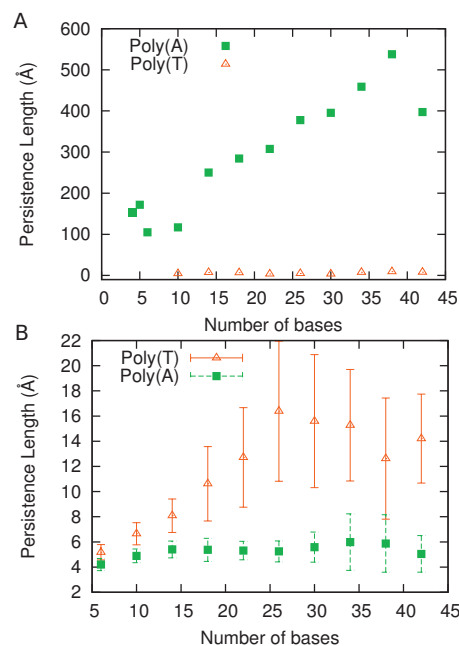

FIG. 5. Persistence length of ssDNA for poly(A) and poly(T) as a function of the number of bases in the polymer chain. Simulations were conducted at temperatures of (a) 0.03ϵ and (b) 0.42ϵ at an ionic concentration of 40 mM for 240 ns. At these temperatures and ionic concentrations the distance between stacked bases is about 4 Å. Note that the temperature in (b) is above the hybridization temperature for dsDNA.

in sequence, as can be seen from the snapshot in Fig. 4(a). Thus, removing base-base interactions [by setting the first term in Eq. (1) to zero] tends to stiffen the strand. On the other hand, at lower temperatures the opposite behavior is observed: stiffness does increase with increasing stacking interactions. We study this effect in detail below.

The double helix is inherently more stable to thermal fluctuations than a single strand, as two backbones wound around each other provide larger elastic modulus. As can be seen from Fig. 4(b), the persistence length of the double strand has the same functional dependence on ionic concentration as ssDNA but is roughly 55 times stiffer at $0.02M$, 25 times stiffer at $0.04M$ and 52 times stiffer at $0.13M$. The stiffness ratio from experimental measurements is ≈ 66.[42] Weakening the RE^2 potential has little effect on the persistence length. That is, due to the extra stability provided by double-stranded hybridization, the double helix shows no collapse on the scale of ~ 100 bp (see inset snapshots) so that weakening base-base interactions only modestly reduces the stiffness due to stacking. The effect can also be seen more evidently from the radius of gyration (Fig. S7).

The present model predicts a larger persistence length for a homogeneous single-strand of adenine (as large as ~ 50 bases or more) than the corresponding homogeneous strand of thymine bases ($\ell_p \approx 2$ bases) at low temperatures, as seen in Fig. 5(a). These results are consistent with the conclusions of Goddard *et al.*,[47] who found larger enthalpic costs for hairpin formation in poly(A) than in poly(T). However, at high temperatures the situation is reversed, and ss-poly(A) has a *smaller* ℓ_p (≈ 1.5 bases) than ss-poly(T) ($\ell_p \approx 4$ bases) [see Fig. 5(b)]. Adenine, being a purine, has a stronger RE^2 stacking interaction (see the z-minima in Table SII); however
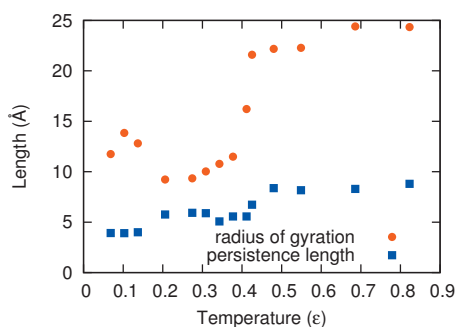
FIG. 6. Persistence length and radius of gyration vs temperature for ssDNA sequence $\text{HET}_{\text{SS,25}}$ at ionic concentration 0.04 mol/l. Simulations were run for 240 ns.

all *A-A* and *A-P* interactions are generally stronger than *T-T* and *T-P* interactions, and at high temperature this induces a greater degree of collapse due to non-local self interactions of the DNA strand. The persistence length shows an increasing trend with the length of the strand at the temperatures and ionic concentrations studied in Fig. 5. This is due to the exaggeration of end effects on shorter strands. The persistence length converged to its infinite length value for stands longer than about $7\ell_p$ at high temperature.

At moderate to high temperatures, we observe collapse of a heterogeneous strand as we lower the temperature, as seen in Fig. 6. We used the radius of gyration R_g to monitor the overall degree of collapse of the DNA, which is given by

$$R_g^2 = \left\langle \frac{1}{N}\sum_{k=1}^{N} (\mathbf{r}_k - \mathbf{r}_{\text{avg}})^2 \right\rangle,$$

where $\mathbf{r}_{\text{avg}} = N^{-1}\Sigma_{k=1}^{N}\mathbf{r}_k$ and the angle brackets represent the thermal average. Small values of R_g correspond to collapsed states. The general increase in persistence length and radius of gyration in the model with temperature contrasts with the temperature dependence of a wormlike chain ($\ell_p \sim T^{-1}$). At higher temperatures thermodynamic states with larger entropy have larger weight, the polymer expands, and the self-interactions that reduce the persistence length become less important. The collapse temperature where the radius of gyration suddenly increases is $\approx 0.4\epsilon$ (see Fig. 6). A similar trend toward collapsed states with increasing ionic concentration can be seen in Fig. S7.

The temperature dependence of persistence length and radius of gyration is not a simple monotonic function. Its complexity is seen by comparing Figs. 5 and 7. The RE^2 interaction in the adenine bases is strong enough to cause a well stacked configuration at low temperatures, whereas for thymine bases, the stacking interaction is too weak [Fig. 7(a)]. At higher temperatures, Fig. 7(b) shows that the stronger RE^2 interaction causes collapse, a fact confirmed in Fig. 4(a).

A slightly stronger or weaker van der Waals potential between bases results in DNA that is either collapsed or expanded, respectively. Our parametrized DNA is poised between an expanded and collapsed state (see Fig. 19), so the actual state of DNA would be highly sensitive to conditions affecting base-base interactions. *In vivo* mechanisms for modulating base-base interactions include DNA methylation

or potentially nucleosome post-translational modifications utilized for gene regulation such as histone phosphorylation, acetylation, or methylation.[48]

B. Twist and stacking of dsDNA

The twist is defined as the average angle that the backbone of the double helix rotates about the helical axis for each successive base pair. Figure 8(a) visualizes the calculation of the local twist at position i along the DNA. To obtain the helical axis at a given position along the double helix, we take the positions of the two sugars opposite the hydrogen-bonded bases at that position, as well as the sugar pairs up to three bases above and below that position. From the sugar coordinates, we compute the principal moments of inertia
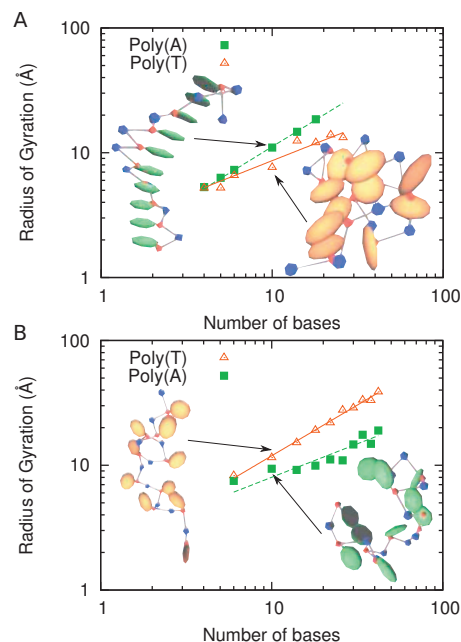

FIG. 7. Radius of gyration of ssDNA for poly(a) and poly(t) as a function of the number of bases obtained from the same set of simulations shown in Fig. 5 [i.e., $c=40$ mM, (a) $T=0.03\epsilon$ and (b) $T=0.42\epsilon$]. R_g exhibits a power law scaling with the number of bases with exponents of (a) 0.54 ± 0.08 for poly(t) and 0.84 ± 0.01 for poly(a), and (b) 0.80 ± 0.02 for poly(t) and 0.45 ± 0.07 for poly(a), respectively. Insets show representative snapshots taken from the simulations (Ref. 46).
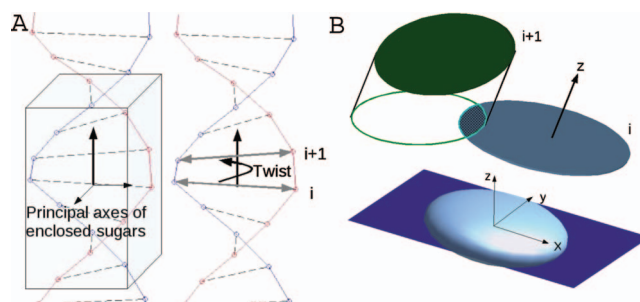

FIG. 8. (a) Calculation of local helical axis and twist (see text). (b) Calculation of the degree of stacking or stacking fraction. Principal axes are found for each ellipse, and a plane normal to the z axis is taken through the center of the base as shown in the bottom figure. This defines an ellipse for each base. The stacking is defined by projecting ellipse $i+1$ onto ellipse i and vice versa and measuring the overlap (see text). The top of (b) shows a visualization of this projection.
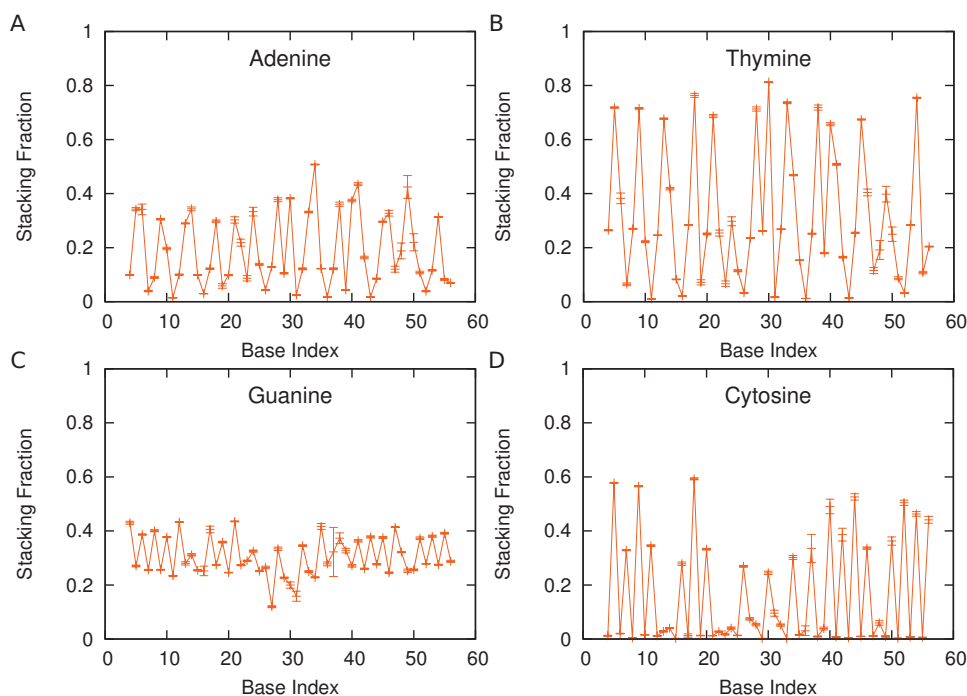
FIG. 9. Time averaged stacking fractions of homogeneous strands in a helix with their complementary strand. (a) and (b) are obtained from simulations of 60 base dsDNA consisting of a homogeneous strand of poly(A) hydrogen bonded to a similar strand of poly(T). The figures show the stacking of the adenine bases with themselves and the thymine bases with themselves, respectively. (c) and (d) are obtained similarly as in (a) and (b), except with guanine and cytosine bases. All simulations were conducted at $T=0.07\epsilon$ and an ionic concentration of 0.04 mol/l. The total simulation time is 800 ns. The average stacking fraction for each of these simulations is 0.1889 ± 0.0185 for adenine, 0.3178 ± 0.0343 for thymine, 0.3065 ± 0.0099 for guanine, and 0.1593 ± 0.0271 for cytosine.

and take the moment of least rotational inertia to be the helical axis. This method fails if the double helix persistence length drops to be on the order of three base pairs or if the helix dehybridizes. However, neither of these scenarios occurs for simulation parameters giving a stable double helix. The sugar-sugar vector rotates around the helical axis as one proceeds along the bases; the angle between the ith and $(i+1)$th sugar-sugar vectors is the *local* twist. This quantity is then averaged along the strand as well as over time. Snapshots during the simulation give the average twist reported in the figures below. We can similarly define the pitch of the DNA as pitch$=2\pi/$twist, i.e., the number of base pairs one must traverse for a full revolution of the helix. Note that the observed crystal structure value for the pitch of B-DNA is 10,[33] giving an expected twist of 36°.

We also develop an order parameter to define how well a given base is stacked to its neighbors. Taking the dot product of z principal axes of the ellipsoids has translational symmetry in the x-y plane of either ellipsoid and so does not capture the concept of stacking. The method we employ instead uses area projections depicted in Fig. 8(b). For each ellipsoid, we take the cross-section of the ellipsoid in the x-y plane of the own principal axes, which is an ellipse. To calculate the stacking fraction between bases i and $i+1$, we project cross-section $i+1$ down onto cross-section i. We then take the average of this value with the equivalent projection of i onto $i+1$ so that our definition of stacking is symmetric. We divide the projected area by the area that would be obtained from the B-DNA crystal structure (a stacking fraction of 0.6) to properly normalize. Because of twist, bases are not perfectly stacked in the crystal structure, so it is possible to see

stacking fractions greater than unity. This definition of stacking mathematically represents the intuitive notion of stacking very well, namely, the degree to which the flat parts of the two objects overlap their areas. The stacking fraction is a geometrical order parameter but correlates strongly with the RE^2 energies between neighboring bases on either of the two strands of the DNA (see Fig. S8). Thus stacking fraction accurately captures the base-base van der Waals stacking energies in addition to quantifying the structural features of DNA.

We found base stacking to be very heterogeneous, with traces of periodicity along the strand having a period of roughly four to five bases (see Fig. 9). Bases seem to stack well in small groups at the expense of poorer stacking in bases nearby along the strand. This results in kinks in the stacking structure of the strand, with the distance between kinks being only a few base pairs. An interesting trend is that the purines, the larger bases with greater stacking interactions, fluctuate far less and show more consistent stacking. On the other hand pyrimidines show larger extremes in their decimated stacking pattern, stacking intermittently more strongly than purines and completely unstacking. The stronger stacking interactions of the purines apparently induce convergence to the average value. The decimation pattern that we observe for the bases is due to frustration between stacking interactions and the other potentials (bond, angle, and dihedral), and reducing these other potentials resulted in less fluctuation and a larger average stacking (see Fig. 10). This situation is reminiscent of a Frenkel–Kontorova model where a competition between two incommensurate length scales corresponding to the equilibrium separation of a 1D
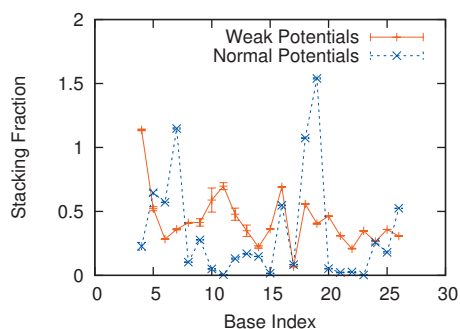
FIG. 10. Time averaged stacking fraction of the bases along a strand for both the all-atom parameterized potentials and another simulation with weakened parameters. Here the bond potentials were weakened by a factor of 0.5 and the angle potentials by 0.1, and the dihedral potentials were disabled. This indicates that the decimated stacking pattern we observe for bases is due to frustration between stacking interactions and the other potentials in the model (bond, angle, and dihedral). These simulations were done with $HET_{DS,30}$ at an ionic concentration of 0.04 mol/l and a temperature of 0.07ϵ.

chain of oscillators and a periodic underlying potential results in frustration-induced domains of oscillators.[49]

To understand which parts of the Hamiltonian control the large scale properties of twist and stacking of dsDNA, we applied global multiplicative scaling factors to the energies of individual potential classes, such as RE^2, bond, angles, dihedrals, and hydrogen bonds. Our simulations show that the potential most sensitively affecting the structure of the double helix is the base-base RE^2 interaction. Increasing the base-base RE^2 potential reduces the contour length of the helix (see Fig. S9), increases the stacking fraction, and tightens the twist over a large range of interaction strength.

Perhaps surprisingly, we found that the base-base attraction also appears to induce twist of the double helix. A picture that has emerged from computational models of DNA structure[50] is that twist results from the competition between electrostatic repulsion of phosphate groups and favorable base-stacking interactions, with stronger stacking interactions favoring alignment of the bases and thus putatively tending to straighten the helix. Indeed, the van-der-Waals-like RE^2 potential is minimized when the relative twist between stacked bases is zero. Figure 11(a) plots the helical twist as a function of RE^2 interaction strength, which modulates stacking interactions. This clearly shows an increase in twist with stacking interaction strength over a large range of the RE^2 interaction. These results were obtained by a zero temperature energy minimization starting from the expected crystal structure of B-DNA. Finite temperature reduces the overall values but does not change the trend. During the simulations, the helical twist relaxed to a degree determined by the overall strength of the RE^2 potentials. At the all-atom parameterized values the twist was 21.2°, whereas in the crystal structure it is 36.0°. Despite this quantitative discrepancy we did find that the potential function [Eq. (1)] reproduced a double-stranded structure with major and minor grooves (Fig. 12). The ratio of the sizes of the grooves minor to major in the coarse-grained model was 0.64, as compared to the experimental number of 0.54. Major and minor grooves persist so long as the twist of the DNA is $\gtrsim 10°$.

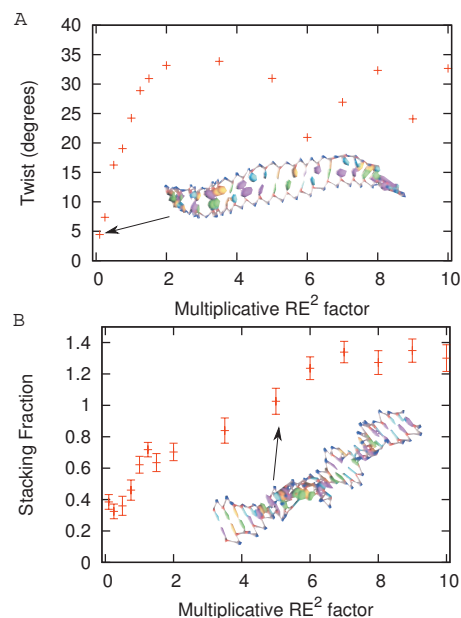The ladder conformation is preferred when the RE^2 po-

FIG. 11. Effect of scaling the RE^2 energies on (a) twist and (b) stacking of double-stranded DNA for a random sequence ($HET_{DS,30}$) at zero temperature and 40 mM ionic concentration. Simulations were started from the standard model (B, 10_1, 0.338)-DNA structure (Ref. 33) and allowed to equilibrate by energy minimization to structures having the values shown at $T=0$. The equilibration time was about 80 ns, implying shallow energetic gradients of collective modes. Insets show representative snapshots taken from the simulations (Ref. 46).

tential is scaled down to zero, where Coulombic interactions are competing solely with local bond, angle, and dihedral potentials and hydrogen bonds. The twisted conformations of the helix bring the phosphate residues closer together rather than farther apart (see Fig. 20). The direct compression of a ladder configuration would also increase the Coulomb energy in inverse proportion to the contour length, as well as frustrating local potentials such as angular potentials (see Fig. 13 and description below). To avoid this energetic cost as DNA is compressed, the system can lower its energy by structurally relaxing into a helical conformation. Upon helix formation, the potential energy of terms such as angles and Coulomb lower more than the RE^2 potential energy raises due to shearing the stacking pattern. Put another way, the stacking interactions do not favor the twisted conformation; they favor proximity of the bases: the total RE^2 energy in a
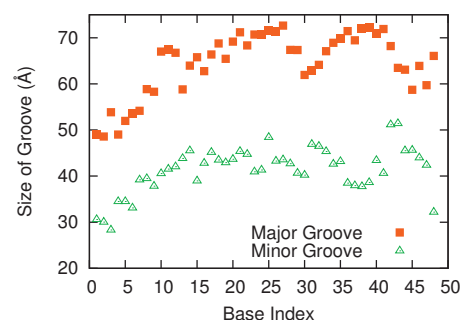

FIG. 12. The size of the major and minor grooves as a function of position along the double helix. This is a finite temperature simulation ($T=0.10\epsilon$) at ionic concentration 0.04 mol/l. The errors are of the order of the symbol size. The double helix makes about two turns over the length of the strand.
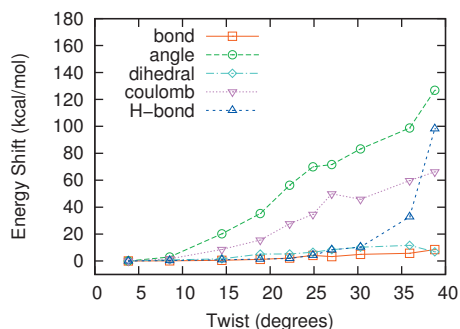
FIG. 13. Behavior of energy scales as functions of helical twist, as induced by increasing stacking interactions. Data plotted here are taken from the same simulations as in Fig. 11. Energies have been shifted to start from zero to depict the relative effects of increasing twist more clearly.

helical B-DNA conformation of $HET_{DS,30}$ is about -205 kcal/mol, while if the twist is set to zero from this conformation by forcing a ladder initial condition, the RE^2 energy is -270 kcal/mol before relaxation. However, the other potentials favor helix formation more so than maintaining the ladder and so break translational symmetry along the DNA contour when forced into proximity by the RE^2 potential. The correlation between RE^2 energy and stacking geometry (see Fig. S8) along with the competition between potentials in the system together imply that as the strength of the RE^2 interaction continues to increase, the system must eventually favor ladder-like configurations as bases are more properly stacked. This can be seen in Figs. 11(a) and 11(b) in the range of multiplicative RE^2 factor from 5 to 10. The twist begins to decrease (albeit with large scatter in the equilibration data) as the stacking fraction continues to monotonically increase above values present in the crystal structure.

The dominant interaction governing the stacking fraction is also the strength of the base-base RE^2 potential. There was significant relaxation of the B-DNA model structure at the all-atom parameterization values. The twist and stacking fraction obtain their B-DNA model values only when base-base interactions are magnified by a factor of 5, as can be seen from Fig. 11(b). This may indicate that cooperative many-body effects beyond the superposition of pairwise LJ interactions are governing the stacking interactions between bases. It is worth noting that large dynamical fluctuations in the DNA structure are also observed in all-atom simulations.[51]

Increasing the stacking energy (via the RE^2 potential) increases the twist, and this effect frustrates nearly all other interaction energies in the system. Figure 13 plots the other five contributions to the energy as a function of the induced helical twist due to increasing the RE^2 energy. The values of the twist were taken from the relationship in Fig. 11(a). We found RE^2 energy to be the optimal parameter to vary the helicity in the subsequent simulations. The increased twist of the double helix frustrates the Coulomb and angle potentials most significantly. As the twist of the helix increases, the energy of the angle potentials increases, indicating that the all-atom parameterized angle potentials favor a ladder-like configuration. Thus the base-base-stacking interactions, not
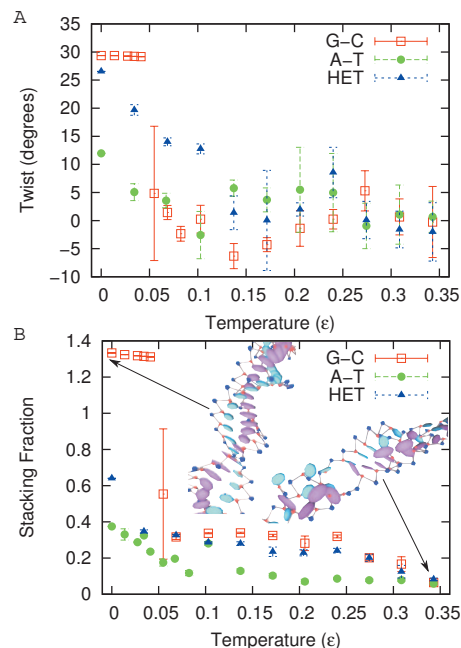

FIG. 14. Behavior of twist and stacking vs temperature. This is a simulation of sequences G_{60}-C_{60}, A_{60}-T_{60}, and $HET_{DS,60}$ simulated at an ionic concentration of 0.04 mol/l for 800 ns. Insets show representative snapshots taken from the simulations.

the angle potentials, govern the tendency to helix formation, a result in agreement with previous coarse-grained studies.[19]

C. Increasing temperature reduces twist and stacking in dsDNA in a sequence-dependent fashion

Figure 14(b) shows the twist as a function of temperature. The sequence G_{60}-C_{60} maintains a twist of approximately $29°$ until $T=0.05\epsilon$ at which point it drops sharply in a manner suggestive of a phase transition, accompanied by large fluctuations manifested by a sudden increase in the statistical error at $T=0.05\epsilon$. The helicity is approximately zero for higher temperatures. The stacking fraction for G_{60}-C_{60} shows similar behavior to the twist [Fig. 14(b)], with a phase transition around $T=0.05\epsilon$. This transition involves a reduction in order within the double-stranded structure and is distinct from dehybridization (the $HET_{DS,60}$ double strand was observed to dehybridize at a temperature around 0.31ϵ and took about 330 ns to dehybridize completely). Above the "untwisting" transition temperature, the stacking is still appreciable, although the twist does not appear to be significant. We found that A_{60}-T_{60} maintains very little twist in our simulations. At zero temperature it has a weak twist of $12°$. This sequence likewise showed the least stacking order among the sequences we studied. The sequence $HET_{DS,60}$ has much smoother behavior with changing temperature. At zero temperature, it has a twist of $26°$, slightly less than homogeneous G-C DNA, and intermediate stacking fraction between G-C and A-T. The phase transition behavior present in G_{60}-C_{60} is smoothed out for the heterogeneous sequence, yet it curiously maintains helicity until a higher temperature. Stacking is intermediate to G-C and A-T sequences at all temperatures and also has a broadened transition.
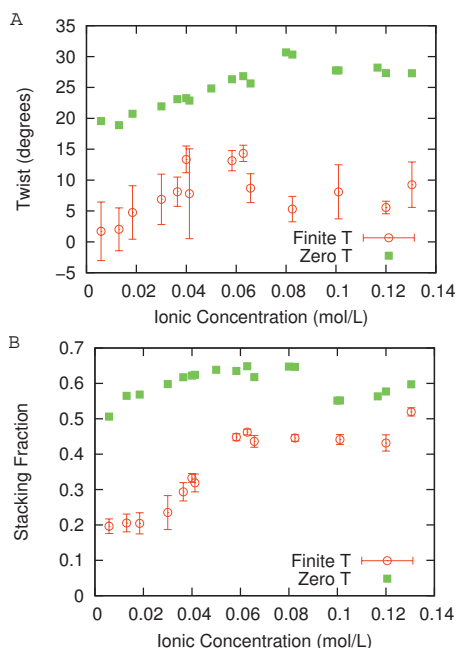
FIG. 15. Twist and stacking of double-stranded DNA vs ionic concentration. Thermal averages at finite T are taken from a 600 ns simulation of $HET_{DS,60}$ at $T=0.10\epsilon$. Zero temperature data are taken from an 80 ns simulation of $HET_{DS,60}$, and data from the equilibrated state are shown.

D. Coulomb interactions oppose both twist and stacking in dsDNA

Figure 15 shows the twist and stacking of the $HET_{DS,60}$ sequence as a function of ionic concentration. Increasing ionic concentration reduces the repulsive Coulomb energy by decreasing the Debye length in Eq. (4). As phosphate backbone charges are more effectively screened in stronger ionic solution, the RE^2 potential starts to dominate over the Coulomb potential, which reduces the contour length. To minimize frustration of the other potentials upon this compression, the helix twists to compensate. Thus an increase in ionic concentration increases the helical twist in an indirect way [see Fig. 15(a)]. The effect saturates as the ionic concentration is increased, whereupon other potentials such as angle and dihedral eventually dominate over the Coulomb potential.

Figure 15(b) shows that the stacking fraction increases with ionic concentration. This can be understood from the interpretation of Fig. 13, where it is seen that the Coulomb energy opposes stacking so that ameliorating it by increasing ionic concentration would increase stacking propensity. As ionic strength is increased, the DNA strand compresses due to the weakening of Coulomb repulsion. Thermal motion is less effective at eliminating stacking in this more compressed state due to the steric constraints manifested in the RE^2 potential. Finite and zero temperature simulations in Fig. 15 obey the same trend.

E. The model shows chiral preference for the right-handed helix

To address the question of whether the present physicochemical based model exhibits chirality, we performed finite temperature simulations of $HET_{DS,30}$ DNA starting from two
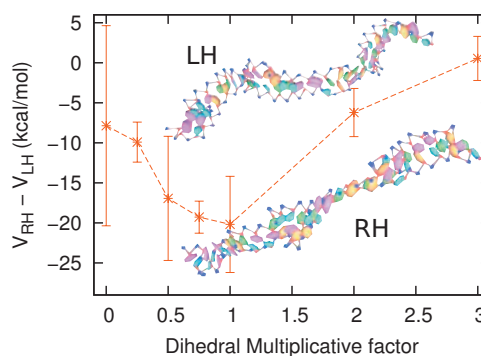

FIG. 16. Difference in the thermal average potential energy between right-handed and left-handed forms of dsDNA (sequence $HET_{DS,30}$), as a function of dihedral strength. Thermal averages are taken from 160 ns simulations at $T=0.03\epsilon$ and ionic concentration 40 mM. Inset images show representative snapshots from the simulations at dihedral multiplicative factor unity.

different initial conditions: one from the (right-handed) B-DNA standard structure and another from a structure with the azimuthal angle between successive base pairs reversed from $+36°$ to $-36°$. The latter gives an "anti-B-form" DNA with left-handed helix as the initial condition. Each configuration was observed to relax to a metastable equilibrium conformation for the duration of the simulation, analogous to either B-form or Z-form DNA.

The difference in the thermal average potential energy [Eq. (1)] between the right-handed and left-handed forms of DNA is plotted in Fig. 16 as a function of the strength of the dihedral potential. The strengths of the dihedral potentials are all simultaneously varied by adjusting an overall multiplicative factor. We chose to vary the dihedral potential because it is an obvious chiral term in the model's potential function. Specifically, the following dihedral potentials show bias toward the right-handed (B-DNA) helix (with the strength of the bias given in brackets in units of kcal/mol): $S_5'P_5'SP_3'$ (0.25), $P_5'SP_3'S_3'$ (0.06), $S_5'P_5'SA$ (0.40), $S_5'P_5'SC$ (0.18), and $S_5'P_5'SG$ (0.62). The potential $S_5'P_5'ST$ biases toward a left-handed (Z-DNA) helix but at a strength of only 0.03 kcal/mol. The remaining potentials are chiral symmetric.

The model shows an energetic bias toward the right-handed B-form of DNA, which at the temperature of the simulation was about 0.67 kcal/(mol·bp) for values of the dihedral potential obtained from all-atom parameterization. Interestingly, the chiral bias toward right-handed DNA was found to be maximal at this value of dihedral strength. Figure 16 also apparently shows a chiral preference even in the absence of dihedral potentials. This may be evidence that the origin of handedness in the model results from a coupling between potentials that individually are achiral, but collectively these potentials result in chiral constituents that when coupled together (e.g., by stacking interactions) yield a preference for the right-handed helix. We elaborate on this further in the discussion section, but note for now that the structure of the phosphate-sugar-base moieties, which constitute the building blocks of DNA, are themselves achiral objects by virtue of their absence of a center of inversion or mirror plane. We also note that at the parameterized values, most of the chiral bias does come from dihedral potentials.

Experimental measurements of the preference of B-DNA over Z-DNA give a free energetic difference of about 0.33 kcal/mol·bp.[52] The above energetic difference of 0.67 kcal/mol·bp in the computational model does not account for entropic differences between the B and Z forms. If one takes the magnitude of the value in the model seriously, it would imply that the Z form has larger entropy than the B form.

As is evident from the snapshots in the inset of Fig. 16, the stiffness of Z-DNA was seen to be softer than that of the B-form, with persistence length ℓ_p reduced by about a factor of 2 at the all-atom parameterized values. Electron microscopy measurements of chain flexibility in Z-DNA have shown moderate increases in persistence length ℓ_p of about 30%;[53] however these measurements were for poly(dG-dC) that had been adsorbed onto a two-dimensional surface, which could introduce new interactions modifying ℓ_p. Since a larger ℓ_p must correspond to stronger stabilizing interactions whose effect would be to increase the bending modulus, one would expect from energetic arguments that the thermodynamically less stable Z-form of DNA would have a shorter persistence length, as we observed in our simulations. On the other hand the increased number of base pairs per turn in the crystal structures implies a larger stacking fraction in Z-DNA; however only by about 10%. Moreover the rise per base pair is almost 50% larger in Z-DNA, implying reduced base-stacking interactions, which would also reduce the stiffness. Reconciling the computational and experimental observations of persistence length of Z-DNA is a topic for future study.

IV. CONCLUSIONS AND DISCUSSION

In this paper, we have introduced a coarse-grained model of DNA using rigid-body ellipsoids to model the stereochemistry of bases. This model captures the steric effects of base stacking, the stability of base pairing hydrogen bonds, screened Coulombic repulsion of the phosphate backbone, and nonlocal interactions between base-base and base-backbone, as well as backbone rigidity due to bond, angle, and dihedral potentials. Local effective potentials along the backbone are obtained from the statistics of all-atom simulations in explicit solvent. Base-base and base-backbone interactions are obtained from best fit between van der Waals interactions in an all-atom model and an anisotropic potential between effective ellipsoids.[29] Hydrogen bonds are modeled by adapting a functional form used in all-atom simulations to ellipsoidal bases, and phosphate-phosphate interactions are modeled through a mean-field screened Coulomb potential with Debye length dependent on ionic concentration.

The radii along the principal axes of the bases as defined by equipotentials of the anisotropic energy function correlate well with the base radii as determined from hydrodynamic diffusion measurements in all-atom simulations. The stiffness constants in the bond and angle potentials correlate well with those predicted by the phonon dispersion relation for a 1D chain of harmonic oscillators.

The model is physicochemistry-based and uses no structural information (i.e., no Gō potentials[54]) to provide bias toward the DNA crystal structure. The potentials result in a stable double-stranded helix with both major and minor grooves and a persistence length for single- and double-stranded DNA comparable to experimental values.

We have introduced quantitative recipes appropriate for the dynamical trajectories in molecular dynamics to quantify structural order parameters in the system such as degree of stacking and amount of twist. We investigated these structural properties along with other quantities such as persistence length, radius of gyration, and chirality for both single- and double-stranded DNA where appropriate, as various environmental factors such as temperature and ionic concentration were varied. We also investigated structural order in ssDNA and dsDNA as internal parameters such as number of bases, base sequence, and stacking strength were varied.

We find that at lower temperatures in the model, ss poly(A) stacks significantly more strongly than ss poly(T), consistent with the conclusions from the Libchaber group.[47] However, at higher temperatures another regime is reached where nonlocal interactions between bases govern the persistence length: poly(A) forms a collapsed globule with shorter persistence length than poly(T), which forms a more expanded globule. The scaling exponents for the radius of gyration with chain length likewise show inverse behaviors in these two temperature regimes. The persistence length of ssDNA initially decreases with increasing temperature in accord with the wormlike chain model; however at higher temperatures where nonlocal interactions become important, the persistence length shows an increasing trend over a large range of temperature, while the radius of gyration of the DNA globule expands through a collapse-transition temperature. In other words, below the crossover temperature, stacking interactions stiffen the chain and increase the persistence length, while above this temperature nonlocal base-base van der Waals interactions (which are inseparable from stacking interactions) soften the chain and decrease ℓ_p.

We also investigated the interplay of forces that results in twist for dsDNA. We find that under typical conditions, base-stacking interactions are the dominant factor in driving twist in spite of the fact that a ladder configuration would minimize base-stacking energy. Increasing the putative base-stacking strength frustrates all other interactions in the system as twist increases, indicating no other interaction could be responsible for inducing twist. However, the RE^2 potential responsible for base stacking is achiral and energetically minimized when bases are stacked directly on top of each other. We thus infer that base stacking enhances the chiral properties of the constituent components in DNA by bringing them in close proximity, resulting in increased twist. Moreover, when base interactions are sufficiently strong ($\sim 5-10X$ their putative value), bases eventually stack more directly on top of each other at the expense of twist. This results in nonmonotonic behavior of the twist as a function of stacking strength. Both twist and stacking increase as Coulomb interactions are more effectively screened. Even in the native, thermodynamically-stable structure, DNA is under stress and thus strained due to competition between the various potentials.

In our model, the structure in hybridized poly(G)-

poly(C) shows different temperature dependence than poly(A)-poly(T), with G_{60}-C_{60} showing much more order as temperature is raised, along with a sudden first-order-like drop in twist and stacking at a transition temperature below the dehybridization temperature. The untwisting with temperature is smoothed out for the heterogeneous sequence in a similar manner to disorder-induced broadening of a phase transition. It is notable how sensitive the qualitative behavior is to the sequence: A-T bases differ only mildly from their G-C counterparts in size and energy scales, yet these differences are enough to determine the presence or absence of critical behavior with respect to the order parameters of helical twist and base pair stacking. The effects of hydrogen bond strength on stacking and twist is a topic for future work.

Base stacking was analyzed at the level of resolution of individual bases, where stacking was found to exhibit an intermittent decimated pattern wherein roughly four to five bases stack well in groups at the expense of poor stacking in nearby bases. This quasiperiodic structure is reminiscent of systems frustrated by incommensurate length scales such as the Frenkel–Kontorova model. Consistent with this notion, decreasing the putative strength of bond, angle, and dihedral interactions in the model resulted in less stacking heterogeneity and an increase in the degree of stacking. The heterogeneous stacking pattern was observed over the total simulation time of about 1 μs. One would anticipate that over longer time scales the specific heterogeneous pattern would shift to other metastable configurations.

Including anisotropic van der Waals interactions through the RE^2 potential introduces a large number of parameters in the model so that the reduction in total number of parameters from that required in all-atom simulations is not dramatic. For each of the 10 base-base interactions, there are six radii, six well depths, an overall energy scale A_{12}, and an atomic length scale σ_c, a total of 140 parameters. Base-backbone potentials introduce another 80 parameters. Including masses, bonds, angles, dihedrals, screened electrostatics, Langevin coefficients, phosphate charge, and temperature gives a total of 382 parameters. To simulate the same system using all-atom potentials requires at least 600 parameters including van der Waals parameters, bonds, angles, dihedrals, and atomic properties such as mass, charge, and diffusion coefficient (this number can increase to thousands if accurate van der Waals potentials are sought).

On the other hand, the number of degrees of freedom in the coarse-grained model is substantially reduced. Each base-sugar-phosphate residue has only nine degrees of freedom in our model once rigid constraints are accounted for, whereas the same residue has approximately 100 degrees of freedom in the all-atom model if hydrogen atoms are treated as rigidly bonded. Implicit solvent can be present in both coarse-grained and all-atom scenarios. The tenfold reduction in the number of degrees of freedom allows the coarse-grained model to explore longer time scale phenomena than would be practically obtainable with all-atom simulations.

Base-base interactions had to be strengthened to reproduce the properties of the crystal structure. Since the putative strength of the base-base interactions resulted from direct all-atom parameterization, this implies that the functional form of the RE^2 potentials may not fully capture the electron correlations governing stacking interactions. These polarization effects likely induce a many-body cooperative component to the stacking interaction for coarse-grained bases, resulting in a much stiffer potential surface for local fluctuations around the native structure. Similarly, many-body interactions between coarse-grained residues in proteins are necessary to effectively capture protein folding rates and mechanisms.[55]

We found that the physicochemical based model showed energetic bias toward a right-handed form of dsDNA helix over a left-handed form. The more stable right-handed form had longer persistence length. While the dihedral potentials (as determined from the equilibrium sampling of a short piece of ssDNA) yield potentials that break chiral symmetry toward the right-handed helix, their role in determining chirality is sufficiently coupled to other interactions in the system that strengthening the dihedral potential alone does not enhance chirality. The chiral energetic bias was largest for the all-atom parameterized values of the dihedral strength, i.e., while decreasing the overall strength of the dihedral potentials diminished chiral preference in the model, increasing dihedral strength also did not enhance chirality but instead diminished it. This effect is likely due to an interplay between stacking and dihedral interactions. It appears that like twist, chirality is induced not so much by the direct inherent energetic preferences of potentials but by an indirect minimization of frustration induced by the forced compression of the system due to base-base-stacking interactions.

Put another way, preference for a right-handed helix over the left-handed helix for dsDNA in our model may arise predominantly from the stacking of chiral constituents: each phosphate-sugar-base-phosphate constituent is a nonplanar object that cannot be transformed by rotations and translations into its mirror image. Chirality is enforced by asymmetries in the equilibrium values of the bond and angle potentials, which determine the minimum energy structure of these molecular constituents, as well as chiral asymmetry in dihedral potentials. Specifically, the molecular building block $P_{3'}$-S-base-$P_{5'}$ has sugar residue at the chiral center, and there is no center of inversion or mirror plane. The subsequent building block is constrained to have its $P_{3'}$ residue at the position of the previous block's $P_{5'}$ residue and is coupled to the previous building block through base-stacking interactions. In this way, the handedness of DNA is induced by coupling these chiral objects together through stacking interactions, analogous to the mechanism behind the right-handed preference of α-helices due to hydrogen-bond-mediated coupling between chiral L-form amino acids.[56] Phospholipid-nucleoside conjugates have also been observed to exhibit spontaneous right-handed helix formation, with no helical preference present for the conjugate with nucleic acid bases removed.[57] This again reinforces the idea that helicity can be induced by coupling chiral constituents together through an achiral force. Such spontaneous formation of handed helical structures from chiral ingredients is also reminiscent of the mechanism by which chiral nematogens, inter-

acting through simple LJ-like potentials, form a cholesteric (twisted nematic) phase in a liquid crystal.[58] Exploration of the origins of DNA handedness including C-G sequence preference and nucleotide pairing as in Z-DNA, as well as more refined structural studies of the connection between atomistic and coarse-grained models in the context of chirality, are topics for future work.

ACKNOWLEDGMENTS

APPENDIX: POTENTIALS, HYDRODYNAMICS, AND FIGURES

1. The RE² potential between ellipsoids

The RE^2 potential, developed by Everaers and Ejtehadi,[28] is a generalization of the ubiquitous LJ interaction to ellipsoidal particles. The extra rotational degrees of freedom accounted for by the RE^2 potential come with a greater computational burden: the RE^2 potential needs five times the number of input parameters per interaction pair (including the cutoff distance) as does the LJ potential.

The first six of these parameters come from the radii of the two ellipsoids, expressed as the following shape tensor:

$$\mathbf{S}_i = \text{diag}(\sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)}). \tag{A1}$$

The anisotropic well depths are expressed as the relative potential well depth tensor. Its entries are dimensionless and are inversely proportional to the well depths in their respective directions,

$$\mathbf{E}_i = \text{diag}(\epsilon_x^{(i)}, \epsilon_y^{(i)}, \epsilon_z^{(i)}). \tag{A2}$$

Another input parameter with a dimension of distance is the atomic interaction radius, σ_c, which characterizes the distance scale for interactions between the atomic constituents to be coarse-grained. The energy scale is given by the Hamaker constant, A_{ij}. Finally, there is the necessary interaction cutoff distance, R_{cut}, as the computation time to compute all $O(N^2)$ RE^2 interactions is too prohibitive and is unnecessary because of how quickly the potential decays with separation.

The RE^2 potential between two ellipsoids, labeled as $i = 1, 2$, can be conveniently written in terms of attractive and repulsive components.[29] In the expressions below, the tensor $\mathbf{A}_i$ is the rotation matrix from the laboratory frame to the rotated principal axis frame of particle i,

$$V_A^{\text{RE}^2}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{r}_{12}) = -\frac{A_{12}}{36}\left(1 + 3\eta_{12}\chi_{12}\frac{\sigma_c}{h_{12}}\right)$$

$$\times \prod_{i=1}^{2}\prod_{e=x,y,z}\left(\frac{\sigma_e^{(i)}}{\sigma_e^{(i)} + h_{12}/2}\right), \tag{A3}$$

$$V_R^{\text{RE}^2}(\mathbf{A}_1, \mathbf{A}_2, \mathbf{r}_{12}) = \frac{A_{12}}{2025}\left(\frac{\sigma_c}{h_{12}}\right)^6\left(1 + \frac{45}{56}\eta_{12}\chi_{12}\frac{\sigma_c}{h_{12}}\right)$$

$$\times \prod_{i=1}^{2}\prod_{e=x,y,z}\left(\frac{\sigma_e^{(i)}}{\sigma_e^{(i)} + h_{12}/60^{1/3}}\right). \tag{A4}$$

The quantities η_{12}, χ_{12}, and h_{12} in Eqs. (A3) and (A4) are defined by

$$\eta_{12}(\mathbf{A}_1, \mathbf{A}_2) = \frac{\det[\mathbf{S}_1]/\sigma_1^2 + \det[\mathbf{S}_2]/\sigma_2^2}{[\det[\mathbf{H}_{12}]/(\sigma_1 + \sigma_2)]^{1/2}},$$

$$\mathbf{H}_{12}(\mathbf{A}_1, \mathbf{A}_2, \hat{\mathbf{r}}) = \frac{1}{\sigma_1}\mathbf{A}_1^T\mathbf{S}_1^2\mathbf{A}_1 + \frac{1}{\sigma_2}\mathbf{A}_2^T\mathbf{S}_2^2\mathbf{A}_2,$$

$$\sigma_i(\mathbf{A}_i, \hat{\mathbf{r}}_{12}) = (\hat{\mathbf{r}}_{12}^T\mathbf{A}_i^T\mathbf{S}_i^{-2}\mathbf{A}_i\hat{\mathbf{r}}_{12})^{-1/2},$$

$$\chi_{12} = 2\hat{\mathbf{r}}_{12}^T\mathbf{B}^{-1}\hat{\mathbf{r}}_{12},$$

$$\hat{\mathbf{r}}_{12} = \mathbf{r}_{12}/|\mathbf{r}_{12}|,$$

$$\mathbf{B} = \mathbf{A}_1^T\mathbf{E}_1\mathbf{A}_1 + \mathbf{A}_2^T\mathbf{E}_2\mathbf{A}_2,$$

$$h_{12} = |\hat{\mathbf{r}}_{12}| - \sigma_{12},$$

$$\sigma_{12} = \left[\tfrac{1}{2}\hat{\mathbf{r}}_{12}^T\mathbf{G}^{-1}\hat{\mathbf{r}}_{12}\right]^{-1/2},$$

$$\mathbf{G} = \mathbf{A}_1^T\mathbf{S}_1^2\mathbf{A}_1 + \mathbf{A}_2^T\mathbf{S}_2^2\mathbf{A}_2.$$

In LAMMPS, an RE^2-like interaction between an ellipsoid and a sphere may be specified by making the radii parameters of the sphere zero. The radius of the sphere is given by σ_c. In this case, the RE^2 interaction between the objects is calculated in the limit that $\mathbf{S}_2 \rightarrow \mathbf{0}$ and $A_{12} \rightarrow \infty$ at a rate of $A_{12} \sim 1/\det(\mathbf{S}_2)$. The relevant energy parameter for this interaction we will call $\tilde{A}_{12} \equiv A_{12}/\rho\sigma_c^3$, where ρ is the number density of the sphere. That is, $\frac{4}{3}\pi \det(\mathbf{S}_2)\rho = 1$. The potential may then be straightforwardly evaluated by substituting $\mathbf{A}_2 = \mathbf{I}$ and $\mathbf{S}_2 = \mathbf{0}$ into the RE^2 potential,

$$\tilde{V}_A(\mathbf{A}_1, \mathbf{I}, \mathbf{r}_{12}) = -\frac{\tilde{A}_{12}}{36}\frac{4\sigma_c^3}{3\pi}\frac{8}{h_{12}}\left(1 + 3\chi_{12}\frac{\sigma_c}{h_{12}}\right)$$

$$\times \prod_{e=x,y,z}\left(\frac{\sigma_e^{(1)}}{\sigma_e^{(1)} + h_{12}/2}\right), \tag{A5}$$

$$\tilde{V}_R(\mathbf{A}_1, \mathbf{I}, \mathbf{r}_{12}) = \frac{\tilde{A}_{12}}{2025}\frac{4\sigma_c^3}{3\pi}\frac{60}{h_{12}}\left(\frac{\sigma_c}{h_{12}}\right)^6$$

$$\times \left(1 + \frac{45}{56}\chi_{12}\frac{\sigma_c}{h_{12}}\right)$$

$$\times \prod_{e=x,y,z}\left(\frac{\sigma_e^{(1)}}{\sigma_e^{(1)} + h_{12}/60^{1/3}}\right). \tag{A6}$$

A more complete discussion of the RE^2 potential including its advantages over the alternative biaxial Gay–Berne potential[59] can be found in the literature.[29]

2. Energetic and hydrodynamic comparison

For our Langevin simulations, we directly use the friction eigenvalues extracted from all-atom simulations. However, it is instructive to compare the effective size of the bases both energetically and hydrodynamically.

The shape-dependent friction coefficients of ellipsoidal bodies can be obtained from the low Reynolds number solution to the Navier–Stokes hydrodynamics equations[60] and can be expressed in terms of the three radii r_i in the principal axes,

$$\zeta_i^{(t)} = \frac{16\pi\eta}{S+G_i}, \tag{A7a}$$

$$\zeta_i^{(r)} = \frac{16\pi\eta(r_j^2+r_k^2)}{3(G_j+G_k)}, \tag{A7b}$$

with the elliptical integrals

$$S = \int_0^\infty d\lambda[(r_1^2+\lambda)(r_2^2+\lambda)(r_3^2+\lambda)]^{-1/2}$$

and

$$G_i = r_i^2\int_0^\infty d\lambda(r_i^2+\lambda)^{-1}[(r_1^2+\lambda)(r_2^2+\lambda)(r_3^2+\lambda)]^{-1/2}.$$

To verify that the continuum hydrodynamic expressions for the friction tensor faithfully reproduce the diffusive motion of a base, we extracted diffusion coefficients from all-atom simulations using the CHARMM27 potential. The all-atom simulations were performed using the NAMD simulation package, with a single base solvated in a $(20\ \text{Å})^3$ box of water in a 0.2 mol/l neutral KCl solution. The simulations were performed for 10 ns in the *NPT* ensemble. Four such simulations were performed, one for each type of base. Temperature was maintained at 310 K by Langevin dynamics, and pressure was maintained at 1 atm.

Hydrodynamic radii are found by best fit of Eq. (A7b) to the observed effective friction coefficients for both rotation and translation in all-atom simulations. Tabulated values of friction coefficients and effective hydrodynamic radii are shown in Table SVI. These hydrodynamic radii may then be compared with energetic radii (for our Langevin simulations, we directly use the friction eigenvalues extracted from all-atom simulations rather than the parameters resulting from the best fit to the hydrodynamics of an ellipsoid).

A plot of hydrodynamic radii versus energetic radii is shown in Fig. 17. The two measures compare well; however the hydrodynamic size tends to be smaller than the energetic size. Either the parameterization scheme we employed resulted in RE^2 potentials that overestimated the range of repulsive forces or the smaller hydrodynamic values may be the result of the breakdown of the no-slip condition at the interface between bases and water, a reasonable scenario
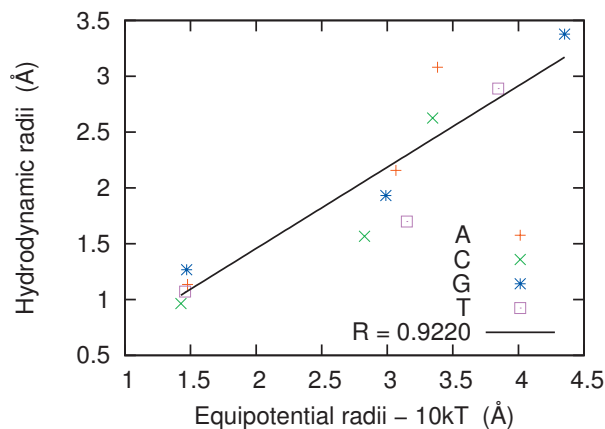

FIG. 17. Hydrodynamic vs energetic radii for effective ellipsoids. Hydrodynamic radii are found by optimizing the measured friction coefficients extracted from all-atom simulations of isolated bases and fitting to Eqs. (A7a) and (A7b) for the rotational and translational friction of an ellipsoid. The values are strongly correlated; however hydrodynamically derived radii tend to be smaller than energetically derived radii. For the thinnest axis, the assumption of continuum hydrodynamics is expected to be least accurate, and relative modifications due to hydration effects are expected to be largest (Refs. 61–65).

given that the size scale of the bases ($\sim$Å) tests the limits of the macroscopic assumptions in continuum hydrodynamics.[61–65]

3. Properties of DNA structure and bond potentials

Figure 18 illustrates properties of DNA bond potentials. Figures 19 and 20 illustrate properties of DNA structure as base-base interactions are varied (see discussion in text).
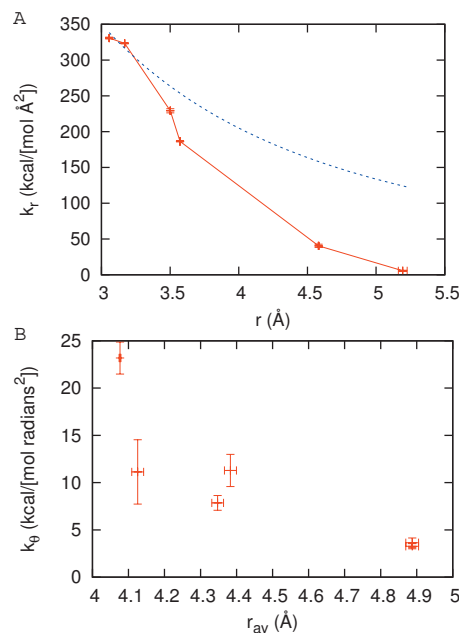

FIG. 18. (a) Stiffness coefficient k_r for bond potentials as a function of the mean distance between the participating residues, along with the phonon dispersion relation for a 1D chain of harmonic oscillators (Ref. 66), $k_{\text{eff}}(\lambda) = k_o \sin^2(\pi a/\lambda)$, with wavelength λ here taken to be the distance r. The correlation coefficient is $r=0.995$ and chance probability $p\sim4\times10^{-5}$. (b) Stiffness coefficient k_θ for angle potentials as a function of the mean length of the two bonds participating in the angle.
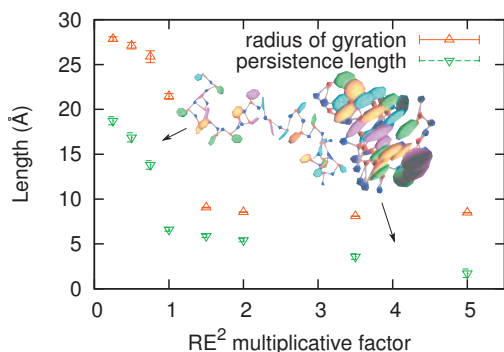
FIG. 19. Persistence length ℓ_p and radius of gyration R_g vs a multiplicative factor weighting the RE^2 energies for a ssDNA sequence $HET_{SS,25}$ at ionic concentration of 0.04 mol/l and temperature of 0.42ϵ simulated for 240 ns. Note that the parameters, resulting from fitting the RE^2 potential to all-atom potentials, result in DNA whose persistence length and collapse are highly sensitive to external conditions. Such a sensitivity could be exploited *in vivo* by either varying the ionic environment or through histone phosphorylation, acetylation, or methylation for the purposes of gene regulation. Insets show representative snapshots taken from the simulations.
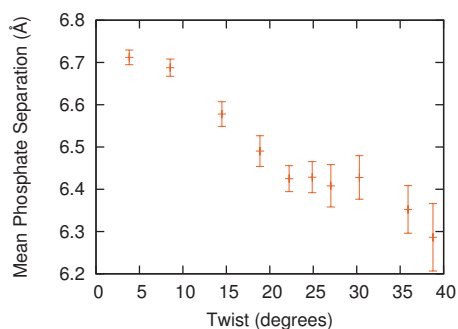

FIG. 20. Mean distance between phosphate groups as a function of helical twist. Data are taken from the same simulations as in Fig. 11.

[1] O. T. Avery, C. M. MacLeod, and M. McCarty, J. Exp. Med. **79**, 137 (1944).

[2] T. E. Cheatham III, Curr. Opin. Struct. Biol. **14**, 360 (2004).

[3] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer, Nat. Struct. Mol. Biol. **14**, 796 (2007).

[4] H. Yin, M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles, Science **270**, 1653 (1995).

[5] S. J. Greive and P. H. von Hippel, Nat. Rev. Mol. Cell Biol. **6**, 221 (2005).

[6] R. J. Davenport, G. J. L. Wuite, R. Landick, and C. Bustamante, Science **287**, 2497 (2000).

[7] K. M. Herbert, A. La Porta, B. Wong, R. Mooney, K. Neuman, R. Landick, and S. Block, Cell **125**, 1083 (2006).

[8] S. Mangenot, M. Hochrein, J. Radler, and L. Letellier, Curr. Biol. **15**, 430 (2005).

[9] P. Grayson, L. Han, T. Winther, and R. Phillips, Proc. Natl. Acad. Sci. U.S.A. **104**, 14652 (2007).

[10] D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, Nature (London) **413**, 748 (2001).

[11] G. Wagner, A. Bancaud, J.-P. Quivy, C. Clapier, G. Almouzni, and J.-L. Viovy, Biophys. J. **89**, 3647 (2005).

[12] E. Nalefski, E. Nebelitsky, J. Lloyd, and S. Gullans, Biochemistry **45**, 13794 (2006).

[13] A. Meller, L. Nivon, E. Brandin, J. Golovchenko, and D. Branton, Proc. Natl. Acad. Sci. U.S.A. **97**, 1079 (2000).

[14] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).

[15] S. Matysiak, A. Montesi, M. Pasquali, A. B. Kolomeisky, and C. Clementi, Phys. Rev. Lett. **96**, 118103 (2006).

[16] K. Luo, T. Ala-Nissila, S.-C. Ying, and A. Bhattacharya, J. Chem. Phys. **126**, 145101 (2007).

[17] F. Trovato and V. Tozzini, J. Phys. Chem. B **112**, 13197 (2008).

[18] M. Techera, L. L. Daemen, and E. W. Prohofsky, Phys. Rev. A **40**, 6636 (1989).

[19] H. L. Tepper and G. A. Voth, J. Chem. Phys. **122**, 124906 (2005).

[20] F. Zhang and M. A. Collins, Phys. Rev. E **52**, 4217 (1995).

[21] K. Drukker, G. Wu, and G. C. Schatz, J. Chem. Phys. **114**, 579 (2001).

[22] T. A. Knotts IV, N. Rathore, D. C. Schwartz, and J. J. de Pablo, J. Chem. Phys. **126**, 084901 (2007).

[23] Y. Ueda, H. Taketomi, and N. Gō, Biopolymers **17**, 1531 (1978).

[24] R. Luo, H. S. R. Gilson, M. J. Potter, and M. K. Gilson, Biophys. J. **80**, 140 (2001).

[25] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, Nucleic Acids Res. **34**, 564 (2006).

[26] See supplementary material at http://dx.doi.org/10.1063/1.3269994 for tables of parameters, plots of potentials, and supplementary figures.

[27] N. Foloppe, J. Alexander, and D. MacKerell, J. Comput. Chem. **21**, 86 (2000).

[28] R. Everaers and M. R. Ejtehadi, Phys Rev. E **67**, 041710 (2003).

[29] M. Babadi, R. Everaers, and M. R. Ejtehadi, J. Chem. Phys. **124**, 174708 (2006).

[30] J. G. Gay and B. J. Berne, J. Chem. Phys. **74**, 3316 (1981).

[31] J. H. Lii and N. L. Allinger, J. Am. Chem. Soc. **111**, 8576 (1989).

[32] D. B. Gordon, S. A. Marshall, and S. L. Mayo, Curr. Opin. Struct. Biol. **9**, 509 (1999).

[33] S. Arnott, P. J. Campbell-Smith, and R. Chandrasekaran, in *Atomic Coordinates and Molecular Conformations for DNA-DNA, RNA-RNA, and DNA-RNA Helices*, Handbook of Biochemistry and Molecular Biology, Vol. 2, 3rd ed., edited by G. D. Fasman (CRC, Cleveland, 1976), pp. 411–422.

[34] A. Fersht, Trends Biochem. Sci. **12**, 301 (1987).

[35] S. S. Mallajosyula, A. Datta, and S. K. Pati, Proceedings of the Sixth International Topical Conference on Optical Probes of Conjugated Polymers and Biosystems, Bangalore, India, January 4–8, 2005 [Synth. Met. **155**, 398 (2005)].

[36] Y. Mo, J. Mol. Model. **12**, 665 (2006).

[37] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).

[38] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[39] J. Kuipers, *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace, and Virtual Reality* (Princeton University, Princeton, NJ, 1999).

[40] Y. Han, A. M. Alsayed, M. Nobili, J. Zhang, T. C. Lubensky, and A. G. Yodh, Science **314**, 626 (2006).

[41] M. Murphy, I. Rasnik, W. Cheng, T. Lohman, and T. Ha, Biophys. J. **86**, 2530 (2004).

[42] S. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).

[43] S. Kuznetsov, Y. Shen, A. Benight, and A. Ansari, Biophys. J. **81**, 2864 (2001).

[44] B. Tinland, A. Pluen, J. Sturm, and G. Weill, Macromolecules **30**, 5763 (1997).

[45] T. T. Nguyen and B. I. Shklovskii, Phys. Rev. E **66**, 021801 (2002).

[46] E. Abrahamsson and S. Plotkin, J. Mol. Graphics Modell. **28**, 140 (2009).

[47] N. L. Goddard, G. Bonnet, O. Krichevsky, and A. Libchaber, Phys. Rev. Lett. **85**, 2400 (2000).

[48] S. Henikoff, Nat. Rev. Genet. **9**, 15 (2008).

[49] S. Aubry, Springer Ser. Solid-State Sci. **8**, 246 (1978).

[50] B. Mergell, M. R. Ejtehadi, and R. Everaers, Phys. Rev. E **68**, 021911 (2003).

[51] M. Levitt, Cold Spring Harb Symp. Quant Biol. **47**, 251 (1983).

[52] L. J. Peck and J. C. Wang, Proc. Natl. Acad. Sci. U.S.A. **80**, 6206 (1983).

[53] B. Revet, J. Malinge, E. Delain, M. Bret, and M. Leng, Nucleic Acids Res. **12**, 8349 (1984).

[54] Y. Ueda, H. Taketomi, and N. Gō, Int. J. Pept. Protein Res. **7**, 445 (1975).

[55] M. R. Ejtehadi, S. P. Avall, and S. S. Plotkin, Proc. Natl. Acad. Sci. U.S.A. **101**, 15088 (2004).

[56] C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland, New York, 1991).

[57] H. Yanagawa, Y. Ogawa, H. Furuta, and K. Tsuno, J. Am. Chem. Soc. **111**, 4567 (1989).

[58] P.-G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Oxford University Press, New York, 1993).

[59] R. Berardi, C. Fava, and C. Zannoni, Chem. Phys. Lett. **297**, 8 (1998).

[60] H. Lamb, *Hydrodynamics* (Dover, New York, 1932).

[61] B. Schneider and H. M. Berman, Biophys. J. **69**, 2661 (1995).

[62] A. H. Elcock and J. A. McCammon, J. Am. Chem. Soc. **117**, 10161 (1995).

[63] M. Feig and B. M. Pettitt, Biophys. J. **77**, 1769 (1999).

[64] J. Garcia de la Torre, M. L. Huertas, and B. Carrasco, Biophys. J. **78**, 719 (2000).

[65] B. Halle and M. Davidovic, Proc. Natl. Acad. Sci. U.S.A. **100**, 12135 (2003).

[66] N. W. Ashcroft and D. N. Mermin, *Solid State Physics* (Thomson Learning, Toronto, 1976).