Brownian dynamics simulations of sequence-dependent duplex denaturation in dynamically superhelical DNA

Steven P. Mielke[a)]
Biophysics Graduate Group, University of California, Davis, California 95616 and Biomedical Division, L-448 Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

Niels Grønbech-Jensen
Department of Applied Science, University of California, Davis, California 95616

V. V. Krishnan
Biomedical Division, L-448 Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, California 94551

William H. Fink
Department of Chemistry, University of California, Davis, California 95616

Craig J. Benham
Genome Center, University of California, Davis, California 95616

The topological state of DNA *in vivo* is dynamically regulated by a number of processes that involve interactions with bound proteins. In one such process, the tracking of RNA polymerase along the double helix during transcription, restriction of rotational motion of the polymerase and associated structures, generates waves of overtwist downstream and undertwist upstream from the site of transcription. The resulting superhelical stress is often sufficient to drive double-stranded DNA into a denatured state at locations such as promoters and origins of replication, where sequence-specific duplex opening is a prerequisite for biological function. In this way, transcription and other events that actively supercoil the DNA provide a mechanism for dynamically coupling genetic activity with regulatory and other cellular processes. Although computer modeling has provided insight into the equilibrium dynamics of DNA supercoiling, to date no model has appeared for simulating sequence-dependent DNA strand separation under the nonequilibrium conditions imposed by the dynamic introduction of torsional stress. Here, we introduce such a model and present results from an initial set of computer simulations in which the sequences of dynamically superhelical, 147 base pair DNA circles were systematically altered in order to probe the accuracy with which the model can predict location, extent, and time of stress-induced duplex denaturation. The results agree both with well-tested statistical mechanical calculations and with available experimental information. Additionally, we find that sites susceptible to denaturation show a propensity for localizing to supercoil apices, suggesting that base sequence determines locations of strand separation not only through the energetics of interstrand interactions, but also by influencing the geometry of supercoiling. © 2005 American Institute of Physics. [DOI: 10.1063/1.2038767]

I. INTRODUCTION

The processes by which double-stranded DNA (dsDNA) carries out its basic functions require both local transformations of its familiar double-helical structure and global deformations of its axis curve. It has become increasingly apparent that the molecule's topology, which couples these events, plays a fundamental role in gene regulation and other mechanisms of central importance to living organisms. For this reason, levels of superhelicity in nearly all eukaryotic, prokaryotic, and viral genomes are carefully regulated *in vivo*.

The topological state of a circular DNA molecule can be quantified by the expression,[1]

$$Lk = Tw + Wr. \qquad (1)$$

Here, Lk represents the linking number (the number of times one backbone strand "links through" the circle formed by the other), Tw represents the helical twist (the number of times either backbone winds around the helix axis), and Wr represents the writhe, or degree of supercoiling (the number of signed crossings of the helix axis in planar projection, averaged over all projection directions). Though Lk is strictly defined only for topologically closed domains, i.e., for closed circular DNA (ccDNA) or open DNA that is looped or anchored, it is, in practice, also a relevant descriptor of the topologies of open domains much longer than the persistence length (approximately 500 Å for B-DNA). In such regions,

---
[a)]Author to whom correspondence should be addressed. Electronic mail: smielke@lifshitz.ucdavis.edu

considerable superhelical stress can accrue, for instance upon torsional loading, due to the resistance of natural bends to translation in the viscous intracellular environment.[2] For a given molecule, the stress produced by deviations of Lk from its relaxed value, Lk_0, is accommodated by changes in Tw, Wr, or both:

$$\Delta Lk = Lk - Lk_0 = \Delta Tw + \Delta Wr. \qquad (2)$$

Here, ΔLk is the superhelicity, which is commonly maintained at negative values in nonthermophilic organisms and can be either quasistatic or fully dynamic; ΔTw corresponds to localized, sequence-dependent twist deformations, such as strand separation, cruciform extrusion, and B-to-Z transition, as well as to continuously distributed overtwist or undertwist; and ΔWr corresponds to bend (supercoiling) deformations, which are integral, for example, to site-specific recombination reactions.[3]

In superhelical stress-induced strand separation, or duplex denaturation, base-pair stacking and hydrogen bonding are disrupted, and the double helix locally unwinds to form two free single strands. Such sites are integral to replication, transcription, recombination, and repair events in prokaryotes, eukaryotes, and viruses.[4–7] For example, the initiation of replication in both prokaryotes and yeast has been shown to require the presence of a site at a precise position that is susceptible to superhelical strand separation.[8,9] Furthermore, local strand separation within promoter regions is a prerequisite for initiation of transcription, since RNA polymerase (RNAP) must gain access to the genetic information carried by the template strand. Although some RNAPs and other regulatory, single-strand-specific DNA-binding proteins actively participate in strand opening, others require a preexisting state of denaturation of the template for binding, and even those that contribute to opening may require at least partial initial destabilization at the binding site. That the initiation of the two central activities of DNA requires local opening of the DNA duplex clearly demonstrates the need for analytical methods that can predict the locations and occasions of strand separation.

The locations of stress-induced duplex destabilization (SIDD) within a torsionally constrained domain depend upon the identity of each base pair within the domain, since the state of each site is coupled to that of every other site by Eq. (1). A transition becomes favored when it relaxes the domain by an amount exceeding its energetic cost, with all base pairs competing to be sites of transition.

To date, only Benham and co-workers have provided a model that accounts for the global coupling of transition behaviors and successfully predicts locations and extents of SIDD as functions of base sequence and imposed superhelicity in long DNA molecules.[5–7] This model evaluates the statistical mechanical equilibrium distribution among states of denaturation and calculates the destabilization properties (such as transition probability and free energy required for base-pair opening) of every location within a topologically closed domain. This method, in conjunction with experimental work involving both prokaryotic and eukaryotic model systems, has implicated SIDD in the regulation of a wide variety of biological processes, including activation of the

*ilv*PG promoter of *E. coli* by integration host factor,[10,11] regulation of initiation of transcription from the *c-myc* oncogene in humans by binding of the far upstream element (FUSE)-binding protein (FBP) to the single-stranded FUSE element,[12] and transcriptional initiation from the CUP1 promoter in yeast.[13]

The success of the statistical mechanical model of Benham and co-workers notwithstanding, *in vivo* processes that alter DNA topology and drive such structural transitions as strand separation are inherently dynamic in character. A well-known example of such a process is transcription within a torsionally constrained domain, where rotation of the transcription ensemble (RNA polymerase, associated factors, and growing mRNA chain) about the helical DNA template is hindered by anchoring or drag forces within the cellular milieu, leading instead to rotation of the template about its axis as it is threaded through the polymerase. When this occurs within a region that is torsionally constrained, for instance because its ends are attached to a protein scaffold, the generated left-handed torque is sufficient to induce positive supercoiling downstream and negative supercoiling upstream from ongoing transcription.[14] When transcription occurs at only one site in plasmid DNA, the supercoils thus generated cancel, and there is no excessive net supercoiling, even when the transcription unit is well anchored.[15] However, when transcription occurs concurrently and divergently from two nearby sites in the same plasmid, or even in an open, linear region of DNA,[16] both act to negatively supercoil the intergenic region, often generating substantial superhelical stress. Such a mechanism has been implicated in a number of regulatory processes, such as coupled transcription in the *ilvYC* operon of *E. coli*, where it may coordinate expression levels during metabolic adjustments associated with growth state transitions.[17]

A theoretical approach for evaluating strand separation within the context of dynamically generated DNA superhelicity could potentially provide valuable new insights into the role of topology in basic DNA functions ranging from gene regulation to self-repair.[18] One such approach is dynamic computer simulation. However, scenarios involving dynamic stress-induced duplex denaturation have not to date been simulated, no doubt in part due to their complexity. Such simulations require the time evolution of hundreds or thousands of base pairs of dynamically superhelical, explicitly double-stranded model DNA under conditions approximating those of the physiological environment. An appropriate model must incorporate identity-specific information about base-pairing energetics, since it must capture the known global coupling of base pairs defining a region that is topologically constrained. Moreover, simulation times must extend out to tens of microseconds and beyond, since these are the time scales on which localized melting of DNA under denaturing conditions is known to occur. Collectively, these requirements present a formidable computational challenge.

Although Brownian dynamics algorithms capable of simulating the long-time supercoiling dynamics of kilobase-length DNA in solution have appeared,[19–21] they model DNA as a single chain of particles embodying the hydrodynamic and elasto-mechanical properties of dsDNA, not explicitly as

two interacting polynucleotide chains capable of separating within a biological context. Furthermore, so far such models have only been used to study equilibrium dynamics; that is, they simulate the dynamic approach to equilibrium conformations of molecules upon which static levels of superhelical stress are imposed, not the more biologically relevant, nonequilibrium situation in which superhelicity itself is fully dynamic. To our knowledge, only the model presented in Ref. 21 has attempted the latter. Motivated by applications in the modeling of double-helical DNA deformations, Moakher and Maddocks[22] have introduced a continuum mechanics formalism for modeling the elastic interaction of two Cosserat rods. However, this formalism represents only a first step toward a deterministic model of superhelicity-driven DNA deformations, and the appropriateness of continuum mechanics for modeling a fundamentally discrete system subjected to the stochastic intracellular environment remains an open question.[23] To our knowledge, no operational, explicitly two-stranded dynamic model for superhelical stress-induced DNA duplex denaturation has been presented.

In what follows, we introduce such a model and present results from simulations of localized, sequence-dependent strand separation in a 147 base pair (bp) region of DNA dynamically subjected to torsional stress. Section II describes the model and simulation method, placing particular emphasis on the interstrand interaction potential through which the model incorporates sequence-specific energetics, allowing it to capture accurately the stress-driven melting of double-helical DNA. Section III discusses results from several long-time simulations in which the base sequence of the model region is systematically altered in order to test the robustness of the method. It is shown that these results are in good agreement both with a successful statistical mechanical method for predicting locations of SIDD and with experiment. It is further shown that the method is capable of providing dynamical, mechanical, and geometrical information unavailable from the former. Section IV offers concluding remarks.

II. METHODS

A. Model

Two interwound polynucleotides are represented by segmented chains, each with N monomer subunits (refer to Fig. 1). The dimensions and relative orientation of the chains are set to approximate the hydrodynamic properties and geometry of closed circular, double-helical DNA. In particular, the subunits are assumed to be Stokes' spheres (beads) of radius R, whose centers initially coincide with the surface of a torus in Cartesian space according to the equations

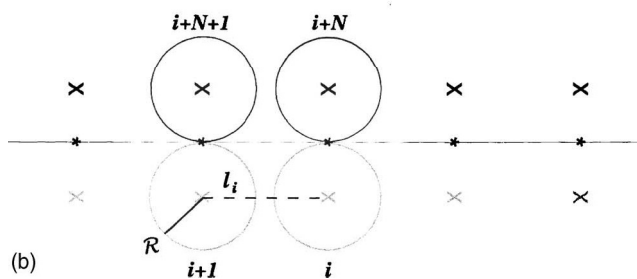$$x_{i+1,i+1+N} = (\alpha \sin(\gamma i \beta + \phi) + \varrho)\cos(i\beta), \tag{3}$$

$$y_{i+1,i+1+N} = (\alpha \sin(\gamma i \beta + \phi) + \varrho)\sin(i\beta), \tag{4}$$

$$z_{i+1,i+1+N} = \alpha \cos(\gamma i \beta + \phi), \tag{5}$$

where i runs from 0 to $N-1$, $\phi=0$ for one chain (hereafter denoted $\mathcal{C}_1$), and $\phi=\pi$ for the other (hereafter denoted $\mathcal{C}_2$).
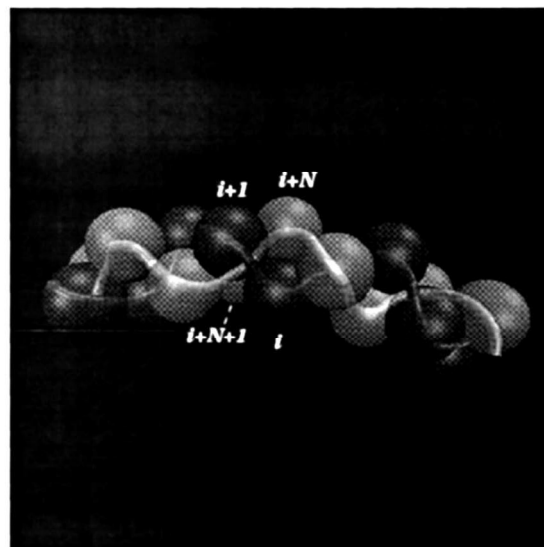
(a)

(b)

FIG. 1. Double-chain model. (a) 3D representation of the interwound bead chains. The particles comprising the ith duplex unit are labeled. (b) 2D schematic. Complementary beads (e.g., i and $i+N$) touch i.e., are separated by twice the bead radius R at mechanical equilibrium. The intrachain separation of particles i and $i+1$ is labeled as l_i. In the actual 3D structure, noncomplementary interchain beads (e.g., i and $i+N+1$) approximately touch at equilibrium.

The cross-sectional radius α is set equal to R, so that complementary beads on opposing chains touch at mechanical equilibrium. The winding number, $\gamma=n(N-1)/10.5$, corresponds to the number of double-helical turns [n base pairs per axis segment, $(N-1)$ axis segments, and 10.5 base pairs per turn in B-DNA]. The bending angle per axis segment is given by $\beta=2\pi/(N-1)$, and ϱ is the axial radius ($2\pi\varrho$ is the contour length of the molecule). In the simulations discussed below, $N=50$ and $n=3$, so that each of the 49 steps along the molecular axis spans three base pairs (10.2 Å in B-DNA), leading to a 147 base-pair structure that is approximately 500 Å in length—about one persistence length of B-DNA under physiological conditions. The bead radius is set to $R=\alpha=6.5$ Å, so that the interchain equilibrium separation of complementary beads, $2R$, equals 13.0 Å. Furthermore, the *intra*chain equilibrium separation of neighbor beads is set to 14.4 Å, and the interchain equilibrium separation of noncomplementary neighbor beads to 13.0 Å. Altogether, this means each duplex unit of contiguous bead pairs (beads i, $i+1$, $i+N$, and $i+N+1$), whose midpoints are initially separated by a distance of 10.2 Å (three base pairs), occupies a volume equivalent to that of a cylinder of the same length, and a diameter of 24 Å, which lies within the range of experimentally measured values of the hydrodynamic diameter

of B-DNA.[24,25] We note that the model does not resolve the major and minor grooves of B-DNA.

The total interaction potential of the $2N$-particle system is given by

$$U_{\text{tot}} = \sum_{i=1}^{N-1} U_s(l_i) + \sum_{i=1}^{N-2} U_b(\theta_i) + \sum_{i=N+1}^{2N-1} U_s(l_i) + \sum_{i=N+1}^{2N-2} U_b(\theta_i)$$
$$+ U_{\text{ic}}(r_{1,1+N+1}) + \sum_{i=2}^{N-1} [U_{\text{ic}}(r_{i,i+N-1}) + U_{\text{ic}}(r_{i,i+N+1})]$$
$$+ U_{\text{ic}}(r_{N,2N-1}) + \frac{1}{2} \sum_{j \neq i=1}^{2N} U_{\text{ev}}(r_{ij}). \tag{6}$$

The individual contributions to the total potential are

$$U_s(l_i) = \frac{K}{2}(l_i - l_0)^2, \tag{7}$$

$$U_b(\theta_i) = \frac{A}{2}(\theta_i - \theta_0)^2, \tag{8}$$

$$U_{\text{ic}}(r_{ij}) = -\varepsilon_{\text{ic}} \frac{\Delta G_{\text{den}}}{2} \left[\frac{1}{\exp(20((r_{ij}/\sigma_{\text{ic}}) - 1.9)) + 1} \right], \tag{9}$$

$$U_{\text{ev}}(r_{ij}) = \begin{cases} \varepsilon_{\text{ev}} \left[\left(\frac{\sigma_{\text{ev}}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{\text{ev}}}{r_{ij}} \right)^6 \right] + \frac{\varepsilon_{\text{ev}}}{4}, & r_{ij} < \sigma_{\text{ev}}\sqrt[6]{2} \\ 0, & r_{ij} \geq \sigma_{\text{ev}}\sqrt[6]{2}. \end{cases} \tag{10}$$

In Eq. (7), which expresses the potential energy associated with the linearly elastic stretching of intrachain segments, $l_i \equiv |\mathbf{r}_{i+1} - \mathbf{r}_i|$ is the length of segment i and l_0 is the equilibrium separation of intrachain particles. The stretching force constant is given by $K = k_B T/\delta^2$, where $k_B T$ is the thermal energy (the simulation temperature is set at 293 K) and δ corresponds to the fluctuation of the average segment length (we use the value $\delta = 0.025 l_0$ in accordance with previous work[21,26]). Equation (8) expresses the potential energy associated with the linearly elastic bending of the chain. The angle $\theta_i \equiv \cos^{-1}[((\mathbf{r}_{i+2} - \mathbf{r}_{i+1})/|\mathbf{r}_{i+2} - \mathbf{r}_{i+1}|) \cdot ((\mathbf{r}_{i+1} - \mathbf{r}_i)/|\mathbf{r}_{i+1} - \mathbf{r}_i|)]$ represents the deviation of segment $i+1$ from colinearity with segment i. Bending is assumed to be isotropic, with $\theta_0 = 0$, for all segments. The bending force constant is given by $A = k_B T(P/l_0)$,[24] where P is the persistence length of a single-stranded DNA [the value $P = 12.7$ Å (Ref. 27) is used in this work].

Stacking and hydrogen bonding of the two polynucleotide chains to form dsDNA are collectively represented by the interchain (ic) potential of Eq. (9), a pairwise potential for the interaction of hydrophobic particles in the low-density regime.[28] We choose this potential because the energetics of DNA double helix formation are dominated by the stacking of hydrophobic base pairs to minimize their exposure to the solvent.[27] For a given particle, say i in $\mathcal{C}_1$, the interchain potential is given by the sum of pairwise interactions with particles $(i+N)-1$ and $(i+N)+1$ in $\mathcal{C}_2$. Components of the force parallel to the helix axis are assumed to correspond to stacking forces, and those transverse to the

axis to H-bonding forces. Examination of Eq. (9) shows that $U_{\text{ic}}(r_{ij}) = -\varepsilon_{\text{ic}}(\Delta G_{\text{den}}/2)$ for $r_{ij} < 1.6\sigma_{\text{ic}}$ and $U_{\text{ic}}(r_{ij}) = 0$ for $r_{ij} > 2.2\sigma_{\text{ic}}$, where r_{ij} is the distance between particles i and j. The parameter σ_{ic} is determined by $1.6\sigma_{\text{ic}} = 13.0$ Å, the interchain equilibrium separation of noncomplementary neighbor beads (i.e., beads i and $i+N+1$, or i and $i+N-1$).

The factor ΔG_{den} in Eq. (9) is the free energy of local denaturation, as characterized by Benham[5] and Bauer and Benham.[6] This is the energy required to transfer base pairs having the torsionally unstressed winding of B-DNA to pairs of monomer units lacking hydrogen bonds and having no net interstrand twist. In the present context, the free energy of denaturation associated with each 3-bp duplex unit is given by

$$\Delta G_{\text{den}}(a,b) = a + \sum_{i=1}^{3} b_i, \tag{11}$$

where a is the free energy needed to nucleate the strand separation transition and b is the free energy needed to separate the ith base pair. One contribution to a is the additional free energy required to disrupt two stacking interactions when the initial base pair is separated; each additional base pair in a run of separation requires only one such disruption, so the nucleation free energy needs to be provided only once in a denatured region, even if that region spans more than one 3-bp segment (this will be discussed in greater detail below). For the free energy of nucleation we use the value $a = 10.2$ kcal/mol.[6] The incremental free energy of separating a given base pair is

$$b_i = \Delta H_i \left(1 - \frac{T}{T_{m_i}} \right). \tag{12}$$

Here, ΔH_i is the enthalpy of denaturation of the ith base pair, T_{m_i} is its transition temperature, and T is the simulation temperature, 293 K. For ΔH_i, we use the sequence-dependent values: $\Delta H_{\text{AT}} = 7.25$ kcal/mol for adenine-thymine (AT) (or TA) base pairs and $\Delta H_{\text{GC}} = 9.02$ kcal/mol for guanine-cytosine (GC) (or CG) base pairs.[5,6] The transition temperature in kelvins is known to depend on base sequence and monovalent cation concentration (relative to $1M$), x, according to

$$T_{m_i} = 354.55 + 16.6 \log(x) + 41 F_{\text{GC}}, \tag{13}$$

where F_{GC} is the fractional G+C content of the local sequence.[29,30] Then the transition temperature associated with denaturation of an AT (or GC) base pair is found by setting $F_{\text{GC}} = 0$ (or $F_{\text{GC}} = 1$). We assume a monovalent cation concentration of $x = 0.01M$. Inserting appropriate values of ΔH_i and T_{m_i} into Eq. (12), one obtains sequence-specific values of the free energy of separation, and, by inserting these into Eq. (11), the total free energy of denaturation characterizing a given 3-bp segment of the model duplex. In Eq. (9), the factor ΔG_{den} then provides the relative differences in energy necessary for simulating localized (to a 3-bp level of resolution), sequence-dependent transitions from a double- to single-stranded state, as described below.

Equation (10) is an expression for excluded volume potential energy, which prevents the overlap and passage of

both neighbor and non-neighbor intra- and interchain bead pairs, thereby preserving the topological constraint. The parameter σ_{ev} is determined by $2R = \sigma_{ev}\sqrt[6]{2}$, where $\sigma_{ev}\sqrt[6]{2}$ is the pairwise separation corresponding to the minimum of $U_{ev}(r_{ij})$ and R is the bead radius. We impose the cutoff, $U_{ev}(r_{ij}) = 0$ for $r_{ij} \geq \sigma_{ev}\sqrt[6]{2}$, so that the potential is strictly repulsive. For the model geometry, we set $\varepsilon_{ev}/\varepsilon_{ic} \sim 10^{-8}k_BT$ in order to engender a circular, double-helical structure that is both stable and torsionally relaxed on average out to late times when no external superhelical stress is imposed on the system.

B. Dynamically imposed superhelicity

In the simulations reported here, we dynamically impose a superhelical stress on the circular model structure in a manner analogous to that in which such stress is imposed on an intergenic region between two closely spaced, divergently oriented transcription units. In particular, the four particles at the ends of the duplex-particles 1, N, $N+1$, and $2N$-are set to rotate with a constant, right-handed angular velocity about the duplex axis endpoints, which remain fixed. The contiguous four particles-2, $N-1$, $N+2$, and $2N-1$-remain coupled to the rotating end particles via the potentials represented by Eqs. (7)–(10). Uncoupling of these particles during the course of simulations is prevented by forbidding the stress-induced melting of axis segments 1 and $N-1$; i.e., by defining particles 2, $N-1$, $N+2$, and $2N-1$ as the ends of the duplex, subjected to applied forces that continuously introduce negative superhelical stress. Then the portion of the duplex susceptible to strand separation runs from axis segment 2 to $N-2$, and consists of 141 base pairs when $N=50$ and each segment corresponds to three base pairs (has an equilibrium length of 10.2 Å). The value of the constant angular velocity is set at $\omega = 40\,434$ s^{-1} in order to generate substantial stress in a reasonable time. This produces 3.2 clockwise rotations of each end of the structure over the course of a 0.5-ms simulation and represents a rotational driving rate at least two orders of magnitude greater than that of which transcription is known to be capable.[2] Future applications of the model will encompass quasidynamic scenarios more reflective of the discontinuous activity of RNAP, and therefore characterized by a more realistic dynamic evolution of Lk.

C. Algorithm for dynamic simulations of localized, sequence-dependent strand separation

Time evolution of the $2N$-particle system is carried out using a Brownian dynamics algorithm, i.e., an algorithm based on numerical integration of the Langevin equations of motion in the overdamped limit. One scheme for such integration leads to the following expression for the evolution of particle positions:

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \frac{D_t}{k_BT}\mathbf{F}_i(t)\,\delta t + \mathbf{S}_i(t). \tag{14}$$

Here, $\mathbf{r}_i$ represents the position of particle i; D_t represents its translational diffusion coefficient (assumed to be identical for all beads and given by the Stokes-Einstein relation

$D_t = k_BT/6\pi\eta R$, where $\eta = 0.01$ P is the solvent viscosity and R is the bead radius); $\mathbf{F}_i$ represents the total deterministic force acting on particle i, derived from Eqs. (7)–(10); $\mathbf{S}_i$ represents a stochastic displacement, due to interaction with heat bath particles, and characterized by $\langle\mathbf{S}_i\rangle = 0$ and $\langle\mathbf{S}_i^2\rangle = 6D_t\delta t$;[31] and δt represents the simulation time step (the value $\delta t = 50$ ps is used in this work). We note that, because the connection between friction on continuum and atomic scales is ambiguous, there may be some discrepancy between simulated and physical times.

The numerical procedure is as follows. The initial positions of all particles are assigned using Eqs. (3)–(5). For each 3-bp chain segment, i, the values of $U_{ic}(r_{i,i+N+1})$ and $U_{ic}(r_{i+N,i+1})$ appropriate for the base sequence of that segment are assigned [see Eq. (11)]. Segments 1 and $N-1$ retain the value appropriate for a triplet chosen to be a combination of Gs and Cs in all cases examined. At each time step, new positions for particles 2 to $N-1$, and $N+2$ to $2N-1$, are calculated from the equations of motion using values of the forces calculated during the previous time step. Particles 1, $N+1$, N, and $2N$ are rotated according to their constant angular velocity about the fixed endpoints of the duplex, as described previously. For each 3-bp chain segment, i, from $i=2$ to $i=N-2$, if $U_{ic}(r_{i,i+N+1}) \neq 0$ and $U_{ic}(r_{i+N,i+1}) \neq 0$, the separations $r_{i,i+N+1}$ and $r_{i+N,i+1}$ are checked. If both $r_{i,i+N+1} > 2.2\sigma_{ic}$ and $r_{i+N,i+1} > 2.2\sigma_{ic}$, at which separation $\nabla U_{ic}(r_{ij}) = 0$, we fix $U_{ic}(r_{i,i+N+1}) = 0$ and $U_{ic}(r_{i+N,i+1}) = 0$, i.e., deem the segment strand separated for the remainder of the simulation. Otherwise, adjacent segments are checked. If either or both are strand separated, $U_{ic}(r_{i,i+N+1})$ and $U_{ic}(r_{i+N,i+1})$ are reassigned using

$$\Delta G_{den}(b) = \sum_{i=1}^{3} b_i; \tag{15}$$

if either segment $i-1$ or $i+1$ is already open, the nucleation free energy a is no longer required to open segment i. If neither segment $i-1$ nor $i+1$ is open, segment i retains its starting values of $U_{ic}(r_{ij})$. Using current information, new values of the forces are calculated, and this procedure is repeated until the desired simulation time is achieved.

In this algorithm, superhelical stress-induced strand separation is modeled as an irreversible, two-state process. Within open regions, i.e., for melted segments whose neighbors are also melted, the assumption of irreversibility on the time scales considered is justified by the fact that double-strand renaturation kinetics are known to be influenced significantly by such factors as friction-imposed limits on winding rates, the formation of competing intrastrand secondary structures, and complications associated with the mutual interpenetration of two single-strand coils prior to nucleation.[32] (This last point is particularly relevant to the present simulations, in which both ends of the duplex are continuously twisted in a right-handed sense to drive the unwinding of the chains.) However, at the ends of open regions, where a junction exists between strand-separated and duplexed secondary structures, one might expect renaturation to compete effectively with denaturation, thereby influencing the final extent

of melting. We are currently extending the model to represent strand separation as a reversible process.

D. Transition probability calculations

Transition profiles were generated using WEBSIDD, the Web-based program of Bi and Benham[33] (http://genomics.ucdavis.edu/benham/sidd/index.php) for predicting locations and extents of SIDD in double-stranded DNA. WEBSIDD accepts base sequence and superhelical stress level as input and produces the transition probability and destabilization energy of each base pair in the sequence. Additionally, parameters such as temperature and salt molarity can be specified. We chose WEBSIDD parameters to agree with our simulation parameters [$T=293$ K and $0.01M$ salt (the default value)] and used the moderate default superhelical stress level (superhelix density), $\sigma=(Lk-Lk_0)/Lk_0=-0.055$. G was entered for all G/C base pairs, and A for all A/T base pairs.

III. RESULTS AND DISCUSSION

In this section, we present results from four simulations in which base sequence was systematically altered in order to probe the ability of the method to capture the sequence dependence of the superhelical stress-induced melting of double-stranded DNA. Although the method is applicable to model structures of arbitrary sequences (of up to about a thousand base pairs in the present implementation), as an illustrative test set we have chosen four structures of nearly homogeneous base sequence, in which a three-segment (9-bp) A+T region was placed at three different locations in a model duplex, otherwise characterized by G+C content. Additionally, a simulation with an all-G+C structure was performed. This scheme was motivated by the fact that sites A+T-rich relative to an encompassing region are known to be susceptible to strand separation, and are therefore potentially of considerable biological relevance. For example, sites that promote initiation of transcription are typically high in A+T content. We note that G+C-rich sequences can potentially form Z-DNA, which requires an alternating purine-pyrimidine dinucleotide (GpC) repeat. However, this feature is not resolved under the current assumption of copolymer energetics (CGC duplex segments are identical to GGG segments, although base-pair-specific energetics can in principle be incorporated by using heteropolymeric values of the incremental free energy, b). We further note that all simulations were performed with identical initial configurations, and the same seed for random number generation, so that all differences in particle trajectories arise strictly from differences in base sequence.

Figure 2 shows four trajectory snapshots from a 0.5-ms simulation in which a circular, 141-bp model double helix was continuously subjected to superhelical stress by twisting its ends, as described in Sec. II. The region marked "τ" indicates particles that undergo rotations at constant angular velocity. The white regions indicate G+C content, i.e., duplex units assigned the value of ΔG_{den} corresponding to triplets consisting of arbitrary combinations of Cs and Gs, and gray regions indicate A+T content, duplex units assigned the value of ΔG_{den} corresponding to triplets consisting of arbi-
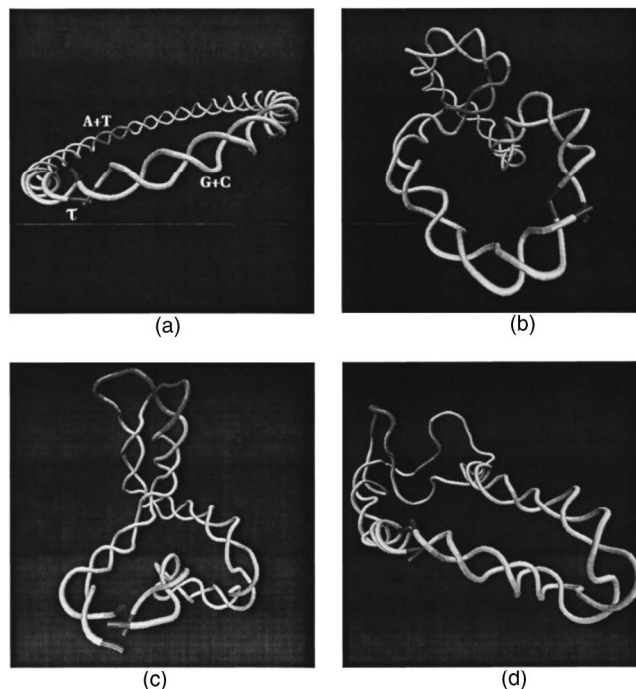

FIG. 2. Trajectory snapshots. The white regions indicate G+C content, gray regions A+T content (base pairs 43-51 in this 141-bp structure), and the region labeled "τ" locations undergoing rotation at constant angular velocity. (a) $t=0$: ringlike initial conformation. (b) $t=200$ μs: increasingly negative superhelicity resulting from continual rotation of both ends of the structure has produced left-handed, plectonemic interwinding. (c) $t=256$ μs: the A+T region, which has localized to the apex of a supercoil, has opened. (d) $t=368$ μs: additional duplex opening has occurred in G+C regions adjacent to the initial site of strand separation, unwinding right-handed twist, and relaxing left-handed supercoils.

trary combinations of As and Ts. For the simulation represented by Fig. 2, the A+T region spans segments 15, 16, and 17 (base pairs 43-51). Figure 2(a) shows the circular, torsionally relaxed initial structure. In Fig. 2(b), a snapshot taken at 200 μs, the structure has plectonemically supercoiled under the applied stress. In Fig. 2(c) (256 μs), the A+T region, which is seen to coincide with a supercoil apex, has melted [close inspection of the data (see Fig. 3) shows that, in fact, segment 18 (CG base pairs 52, 53, and 54) has also melted by this time]. In Fig. 2(d) (368 μs) it can be seen that several additional segments have separated, altogether 16 by this time, relaxing both right-handed helical twist and left-handed plectonemic interwinding.

Figure 3 compares melting profiles generated during our simulations [Figs. 3(a), 3(c), 3(e), and 3(g)] with transition profiles generated using WEBSIDD. Figure 3(a) shows the melting profile corresponding to the simulation of Fig. 2, in which duplex segments 15-17 are considered to be of A+T content. The ordinate displays the time in milliseconds and the abscissa displays the base-pair position. The white regions indicate locations that are unmelted at a given time under the criteria described in Sec. II, and the black regions indicate locations that are irreversibly melted. The gray bar demarcates the position of initial melting, A/T segment 17 (base pairs 49, 50, and 51), which strand separates at 248 μs. Following this event, the profile assumes the form expected for the cooperative melting of dsDNA under denaturing con-
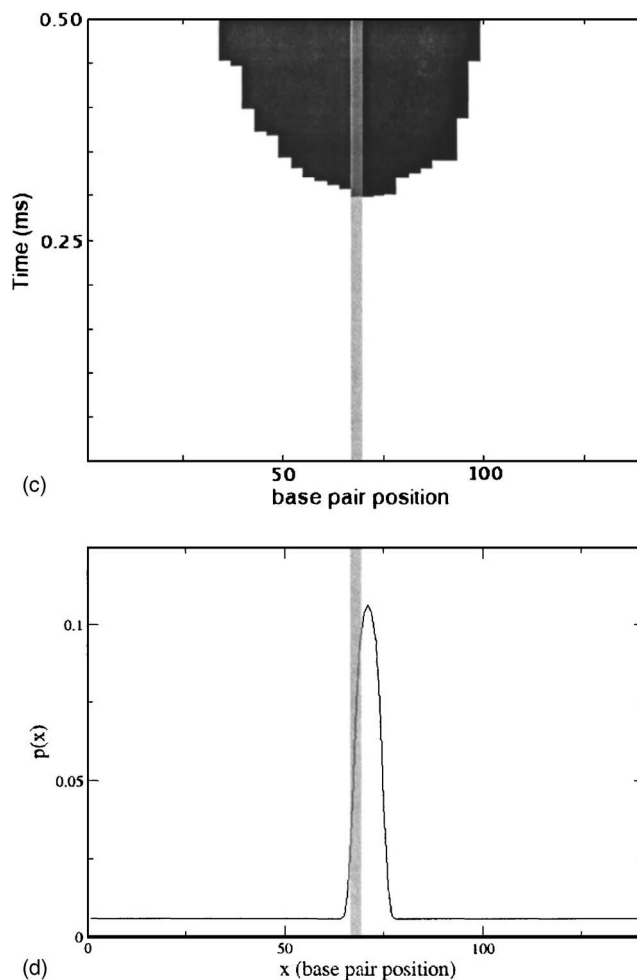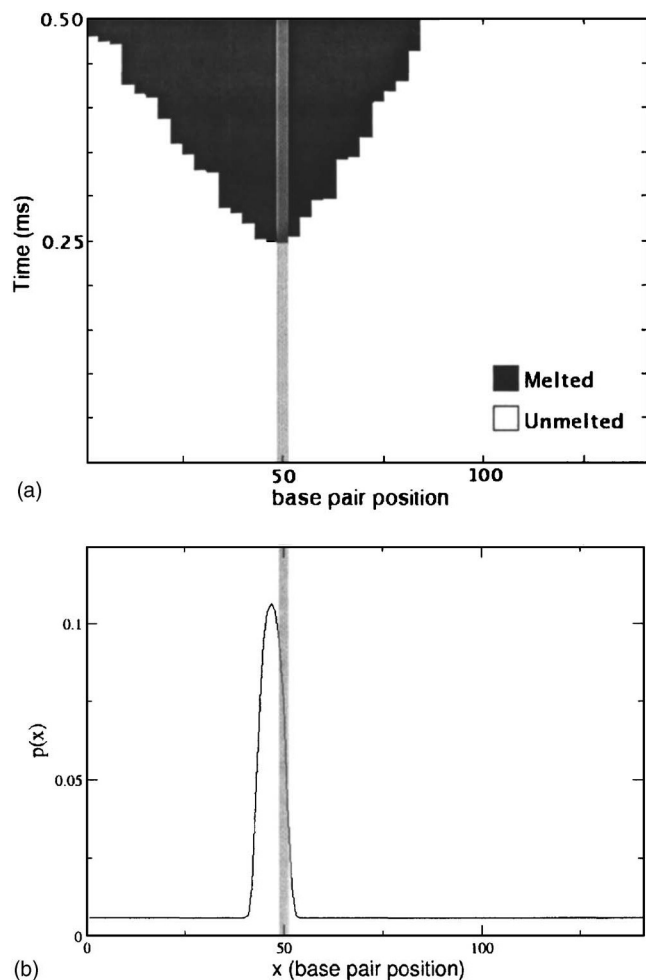
(a)

(b)

(c)

(d)

FIG. 3. Comparison of melting profiles with transition probabilities. (a), (c), (e), and (g) plot the time in milliseconds vs base pair position. The black areas indicate regions of the model structure that have irreversibly melted, and white areas indicate regions that are unmelted. The gray bar demarcates the 3-bp site of initial melting. (b), (d), (f), and (h) plot the probability of duplex opening (at superhelix density $\sigma = -0.055$) vs base pair position, calculated using WEBSIDD (see discussion). The gray bar again demarcates the site of initial melting. (a) and (b) correspond to a 141 base pair sequence in which base pairs 43-51 are of A+T content, and the remaining base pairs are of G+C content. In the sequence represented by (c) and (d), the A+T region spans base pairs 67-75, and in the sequence represented by (e) and (f) this region spans base pairs 94-102. The sequence represented by (g) and (h) is entirely of G+C content. For all sequences containing an A+T region, the site of initial strand separation coincides with the transition probability profile predicted by WEBSIDD. Comparison of (g) and (h) suggests that, in the case of all-G+C content, factors other than the base sequence, such as supercoil geometry, determine the location of the initial strand opening. The time of onset of melting in all dynamic simulations is ~100 μs after $\sigma = -0.055$ is reached.

FIG. 3. (*Continued*).

ditions, i.e., after disruption of segment 17, which requires breaking the two stacking interactions associated with the first base pair to open, contiguous segments open more rapidly at a lower threshold, until, by 0.5 ms, a total of 28 segments (84 base pairs or 8 turns of B-DNA) has opened. We note this value is larger than the expected upper limit on the extent of opening, 22.4 segments (67.2 base pairs corresponding to the 6.4 superhelical turns introduced at the ends of the model structure). This may be a consequence of the assumption of irreversibility, which does not allow renaturation processes to compete with melting, and thereby limit

the final extent of strand separation. For comparison, Fig. 3(b) shows the transition profile for the same sequence, calculated using WEBSIDD. The abscissa is again the base-pair location, and the ordinate is the transition probability, delineating those regions of the superhelical domain where duplex opening occurs with a significant probability. For this sequence, under a superhelix density of -0.055, the profile localizes at base pair 47, the midpoint of the A+T region, and has a maximum value of $p(x) = 0.11$. The gray bar again demarcates the initial position of melting within the model structure [cf. Fig. 3(a)] and is seen to lie within the transition profile predicted by WEBSIDD. Initial melting in the dynamic simulation occurs when $\sigma = -0.23$, about 200 μs after the value of -0.055 is reached, a time of the order of magnitude of the rate of nucleation of dsDNA melting under conditions of thermal or ionic denaturation based on oligonucleotide studies.[32] Over the 0.5-ms course of the simulation represented by Fig. 3(a), σ falls to a value of -0.47, as expected for 141 base pairs of B-DNA (10.5 base pairs per turn) underwound by 6.4 turns.

Figures 3(c) and 3(d) [3(e) and 3(f)] correspond to simulations in which the A+T region was moved to segments 23-25 (base pairs 67-75) [32-34 (base pairs 94-102)]. Locations and times of initial melting are A/T segments 23 (base pairs 67-69) [Fig. 3(c)] and 33 (base pairs 97-99) [Fig. 3(e)]
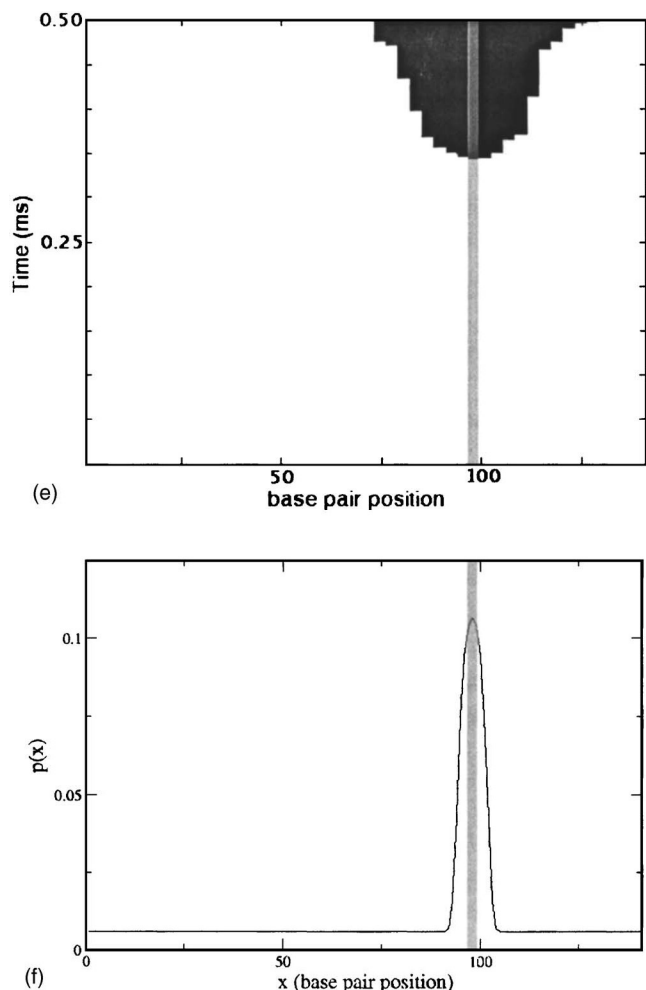
FIG. 3. (*Continued*).

FIG. 3. (*Continued*).

at about 300 and 344 μs, respectively. These events are again indicated by a gray bar in the figures. By 0.5 ms, 23 segments (69 base pairs) have opened in the simulation represented by Fig. 3(c), and 19 segments (57 base pairs) have opened in the simulation represented by Fig. 3(e). In both simulations, initial melting occurs when $\sigma \approx -0.3$, about 240 μs [Fig. 3(c)] and 284 μs [Fig. 3(e)] after a value representative of substantial superhelical stress (-0.055) is reached. Figures 3(d) and 3(f) show transition profiles at this superhelix density, calculated using WEBSIDD, corresponding to Figs. 3(c) and 3(e), respectively. Again it is seen, in each case, that the location of initial strand separation coincides with the region of the matching sequence predicted by WEBSIDD to have a significant probability of opening, with a maximum value, $p(x)=0.11$, at the midpoint of the region. In both simulations, a value $\sigma=-0.47$ is reached by 0.5 ms.

Figures 3(g) and 3(h) correspond to a simulation in which the model structure consisted entirely of G+C content. In this simulation, initial melting occurs at segment 31 (base pairs 91-93) at about 384 μs, the latest time of nucleation among the simulations presented. By 0.5 ms, a total of 21 segments (63 base pairs) has opened. After the superhelix density reaches a value $\sigma=-0.055$, about 324 μs elapse be-
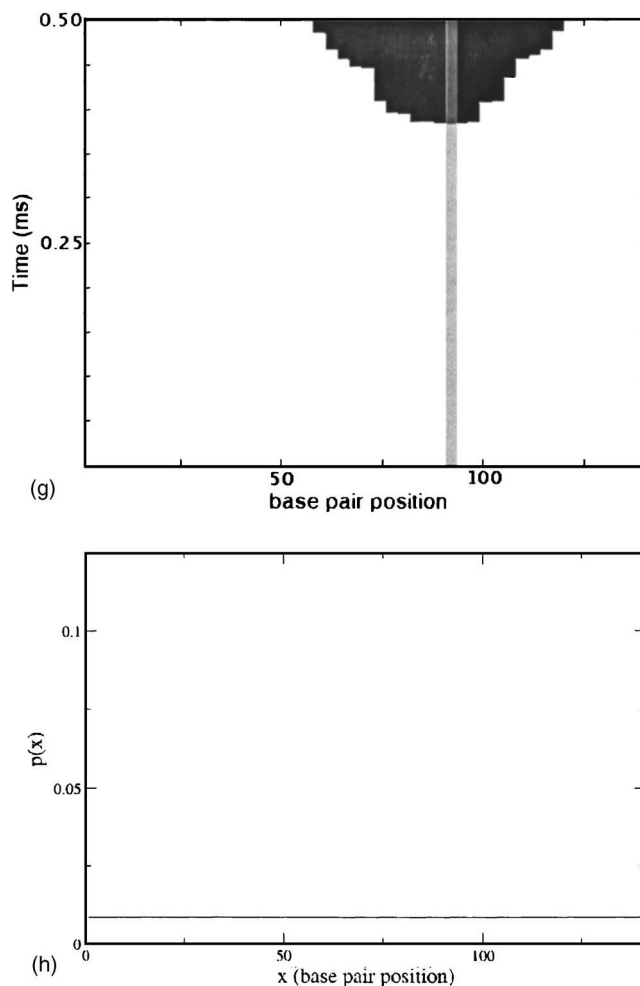
fore this melting event takes place, when $\sigma=-0.35$. Figure 3(h) shows a transition profile calculated by WEBSIDD for a base sequence of 141 guanine-cytosine pairs upon which a superhelix density $\sigma=-0.055$ has been imposed. It is immediately obvious from the figure that, as anticipated, no site within the homogeneous region is predicted to have a significant probability of opening; opening at every location is predicted to occur with the same probability, $p(x)=0.0085$. Although simulated melting appears to be delayed somewhat in the homogeneous model structure, it nevertheless occurs under the substantial imposed stress. This prompts a consideration of factors in addition to sequence that may determine locations of strand separation. To this end we point out that in all four simulations, at the time of initial strand opening, the first segment to open resides in the terminal loop of a supercoil, where bending of the duplex axis is most extreme, as opposed to a more linear region separating loops. Figure 4 shows a snapshot near the time of duplex opening from each simulation. Figure 4(a) reproduces Fig. 2(c). Figures 4(a), 4(b), and 4(c) correspond to the simulations represented by Figs. 3(a), 3(c), and 3(e), respectively, and Fig. 4(d) corresponds to the simulation represented by Fig. 3(g). It is evident that the first segments to open have localized within a loop in all four cases. In Fig. 4(d) the first segment to open is of G+C content; in the simulation repre-
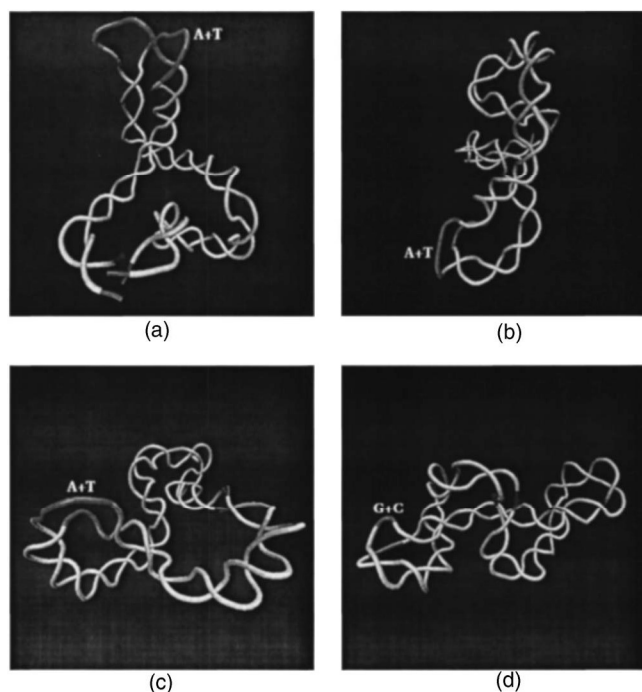
FIG. 4. Localization of denatured regions to loops. (a) [cf. Fig. 2(c)] is a snapshot taken from the simulation corresponding to the melting profile of Fig. 3(a) near the time of initial melting. Similarly, (b) corresponds to Fig. 3(c), (c) to Fig. 3(e), and (d) to Fig. 3(g). A+T sites are the first to open in structures in which they are present. In (d), the entire structure is of G+C content. In all simulations, initial duplex opening occurs in the terminal loop of a supercoil, suggesting both that the reduced effective bending stiffness associated with A+T regions organizes supercoil geometry and that this geometry influences the site of initial opening.

sented by Figs. 3(g) and 4(d), loop formation is a stochastic process that may influence the location of initial strand opening. In Figs. 4(a)–4(c) the first segments to open lie within the A+T site, suggesting the possibility that, under superhelical stress, the reduced bending stiffness of A+T sites, even when intrinsically straight and in the absence of sequence-dependent structural transitions,[34] increases the likelihood that such sites will reside in terminal loops of supercoils, where relatively high deformation energies likely contribute to strand opening, i.e., base sequence may determine locations of strand separation not only through the energetics of interstrand interactions, but also by influencing the geometry of supercoiling. Verification of this speculative interpretation of the results must await elucidation of the nontrivial relationship between sequence-dependent energetics and local bending stiffness in the model.

In order to quantify topological and geometrical changes, as well as track the superhelical stress level, during the simulations, we calculate approximate, time-averaged values of the linking number Lk, writhe Wr, twist Tw, and superhelical density σ using discretizations of the Gauss integral:[35]

$$G = \frac{1}{4\pi} \int_{C_1} \int_{C_2} \frac{\mathbf{e} \cdot (\mathbf{T}_2 \times \mathbf{T}_1)}{r^2} ds_1 ds_2. \qquad (16)$$

Here, in the case of Lk, C_1 and C_2 are the two backbone curves, T_1 and T_2 are unit vectors tangent to those curves at two points, P_1 and P_2, whose position vectors are $\mathbf{r}_1$ and $\mathbf{r}_2$,
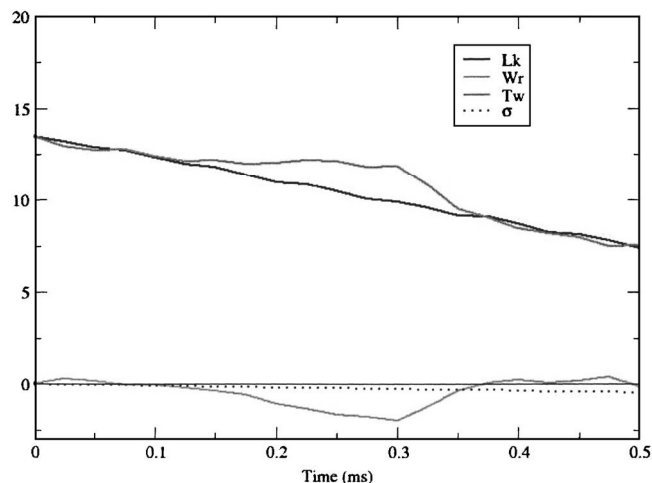

FIG. 5. Time dependence of topological and geometrical properties. Calculations of the linking number (Lk), writhe (Wr), twist (Tw), and superhelix density (σ) vs time (in milliseconds) correspond to the simulation represented by Figs. 3(c) and 4(b). At early times, continually decreasing Lk is partitioned nearly exclusively to Tw, and the initially planar duplex axis maintains $Wr \approx 0$. At just after 0.1 ms, a mechanical threshold is reached, after which Wr becomes increasingly negative (reaching a maximum value near −2.3) as the structure plectonemically supercoils, while Tw maintains an approximately constant average value. Then, at approximately 0.3 ms [cf. Fig. 3(c)], localized strand separation initiates, leading to the unwinding of about five turns ($\Delta Tw=-5.2$) and a relaxation to $Wr \approx 0$ ($\Delta Wr=-2$) by 0.5 ms. Over the course of the simulation, a value of the superhelix density σ =−0.47 is reached.

$\mathbf{e}=(\mathbf{r}_2-\mathbf{r}_1)/|\mathbf{r}_2-\mathbf{r}_1|=(\mathbf{r}_2-\mathbf{r}_1)/r$, and ds_1 and ds_2 are arc length elements on C_1 at P_1 and on C_2 at P_2. In the case of Wr, the double integral is over the single-axis curve of the molecule. Our procedure for calculating Lk, Wr, Tw, and σ is as follows. Once per 1000 time steps, Lk and Wr are calculated from discrete expressions based on Eq. (16). (Backbone elements are defined as the segments separating intrachain particles, and axis elements as the segments separating the midpoints of complementary interchain particles in both open and closed regions.) Then, Tw and σ are calculated from the expressions $Tw=Lk-Wr$ and $\sigma=(Lk-Lk_0)/Lk_0$, where $Lk_0 \approx 14$ is the value of Lk obtained from the initial ($t=0$) structure. The values of Lk, Wr, Tw, and σ thus obtained are summed as the simulation proceeds, and an average value of each is calculated once every 500 000 time steps (after every 25-μs interval). Figure 5 shows the results of these calculations for the simulation corresponding to Figs. 3(c) and 4(b), in which an A+T site resides at positions 23-25. During the first half of the simulation, the dynamic behavior of the driven system, as evidenced by Fig. 5, is in good agreement with previous results.[21] Specifically, at early times, Lk manifests principally as Tw, and $Wr \approx 0$, because the duplex axis remains planar on average. As undertwist propagates continuously through the structure, Tw decreases at the same rate as Lk, which decreases at a constant rate throughout the simulation, because the double helix is underwound at constant angular velocity. As the simulation proceeds, a mechanical threshold is reached, continual underwinding drives plectonemic supercoiling, and changes in Lk manifest principally as Wr, which becomes increasingly negative, while Tw maintains an approximately constant av-

erage value. Then, at about 0.3 ms, a sudden fall in values of Tw, and corresponding rise in values of Wr, clearly demarcates the onset of strand separation; unwinding of the duplex (a decrease in right-handed twist) at a rate exceeding the constant rate of change of Lk is accompanied by the relaxation of supercoils (an increase in left-handed writhe) as a consequence of the coupling embodied by Eq. (2). Over the remainder of the simulation, $\Delta Tw = -5.2$ and $\Delta Wr = 2.0$. As supercoils relax and values of Wr approach zero, changes in Lk are once again partitioned primarily to Tw. It can be seen from Fig. 5 that values of σ reach -0.47 by 0.5 ms, as anticipated.

IV. CONCLUSION

We have presented a method for performing dynamic simulations of the superhelical stress-induced denaturation of long (hundreds of base pairs and longer) DNA molecules in solution. Unlike existing discrete-chain models for the large-scale, long-time supercoiling dynamics of double-stranded DNA, in which the molecule is represented as a single polymer chain subjected to a static level of imposed torsional stress, our model represents the double helix as explicitly two-stranded, and imposes superhelical stress dynamically in a manner analogous to that in which it is imposed by real topology-altering processes, such as simultaneous transcription from two divergently-oriented promoters. Sequence dependence is incorporated via an interchain potential that accounts for relative, sequence-specific differences in the energy required for local denaturation and allows this process to be modeled as an irreversible transition from a double- to single-stranded state. We are currently extending the model to permit renaturation at junctions between open and double-stranded regions.

We have carried out an initial set of computer experiments in which the base sequence of the model system is systematically changed in order to probe the accuracy with which the method predicts locations, times, and extents of strand separation. From the results of these experiments, we conclude: (1) The model predicts sequence-dependent locations of strand separation in agreement with the statistical method of Benham and co-workers, which has been widely implemented in conjunction with experimental work to predict successfully locations and extents of SIDD and to elucidate a diverse array of regulatory mechanisms. Although the extents of strand separation predicted by the dynamic model are in reasonable agreement with the upper limit set by the amount of unwinding (6.4 turns), accommodation of renaturation would likely restrict the sizes of open regions and preclude excessive melting. (2) The model predicts times to initial duplex opening in order-of-magnitude agreement with experimentally determined rates of nucleation of DNA melting transitions, i.e., initial opening occurs ~ 100 μs after the imposed level of superhelicity becomes denaturing. (3) The model suggests the possibility that base sequence determines the locations of strand opening both by setting the energetics of interstrand interactions and by influencing the geometry of supercoil formation. Strand separation initiates at A+T-rich sites in all simulations in which these are present, as pre-

dicted by the statistical method, and in looped regions in all simulations; when present, A+T sites localize to looped regions prior to duplex opening. This illustrates the ability of the present model explicitly to capture mechanical and deterministic features of stress-induced local denaturation for which a strictly equilibrium statistical model cannot account. (4) The model captures anticipated topological and geometrical responses of the driven system, such as the accompaniment of rapidly decreasing local twist upon denaturation by a compensating relaxation of global writhe in order to maintain a constant rate of decrease of the linking number. Based on these successes, we have begun extending the method to novel, kilobase-length systems of heterogeneous base sequence and direct biological relevance.

[1] J. White, Am. J. Math. **91**, 693 (1969).
[2] P. Nelson, Proc. Natl. Acad. Sci. U.S.A. **96**, 14342 (1999).
[3] N. L. Craig, Annu. Rev. Genet. **22**, 77 (1988).
[4] C. J. Benham, Proc. Natl. Acad. Sci. U.S.A. **76**, 3870 (1979).
[5] C. J. Benham, J. Mol. Biol. **225**, 835 (1992).
[6] W. R. Bauer and C. J. Benham, J. Mol. Biol. **234**, 1184 (1993).
[7] C. J. Benham and C. Bi, J. Comput. Biol. **11**, 519 (2004).
[8] D. Kowalski and M. J. Eddy, EMBO J. **8**, 4335 (1989).
[9] R. Y. Huang and D. Kowalski, EMBO J. **12**, 4521 (1993).
[10] S. Sheridan, C. J. Benham, and G. W. Hatfield, J. Biol. Chem. **273**, 21298 (1998).
[11] S. Sheridan, C. J. Benham, and G. W. Hatfield, J. Biol. Chem. **274**, 8169 (1999).
[12] L. He, J. Liu, I. Collins, S. Sanford, B. O'Connell, C. J. Benham, and D. Levens, EMBO J. **19**, 1034 (2000).
[13] B. Leblanc, C. J. Benham, and D. Clark, Proc. Natl. Acad. Sci. U.S.A. **97**, 10745 (2000).
[14] L. F. Liu and J. C. Wang, Proc. Natl. Acad. Sci. U.S.A. **84**, 7024 (1987).
[15] V. A. Stupina and J. C. Wang, Proc. Natl. Acad. Sci. U.S.A. **101**, 8608 (2004).
[16] F. Kouzine, J. Liu, S. Sanford, H. J. Chung, and D. Levens, Nature Structural & Molecular Biology **11**, 1092 (2004).
[17] M. L. Opel and G. W. Hatfield, Mol. Microbiol. **39**, 191 (2001).
[18] L. Grossman and S. Thiagalingam, J. Biol. Chem. **268**, 16871 (1993).
[19] K. Klenin, H. Merlitz, and J. Langowski, Biophys. J. **74**, 780 (1998).
[20] J. Huang and T. Schlick, J. Chem. Phys. **117**, 8573 (2002).
[21] S. P. Mielke, W. H. Fink, V. V. Krishnan, N. Grønbech-Jensen, and C. J. Benham, J. Chem. Phys. **121**, 8104 (2004).
[22] M. Moakher and J. H. Maddocks, Archive for Rational Mechanics and Analysis (2005).
[23] I. Klapper and H. Qian, Biophys. J. **74**, 2504 (1998).
[24] P. Hagerman and B. Zimm, Biopolymers **20**, 1481 (1981).
[25] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, 2000).
[26] G. Chirico and J. Langowski, Biopolymers **34**, 415 (1994).

[27] V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Physical Chemistry of Nucleic Acids* (Harper & Row, New York, 1974).

[28] H. Noguchi and T. Masako, Phys. Rev. E **64**, 041913 (2001).

[29] J. Marmur and P. Doty, J. Mol. Biol. **5**, 109 (1962).

[30] C. Schildkraut and S. Lifson, Biopolymers **3**, 195 (1965).

[31] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).

[32] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (W. H. Freeman, New York, 1980).

[33] C. Bi and C. J. Benham, Bioinformatics **20**, 1477 (2004).

[34] H. Tsen and S. D. Levene, Proc. Natl. Acad. Sci. U.S.A. **94**, 2817 (1997).

[35] J. H. White, in *Mathematical Methods for DNA Sequences*, edited by M. S. Waterman (CRC, Boca Raton, FL, 1989), pp. 225-253.