Macroscopic modeling and simulations of supercoiled DNA with bound proteins

Jing Huang[a] and Tamar Schlick[b]
Department of Chemistry and Courant Institute of Mathematical Sciences,
New York University and Howard Hughes Medical Institute, New York, New York 10012

General methods are presented for modeling and simulating DNA molecules with bound proteins on the macromolecular level. These new approaches are motivated by the need for accurate and affordable methods to simulate slow processes (on the millisecond time scale) in DNA/protein systems, such as the large-scale motions involved in the *Hin*-mediated inversion process. Our approaches, based on the wormlike chain model of long DNA molecules, introduce inhomogeneous potentials for DNA/protein complexes based on available atomic-level structures. Electrostatically, treat those DNA/protein complexes as sets of effective charges, optimized by our discrete surface charge optimization package, in which the charges are distributed on an excluded-volume surface that represents the macromolecular complex. We also introduce directional bending potentials as well as non-identical bead hydrodynamics algorithm to further mimic the inhomogeneous effects caused by protein binding. These models thus account for basic elements of protein binding effects on DNA local structure but remain computational tractable. To validate these models and methods, we reproduce various properties measured by both Monte Carlo methods and experiments. We then apply the developed models to study the *Hin*-mediated inversion system in long DNA. By simulating supercoiled, circular DNA with or without bound proteins, we observe significant effects of protein binding on global conformations and long-time dynamics of the DNA on the kilo basepair length. © *2002 American Institute of Physics.*
[DOI: 10.1063/1.1511506]

I. INTRODUCTION

Proteins maneuver DNA structures in many biological processes of great significance. The binding of most proteins often distorts the local DNA structure and affects DNA's flexibility.[1–3] For many fundamental processes, including transcription, replication, and repair of DNA, such distortions of DNA are intimately linked to function.

Many biologically active DNA molecules are topologically closed (or behave as such) and hence are naturally supercoiled. DNA supercoiling, demonstrated for the first time in 1965 by Vinograd,[4] profoundly influences both DNA conformations and DNA's biological functions.[5–11] The effect of protein binding on a supercoiled DNA can be more complex since not only may local structure be altered but also global geometric properties that depend on the supercoiling geometry (e.g., writhing and twisting).

Atomic-level structures of protein/DNA complexes offer many detailed features of protein/DNA interactions[2] but reveal little information about the geometric and dynamic effects on supercoiled DNA. Modeling and simulation can provide insights into such structural and dynamics details of large-scale DNA. Indeed, long DNA systems, on the scale of thousands of base pairs, have been successfully modeled using a discrete chain model for Brownian dynamics

studies,[12–14] in which the DNA is represented by a uniformly charged elastic polymer immersed in an electrolytic viscous solvent. Monte Carlo methods, as well as Langevin and Brownian dynamics simulations, have been broadly employed to study the equilibrium and dynamic properties of supercoiled DNA.[13–18]

It remains a challenge to model the complex and inhomogeneous effect of proteins on DNA. The size of such systems demands a description on the polymer level, but the local structural distortions require special attention to detail. In this paper, we introduce and apply a method for modeling proteins bound to long DNA. Our model consists of fitting the excluded volume of the molecular protein surface, accounting for the effective charge distribution on that surface, and applying elements of an inhomogeneous elastic model coupled to nonidentical bead hydrodynamics. Using this economic, macroscopic model, we can perform Brownian dynamics simulations to analyze the large-scale DNA motions in processes where DNA and proteins are intimately coupled.

In the first section, we outline the theory and methods which represent DNA-bound proteins based on their macroscopic properties. Details involving the construction of a model for the *Hin*-mediated inversion system[19–21] are presented next. In the Results and Discussions sections, we validate our methods and computations by comparing *results of Monte Carlo* simulations and *experiments* with regard to the translational diffusion coefficient and radius of gyration. We then analyze our simulations on both the homogeneous

---

[a]Electronic mail: jingh@biomath.nyu.edu
[b]Author to whom correspondence should be addressed. Electronic mail: schlick@nyu.edu

         8573

model for average-sequence DNA and the inhomogeneous model for DNA bound to proteins to study the *Hin*-mediated inversion system. This systematic comparison leads to a discussion of the effect of protein binding on DNA dynamics as well as the role of supercoiling on the recombination reaction. The models and methods developed here are general and can be applied to other processes involving supercoiled DNA bound to proteins.

II. METHODS

A. Introduction

Our previous homogeneous model based on the wormlike chain and bead model for polymers describes a supercoiled DNA molecule based on its average properties and has proven to be reliable in reproducing various equilibrium and dynamic properties as well as providing new information on DNA kinetics.[14,22] However, this model does not account for inhomogeneous effects such as those induced by protein binding. Here we develop a model for this purpose, combining aspects of mechanical modeling, hydrodynamics, and electrostatics interactions. The combination of these refinements allows us to properly model and simulate supercoiled DNA molecules bound to proteins to a first approximation.

Our work is motivated by the *Hin*-mediated inversion system, a specific recombination reaction. *Hin* invertase,[2,21,23] an enzyme from *Salmonella typhimurium*, catalyzes the site specific inversion of a 996 bp (base pair) DNA segment flanked by two *hix* sites on a supercoiled DNA substrate.[24,25] Figure 1(A) illustrates our model system for this reaction: a supercoiled DNA substrate of about 5.3 kilobasepair (kbp) length as in the experiments,[25] with two *Hin* protein dimers bound to *hix* sites and two *Fis* dimers bound to an enhancer site. The protein *Fis*[26] ("Factor for Inversion Stimulation") substantially enhances the process.[27,20,28] *In vivo*, other proteins might also influence the inversion but are likely secondary in importance. Indeed, *in vitro* experiments have demonstrated that *Hin* and *Fis* alone are sufficient for the inversion reaction,[29] except in the case when the enhancer is located within 100 bp from one *hix* site, in which an additional host factor, the HU protein, is also required.[30] Therefore, our model represents a sufficiently reasonable description for this inversion process when the enhancer is at least 100 bp away from *hix* sites. Modeling this necessarily simplified system is already quite complex (as we describe in this paper) but can address specific mechanistic questions of the reaction.[31] The inversion leads to the alternative expression of two flagellin genes, that is, it controls a transition between two different kinds of flagellar antigens at frequencies from 10^{-5} to 10^{-3} per bacterial division.[32] Such a transition is important for the bacterial population in order to escape the immune system of the host. The question we will address in future simulations entails the kinetics of the recombination reactions and how the DNA topology and geometry affect the biological outcome.

B. Modeling supercoiled DNA

We model a circular DNA molecule as a closed, discrete wormlike chain. The potentials for DNA we introduce are based on the average elastic properties of DNA molecules with mixed sequences. Such homogeneous tendencies serve as a good approximation for most B-DNA segments of mixed base composition. Later we discuss a more complex, inhomogeneous model for describing the DNA segments involving intrinsic bends as well as protein-induced bends.

The DNA model is described by the following potentials: E^b, an isotropic bending potential characterized by the bending persistence length p; E^t, an isotropic twisting elastic potential; E^s, a stretching elastic potential; E^e, a screened electrostatic potential in the form of Debye–Hückel, dependent on the monovalent salt concentration of the medium; and E^v, an excluded volume potential describing the molecular surface. In Table I we summarize all the terms and symbols used in our work, and in Table II we provide the list of parameters and their values.

To represent a DNA molecule as a discrete wormlike chain, we introduce k linear elastic segments for each Kuhn statistical length l_e. The equilibrium length l_0 for each such linear segment equals to l_e/k. By this approach, we can model a circular DNA of n Kuhn statistical lengths as a closed chain with $N = kn$ vertices and N linear elastic segments.[12,13,15,33] Our previous studies[13,15] have indicated that the value of k does not affect the equilibrium properties of DNA conformations as long as $k \geq 10$; thus, we use $k = 10$ throughout this work.

A given conformation for a N vertex chain is specified by the set of N position vectors $\mathbf{r}_i$ ($i = 1, \ldots, N$) for the vertices. For each vertex i, we construct a local body-fixed coordinate (bfc) frame $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$, where

$$l_i = |\mathbf{r}_{i+1} - \mathbf{r}_i|, \tag{1}$$

$$\mathbf{a}_i = (\mathbf{r}_{i+1} - \mathbf{r}_i)/|\mathbf{r}_{i+1} - \mathbf{r}_i|, \tag{2}$$

and in which l_i is the segment length between the ith and $i+1$th vertices, and the vector $\mathbf{a}_i$ is a normalized vector pointing along the global helix axis [see Fig. 2(A)]. The vector $\mathbf{b}_i$ is a normalized vector along the short axis of the basepair plane pointing to the direction of minor groove (i.e., perpendicular to the vector $\mathbf{a}_i$). Finally, the vector to complete a right-handed (rh) coordinate system $\mathbf{c}_i$ is defined as their cross product

$$\mathbf{c}_i = \mathbf{a}_i \times \mathbf{b}_i. \tag{3}$$

1. Bending potential

The equilibrium configuration for a standard B-DNA molecule is straight on average. In a discrete wormlike model, the deformation of a chain away from the straight, equilibrium conformation is modeled as a bending energy cost proportional to the elastic bending constant $g/2$ times the sum of squares of bending angles, where g is the bending rigidity constant,

$$E^b = \frac{g}{2} \sum_{i=1}^{N} \theta_i^2, \tag{4}$$

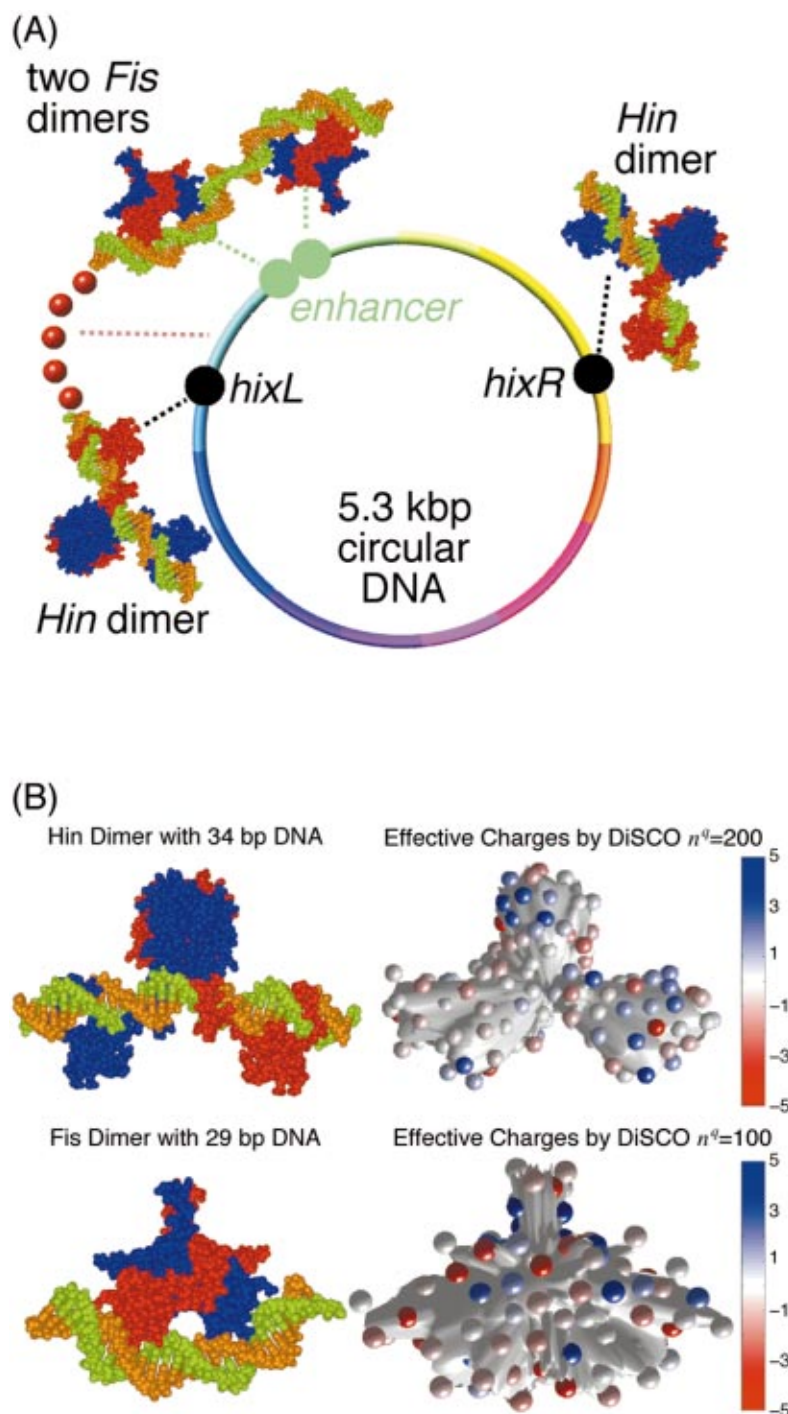$$\theta_i = \arccos(\mathbf{a}_{i-1} \cdot \mathbf{a}_i), \tag{5}$$

J. Chem. Phys., Vol. 117, No. 18, 8 November 2002

Modeling supercoiled DNA with proteins   8575

(A)

two Fis dimers

Hin dimer

enhancer

hixL   hixR

5.3 kbp circular DNA

Hin dimer

(B)

Hin Dimer with 34 bp DNA

Effective Charges by DiSCO $n^q = 200$

Fis Dimer with 29 bp DNA

Effective Charges by DiSCO $n^q = 100$

FIG. 1. (Color) The macroscopic model for supercoiled DNA with bound proteins in the *Hin*-mediated inversion system. (A) The coloring within the 5.3 kbp circular DNA substrate is only for reference, indicating the relative basepair positions, which we follow during the dynamic simulations. The two black circles indicate the two *hix* sites where two *Hin* dimers can independently bind, and the two green circles indicate the enhancer sequence, which includes two *Fis*-dimer binding sites. The two *hix* sites are separated by a DNA sequence of 996 bp, while the *hixL* and the enhancer are separated by about 110 bp. The atomic-level models of the *Hin* dimer and the *Fis* dimer bound to DNA segments are as described in Refs. 23, 26, 42, and 43, as determined from x-ray crystallography and crosslinks data. (B) Illustration of the effective charges constructed using the DiSCO package to model the *Hin* and *Fis* complexes. A total of 200 effective charges for the *Hin*-dimer/DNA and 100 effective charges for the *Fis*-dimer/DNA complexes are optimized on the virtual surface 30 Å from the molecular surface. The total effective charge is $-18.0\ e$ on a *Hin* dimer with 34 bp DNA and $-44.3\ e$ on a *Fis* dimer with 29 bp DNA, comparing with $-99.4\ e$ for a free 30 bp DNA at 0.2 M monovalent salt concentration.

$$g = \frac{A}{l_0} = \frac{p k_B T}{l_0}. \tag{6}$$

Here, E^b is the homogeneous bending potential, θ_i is the bending angle around the vertex i between the segment vector $\mathbf{a}_i$ and $\mathbf{a}_{i+1}$ (see Fig. 2), $k_B T$ is the Boltzmann factor, A is the value of conventional bending rigidity in experimental measurements according to the curvature-squared integral $[E^b = A/2\int \kappa^2(s)ds]$, and p is the bending persistence length. Experiments support that the persistence length p is around 50 nm ($A = 2.0 \times 10^{-19}$ erg cm) for the B-DNA

double-helix under normal thermal fluctuations at a monovalent salt solution ≥ 0.01 M.[34,35] We use this value throughout our study.

2. Torsional potential

In addition to the bending elasticity, a DNA molecule has twisting elasticity.[36] The twisting energy is described by the torsional rotation angles around the helix axis. That is, an equilibrium configuration of B-DNA has a helix repeat of 36 Å (10.5 base pairs/turn); deviations of this twisting rate

TABLE I. List of symbols.

Symbol	Definition
$\mathbf{r}_i$	coordinates of the ith vertex
$\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$	local coordinate system on the ith vertex
$\{\alpha_{i,i+1}, \beta_{i,i+1}, \gamma_{i,i+1}\}$	Euler angles for the transformation from $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$ to $\{\mathbf{a}_{i+1}, \mathbf{b}_{i+1}, \mathbf{c}_{i+1}\}$
$\phi_{i,i+1}$	torsional angle for the transformation from $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$ to $\{\mathbf{a}_{i+1}, \mathbf{b}_{i+1}, \mathbf{c}_{i+1}\}$
l_i	segment length between ith and $i+1$th vertices
$\mathbf{r}_{ij}$	displacement between the ith and jth vertices
E^b	elastic bending potential
E^t	elastic twisting potential
E^s	elastic stretching potential
E^e	electrostatic potential
E^v	excluded volume potential
g	bending rigidity constant
θ_i	bending angle around the vertex i
s	twisting rigidity constant
ϕ_0	homogeneous intrinsic twist in one model segment
Γ_i	roll-like bending angle
Υ_i	tiltlike bending angle
n_i^q	number of charge points used to describe a complex located on ith vertex
q_i^k	value of the kth effective charge assigned for the complex at the ith vertice
$\{\mathbf{x}_i^k, \mathbf{y}_i^k, \mathbf{z}_i^k\}$	position of the kth effective charge for the complex at the ith vertice based on the local coordinates $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$
$\mathbf{r}_i^k$	position of the kth effective charge for the complex at the ith vertice in the global coordinates
r_{ij}^{kl}	distance between $\mathbf{r}_i^k$ and $\mathbf{r}_j^l$
δ_{ij}	collision criteria for excluded volume potential
t	time
$\mathbf{D}^t$	hydrodynamic diffusion tensor for nonidentical bead model
$\mathbf{T}_{ij}$	hydrodynamic tensor for interactions between the ith and jth beads
η_0	solvent viscosity
ρ_i	hydrodynamic radius of the ith bead
$\mathbf{I}$	3×3 identity matrix
η_i	rotational friction coefficent of the ith segment
$\mathbf{F}^t, \mathbf{r}^t$	collective force and position vectors at time t
τ_i^t, ϕ_i^t	torque and torsional angle of the ith segment at time t
$\mathbf{R}^t, \omega_i^t$	random displacements and rotations in Brownian dynamics algorithms at time t
$\delta_{t,t'}$	Kronecker delta function of t and t'
D_t	translation diffusion coefficient
$\mathbf{m}(t)$	position of the center of the mass at time t
σ	DNA superhelical density
R_g	radius of gyration
W_r	writhe of the circular DNA
T_{c1}	site juxtaposition time for two *hix* sites
τ_{i_2}	autocorrelation time for juxtaposition events
I_D	set of vertices describing free DNA
I_P	set of vertices describing DNA/protein complexes

from equilibrium are regulated by the torsional potential, described by the parameters constructed as follows.

To describe the torsional twist, we use the local coordinate frame defined on the ith vertex $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$. The transformation from one frame $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$ to the next $\{\mathbf{a}_{i+1}, \mathbf{b}_{i+1}, \mathbf{c}_{i+1}\}$ can be quantitatively defined by a set of Euler angles $\{\alpha_{i,i+1}, \beta_{i,i+1}, \gamma_{i,i+1}\}$ based on the following procedure: First, we define

$$\beta_{i,i+1} = \arccos(\mathbf{a}_i \cdot \mathbf{a}_{i+1}). \tag{7}$$

The value of $\beta_{i,i+1}$ coincides with the bending angle θ_{i+1}. Next we define $\alpha_{i,i+1}$ as

$$\alpha_{i,i+1} = \begin{cases} \alpha_c & \text{if } \mathbf{a}_{i+1} \cdot \mathbf{c}_i \geq 0, \\ -\alpha_c & \text{if } \mathbf{a}_{i+1} \cdot \mathbf{c}_i < 0, \end{cases} \tag{8}$$

$$\alpha_c = \arccos\left(\frac{\mathbf{a}_{i+1} \cdot \mathbf{b}_i}{\sin(\beta_{i,i+1})}\right). \tag{9}$$

Here, α_c is defined in the interval of $[0, \pi]$, and $\mathbf{a}_{i+1} \cdot \mathbf{c}_i$ determines the sign of the angle $\alpha_{i,i+1}$. Finally, we compute $\gamma_{i,i+1}$ from the relations

$$\gamma_{i,i+1} = \begin{cases} \varphi & \text{if } \sin \varphi \geq 0, \\ -\varphi & \text{if } \sin \varphi < 0, \end{cases} \tag{10}$$

$$\varphi = \arccos\left(\frac{\mathbf{b}_i \cdot \mathbf{b}_{i+1} + \mathbf{c}_i \cdot \mathbf{c}_{i+1}}{1 + \mathbf{a}_i \cdot \mathbf{a}_{i+1}}\right), \tag{11}$$

$$\sin \varphi = \frac{\mathbf{b}_i \cdot \mathbf{b}_{i+1} - \mathbf{c}_i \cdot \mathbf{c}_{i+1}}{1 + \mathbf{a}_i \cdot \mathbf{a}_{i+1}}. \tag{12}$$

TABLE II. Elastic, geometric, and electrostatic parameters used in the DNA/protein model.

Parameter	Definition	Value
N	number of the vertices in the circular DNA model	176 (5.28 kbp)
l_0	equilibrium segment length	10 nm
l_e	Kuhn statistical length of DNA	100 nm
k	number of elastic segment for each Kuhn statistical length	10
p	persistence length for regular B-DNA	50 nm
A	bending rigidity constant	2.0×10^{-19} erg cm
C	twisting rigidity constant	3.0×10^{-19} erg cm
k_B	Boltzmann's constant	1.38×10^{-23} J/K
T	absolute temperature	298 K
χ_0	number of the base pairs in each B-DNA double helix turn	10.5
d_B	distance between two adjacent base pairs along the B-DNA axis	3.4 Å
g	bending rigidity constant of average DNA	pk_BT/l_0
h	stretching rigidity constant	$100k_BT/l_0^2$
i_{hin}^1, i_{hin}^2	vertices to which a *Hin* dimer can bind	1, 34
i_{fis}^1, i_{fis}^2	vertices to which a *Fis* dimer can bind	5, 6
Γ_{ihin}^0, Υ_{ihin}^0	equilibrium tiltlike and roll-like bends on an *ihin* vertex by a *Hin* dimer	$-18°$, $0°$
Γ_{ifis}^0, Υ_{ifis}^0	equilibrium tiltlike and roll-like bends on an *ifis* vertex by a *Fis* dimer	$60°$, $0°$
g_{ihin}^{Γ}, g_{ihin}^{Υ}	computational tiltlike and roll-like bending rigidity for *ihin* vertices	g, g
g_{ifis}^{Γ}, g_{ifis}^{Υ}	computational tiltlike and roll-like bending rigidity for *ifis* vertices	$3.0\,g$, $3.0\,g$
c_S	monovalent salt concentration	0.2 M
κ	inverse Debye length (salt-dependent, here for 0.2 M)	1.477 nm^{-1}
λ	effective linear charge density of double helix (here for 0.2 M)	$40.9\,e/$nm
ϵ	relative dielectric constant of aqueous medium	80
n_i^q	number of effective charges for a DNA segment, a *Hin*-dimer/ DNA complex, and a *Fis*-dimer/DNA complex, respectively	5, 200, 100
μ	computational short-range repulsion force	35 pN
Δt	time step for Brownian dynamic simulations	600 ps
ρ_0	hydrodynamic radius of a DNA segment of l_0	2.24 nm
d_0	radial distance criterion for site juxtaposition	10 nm

The torsional angle around the $i+1$th vertex, or the transformation from the local coordinates $\{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i\}$ to $\{\mathbf{a}_{i+1},\mathbf{b}_{i+1},\mathbf{c}_{i+1}\}$ can be defined by $\phi_{i,i+1}$,

$$\phi_{i,i+1} = \alpha_{i,i+1} + \gamma_{i,i+1}. \qquad (13)$$

We can thus define the elastic torsional potential as a function of $\phi_{i,i+1}$, which is proportional to a twisting elastic constant times the square of the torsional deviation from the equilibrium rate,[33]

$$E^t = \frac{s}{2} \sum_{i=1}^{N} (\phi_{i,i+1} - \phi_0)^2, \qquad (14)$$

$$s = \frac{C_0}{2l_0}. \qquad (15)$$

In these equations, E^t is the twisting potential, s is the twisting rigidity constant, ϕ_0 is the homogeneous intrinsic twist of one model segment, and C_0 is the value of conventional twisting rigidity in experimental measurement according to $E^t = C/2 \int \phi^2(s)ds$. Various experiments indicate that $C_0 \sim 3.0 \times 10^{-19}$ erg cm,[36,37] the value used throughout this work.

The equilibrium value ϕ_0 can be extracted from parameters for the standard B-DNA model,

$$\phi_0 = \frac{2\pi l_0}{\chi_0 d_B} \qquad (16)$$

in which χ_0 is the number of the base pairs in a 360° turn in the double helix and d_B is the distance between two adjacent base pairs along the DNA double helix axis. The common accepted values are $\chi_0 = 10.5$ and $d_B = 3.4$ Å.

3. Stretching potential

DNA segments are fairly rigid on the scale of dozens of basepairs. We use a computational harmonic stretching potential[38] of the form

$$E^s = \frac{h}{2} \sum_{i=1}^{N} (l_i - l_0)^2, \qquad (17)$$

where l_i is the segment length between ith and $i+1$th vertices as defined in Eq. (1) and h is the stretching rigidity constant. We choose $h = 100k_BT/l_0^2$ so that the variance of l_i is close to $l_0^2/100$ in the simulations.[14] Results from recent micromanipulation experiments that applied pico-Newton forces to a single double-stranded DNA validate that the longitudinal stretching of DNA can be described by thermal fluctuations at room temperature using a harmonic stretching potential.

C. Modeling DNA segments bound to proteins: Local bending

The binding of proteins to DNA usually introduces local distortions to the DNA binding sites and affects the elastic
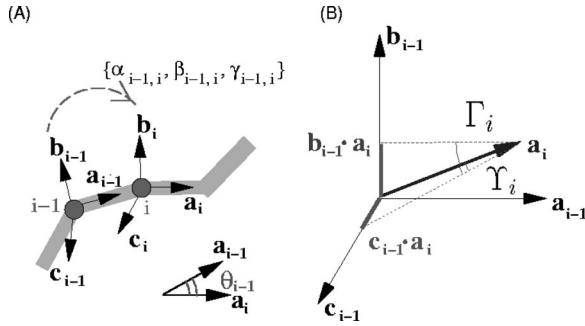
FIG. 2. Local coordinate systems fixed on vertices, with symbols describing the transformations from one reference frame to another. (A) On the ith vertex, we define a local body-fixed coordinate (bfc) frame $\{\mathbf{a}_i,\mathbf{b}_i,\mathbf{c}_i\}$ with Euler angles $\{\alpha_{i,i+1},\beta_{i,i+1},\gamma_{i,i+1}\}$ that describe the transformation from the bfc frame i to $i+1$. (B) Two angles, Γ_i and Υ_i, are used to describe the directional bends, as follows. The angle Γ_i describes the projection of the bend from $\mathbf{a}_{i-1}$ to $\mathbf{a}_i$ along the direction of $\mathbf{b}_{i-1}$, while the angle Υ_i describes the projection of the same bend along the direction of $\mathbf{c}_{i-1}$.

flexibility of local DNA segments. Our mechanical model incorporates these effects by using an inhomogeneous bending energy instead of the homogeneous bending energy, which we construct as follows.

We divide the elastic bending potential into two components, the tiltlike bending, and the roll-like bending. In a DNA model at the base-step level, the tilt defines the opening angle between two basepair planes with respect to the short axis of the basepair plane, while the roll defines the deformation with respect to the long axis of the basepair plane.[39] We similarly define here the tiltlike bending angle, Γ_i, as the bending from $\mathbf{a}_{i-1}$ to $\mathbf{a}_i$ with respect to the direction of $\mathbf{b}_{i-1}$, and roll-like bending, Υ_i, as the bending with respect to the direction of $\mathbf{c}_{i-1}$ [Γ_i and Υ_i are illustrated in Fig. 2(B)]:

$$\Gamma_i = \sin^{-1}(\mathbf{a}_i \cdot \mathbf{b}_{i-1}), \tag{18}$$

$$\Upsilon_i = \sin^{-1}(\mathbf{a}_i \cdot \mathbf{c}_{i-1}). \tag{19}$$

For an inhomogeneous (i.e., protein/DNA) site, the local bending energy E_i^b can thus be defined as a function of Γ_i and Υ_i instead of θ_i in Eq. (4),

$$E_i^b = \frac{g_i^\Gamma}{2}(\Gamma_i - \Gamma_i^0)^2 + \frac{g_i^\Upsilon}{2}(\Upsilon_i - \Upsilon_i^0)^2, \tag{20}$$

where Γ_i^0 and Υ_i^0 are equilibrium values for bending angles along two perpendicular directions introduced by inhomogeneous factors, such as protein binding and DNA sequence effects. The corresponding bending rigidities g_i^Γ and g_i^Υ describe the bending flexibility of Γ_i and Υ_i, respectively. The values of g_i^Γ and g_i^Υ can be calculated in theory from experimental values of the local bending persistence length p_i based on

$$g_i = p_i k_B T / l_0. \tag{21}$$

If such experimental measurements are not available, we estimate values for the bending rigidities based on similar systems.

For the *Hin*-mediated inversion system, we aim to model the circular DNA substrate of length 5.3 kbp[24] shown in Fig. 1. We use a chain of N vertices ($N=176$), in which each segment represents 30 bp of DNA, for a combined contour

length for the DNA of 5.28 kbp. Two *Hin* dimers can independently bind to two *hix* recombination sites,[23] which are separated by $\sim$990 bp of DNA. These two *hix* sites are thus located on two vertices i_{hin}^1 and i_{hin}^2: $i_{hin}^1=1$, $i_{hin}^2=34$. The two *Fis*-dimer binding sites are located at a 60 bp enhancer sequence. In the wild-type system,[20,40] the enhancer region extends from sites $+103$ to $+163$ from the center of the *hix* L site, so we model the enhancer with two *Fis*-dimer binding sites by two vertices i_{fis}^1 and i_{fis}^2: $i_{fis}^1=5$, $i_{fis}^2=6$.

Effects of the *Hin*-dimer and *Fis*-dimer binding on their respective sequences have been well characterized by various experiments. In the complex between the *hix* DNA sequence and the *Hin* dimer, the DNA is slightly bent ($\sim$18°), widening the minor groove at the center of the *hix* site.[41–43] We use this result to calculate the equilibrium values for the directional bending of the *Hin*-dimer binding sites as

$$\Gamma_{ihin}^0 = -18°, \tag{22}$$

$$\Upsilon_{ihin}^0 = 0°, \tag{23}$$

$$i_{hin} = 1 \text{ or } 34. \tag{24}$$

The *Fis* dimer is known to severely bend the DNA segment. The position, degree, and the direction of this DNA deformation have been systematically measured by electrophoretic mobility and phasing analysis.[44] It is found that a *Fis*-dimer bends DNA 60° on average towards the minor groove of the DNA helix at the center of both *Fis* binding sites.[43–45] We can thus estimate equilibrium values for the inhomogeneous bending potential [Eq. (20)] to obtain

$$\Gamma_{ifis}^0 = 60°, \tag{25}$$

$$\Upsilon_{ifis}^0 = 0°, \tag{26}$$

$$i_{fis} = 5 \text{ or } 6. \tag{27}$$

Gel mobility experiments show that the bends introduced by the *Fis* dimer lead to greater rigidity than average sequences at room temperature;[46,45] rigidity has, nonetheless, not been quantified by experiments. Since the value of the bending rigidity constant g is based on average B-DNA sequences, bending reference values for $\{g_{ifis}^\Gamma, g_{ifis}^\Upsilon\}$ should be greater than g and can be roughly estimated to be in the order of 2 to 10 times greater.[46,45] We use $3.0g$ to approximate bending rigidities of the *Fis*-dimer binding sites.

In our model, we use as a first approximation the same homogeneous twisting and stretching elastic potentials as for DNA segments of average sequences (unbound to proteins). These potentials can be similarly replaced by inhomogeneous potentials later if significant inhomogeneous effects are identified and quantified by experiments.

D. Modeling DNA segments bound to proteins: Electrostatics

We use our recently developed method, DiSCO (discrete surface charge optimization),[47] to model the electrostatic properties of a DNA segment bound to proteins based on atomic-level structures of DNA/protein complexes. This method economically describes the electrostatic field predicted by Poisson–Boltzmann theory using a discrete set of

Debye–Hückel charges distributed on a virtual surface near the macromolecular surface. DiSCO was applied to study the folding of chromatin on the macromolecular level.[48]

The procedure in DiSCO relies on the linear behavior of the Poisson–Boltzmann equation in the far zone. In this range, we can superimpose contributions from a number of atoms and approximate the electrostatic potential quickly and efficiently through the cumulative contribution from the set of effective charges. This approach allows us to use only a few hundred effective charges instead of tens of thousands or more atom charges in the macromolecular structure to approximate the electrostatics.[47] This approximation significantly reduces the calculation of long-range electrostatic interactions, a costly aspect of the dynamic simulations.

To determine the positions and values of effective charges required for the DiSCO approximation, we first construct a surface S which encloses the macromolecular surface of the crystal structure. We then locate the positions of a specific number of effective charges $\{\mathbf{x}_i^k, \mathbf{y}_i^k, \mathbf{z}_i^k\}$ by evenly distributing them on the virtual surface. We assign a set of initial charges $\{q_i^k\}$ using the AMBER force field.[49] The calculation of the electrostatic field is sampled on a grid implemented by the Delphi solver.[50] We formulate the residual $R(E, E', q_i^k)$ as the relative difference between E (the original electrostatic field of the macromolecular structure predicted by the Poisson–Boltzmann equation) and E' (the field predicted by the Debye–Hückel equation and effective charges $\{\mathbf{x}_i^k, \mathbf{y}_i^k, \mathbf{z}_i^k, q_i^k\}$) over the set of points V_i' on the exterior of surface S (see Fig. 1C of Ref. 47):

$$R(E, E', q_i^k) = \frac{1}{N_v'} \sum_{V_i'} \frac{\|E - E'\|}{\|E\|}. \tag{28}$$

Here we introduce a parameter d_V to define V_i', which includes all the points in the grid outside of the surface S for which the minimum distance to S is greater or equal to d_V. The minimum distance d_V is set as 30 Å, and N_V is the number of grid points in V_i'. The relationship between the macromolecule, the surface S, and the minimization region V_i' is clearly illustrated in Fig. 1 of Ref. 47. This residual function $R(E, E', q_i^k)$ is minimized in DiSCO by varying the charges values $\{q_i^k\}$ as independent variables. For this minimization task, we use the efficient truncated Newton package (TNPACK).[16,51]

In our study, we model the complex of a *Hin* dimer with its binding site (34 bp DNA) and the complex of a *Fis* dimer with its binding site (29 bp DNA) based on atomic structures provided to us by Johnson and Haykinson (private communication). We test different numbers of effective charges and report the relationship between the number of effective charges and the associated final percentage errors (minimized R values) in Table III. The shape of the *Hin*-dimer/DNA system is more irregular and thus requires a larger number of effective charges. Following experimentation, we set the number of effective charge points $n^q = 100$ for modeling each *Fis*-dimer/DNA complex and $n^q = 200$ for modeling each *Hin*-dimer/DNA complex. We display the positions and effective charges for both *Hin*-dimer/DNA and *Fis*-dimer/DNA structures along with the original atomic structures in Fig. 1(B).

TABLE III. Optimization results for *Hin*-dimer/DNA and *Fis*-dimer/DNA complexes by DiSCO in 0.2 M monovalent salt. Optimizations are performed on the electrostatic field for all the grid points at least 30 Å away from the molecular surface. The errors indicate the difference between the electrostatic field by effective charges and the field by atomic partial charges in original structures. See Ref. 47 for details.

Complex/n^q:	100	200	300	400	500
Fis dimer with 29 bp DNA	4.4%	2.4%	2.0%	2.0%	1.9%
Hin dimer with 34 bp DNA	15.1%	5.9%	3.6%	2.9%	3.1%

For those vertices associated with free DNA sites, we place point charges located on the DNA chain segments. The number of point charges placed on each segment, n^q, is chosen to approximate continuous charges with the same linear density. The value of n^q increases as the Debye length, $1/\kappa$, decreases. We choose $n^q = 5$ for the environment of 0.2 M monovalent salt concentration, as our simulation results do not depend on n^q as long as $\{n^q \geqslant 5\}$.[14]

Therefore, each site i (located at $\mathbf{r}_i$), which either describes a free DNA part or a DNA/protein site, is assigned to a set of n_i^q charges $\{q_i^k, k = 1, \dots, n_i^q\}$. The position of the kth effective charge for the complex at the ith vertice is defined as $\mathbf{r}_i^k$ in our global coordinate representation. For sites associated with free DNA, the values of $\mathbf{r}_i^k$ can be calculated based on linear interpolation between nearby vertices

$$\mathbf{r}_i^k = \left(\frac{k}{n_i^q} - \frac{1}{2}\right)\mathbf{r}_{i+1} + \left(\frac{3}{2} - \frac{k}{n_i^q}\right)\mathbf{r}_i \quad \text{if } k \geqslant \frac{n_i^q}{2}, \tag{29}$$

$$\mathbf{r}_i^k = \left(\frac{1}{2} - \frac{k}{n_i^q}\right)\mathbf{r}_{i-1} + \left(\frac{1}{2} + \frac{k}{n_i^q}\right)\mathbf{r}_i \quad \text{if } k < \frac{n_i^q}{2}, \tag{30}$$

$$q_i^k = \lambda l_0 / n_i^q, \tag{31}$$

in which $\{\mathbf{r}_i^k, q_i^k\}$ are the coordinates and charge values for the kth effective charge associated with the ith vertex, and λ is the effective linear charge density of the double helix. For free DNA segments, we have $n_i^q = 5$ and $\lambda = 40.9 \ e/\text{nm}$ under 0.2 M monovalent salt concentration.

For sites associated with DNA/protein complexes, the relative position of an effective charge to the ith vertex $\mathbf{r}_i$ is first determined in the DiSCO modeling procedure as $\{\mathbf{x}_i^k, \mathbf{y}_i^k, \mathbf{z}_i^k\}$ in the local right-handed coordinate system of the ith vertex, which is defined by $\{\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i\}$. The local coordinates $\{\mathbf{x}_i^k, \mathbf{y}_i^k, \mathbf{z}_i^k\}$ can be transformed to the global coordinates as $\mathbf{r}_i^k, k = 1, \dots, n_i^q$:

$$\mathbf{r}_i^k = \mathbf{r}_i + \mathbf{a}_i \cdot \mathbf{x}_i^k + \mathbf{b}_i \cdot \mathbf{y}_i^k + \mathbf{c}_i \cdot \mathbf{z}_i^k. \tag{32}$$

The electrostatic potential is thus specified as the sum of Debye–Hückel potentials between pairs of non-neighbor effective charges,

$$E^e = \sum_{j>i+1}^{N} \sum_{k=1}^{n_i^q} \sum_{l=1}^{n_j^q} \frac{q_i^k q_j^l \exp(-\kappa r_{ij}^{kl})}{\epsilon r_{ij}^{kl}}, \tag{33}$$

$$r_{ij}^{kl} = |\mathbf{r}_i^k - \mathbf{r}_j^l|, \tag{34}$$

where κ denotes the inverse Debye length (salt dependent), and ϵ is the relative dielectric constant of the aqueous medium. The value of $\{\mathbf{r}_i^k, q_i^k\}$ represents the coordinates and charge value of the kth effective charge describing either the

DNA or DNA/protein complex centered at the ith vertex, and r_{ij}^{kl} is the distance between two effective charges located at $\mathbf{r}_i^k$ and $\mathbf{r}_j^l$. The electrostatic potential is replaced by an excluded volume potential in the region of physical overlapping (see next section).

E. Modeling DNA segments bound to proteins: Excluded volume potential

The energy of a short-range repulsion, E^v, is added to the potential function to prevent unrealistic crossings and collsions between segments. We exclude such events to maintain a fixed topology of the model chain. Basically, we estimate whether two volumes collide by measuring the distance between all pairs of effective charges according to

$$E^v = -\sum_{i>j}^{N} \sum_{k}^{n_i^q} \sum_{l}^{n_j^q} \mu r_{ij}^{kl} \quad \text{if } r_{ij}^{kl} < \delta_{ij}, \tag{35}$$

$$= 0 \quad \text{if } r_{ij}^{kl} > \delta_{ij}. \tag{36}$$

Here, r_{ij}^{kl} is the distance between a pair of effective charges as defined in Eq. (34), n_i^q is the number of effective charges assigned for the ith site, and δ_{ij} is the collision criterion (discussed in the next paragraph). The computational parameter μ describes the strength of the repulsive potential, and its value can be set by simulation tests.

The value of δ_{ij} depends on whether the ith and jth vertices belong to the set of vertices for free DNA segments, I_D, or the set for DNA/protein complexes, I_P. The major difference between DNA segments and protein/DNA complexes is that, for a free DNA segment, effective point charges are located on the axis (which is about 1 nm from the molecular surface), while for protein/DNA complexes, the effective point charges are located on a virtual surface (which is about 0.1 nm inside the real molecular surface). Thus, we use different values of δ_{ij} to describe the collision criteria between two molecular surfaces:

$$\delta_{ij} = 2.0 \text{ nm} \quad \text{if } i,j \in I_D,$$

$$= 1.1 \text{ nm} \quad \text{if } i \in I_D \text{ and } j \in I_P,$$

$$= 0.2 \text{ nm} \quad \text{if } i,j \in I_P. \tag{37}$$

The value for the parameter μ is 35 pico-Newton (pN) in our study. With this setting, the frequency of segment passing events is less than 10^{-5} per simulation step ($\Delta t = 600$ ps, see next section). We monitor those segment passing events as in our previous studies,[13,14] by identifying any discontinuous changes in the writhing number of the chain.[52]

F. Brownian dynamics (BD) simulations with hydrodynamics

We use the second-order BD algorithm[53] with modifications to improve the efficiency by a less-frequent updating of the hydrodynamic diffusion tensor than the systematic forces (e.g., every 10 time steps). This approach was proposed and tested in our previous works.[13,14,54] The first-order BD algorithm generates the new position vector $\mathbf{r}^{t+\Delta t}$ and rotation angles $\phi_i^{t+\Delta t}$:

$$\mathbf{r}^{t+\Delta t} = \mathbf{r}^t + \frac{\Delta t}{k_B T} \mathbf{D}^t \cdot \mathbf{F}^t + \mathbf{R}^t, \tag{38}$$

$$\phi_i^{t+\Delta t} = \phi_i^t + \frac{\Delta t}{\xi_i} \tau_i^t + \omega_i^t, \tag{39}$$

where Δt is a numerical time step (600 ps here), $\mathbf{r}^t$ is the collective position vector for the N vertices $\{\mathbf{r}_i\}$, $i = 1, \ldots, N$ at time t, $\mathbf{F}^t$ is the collective force vector for the systematic forces applied to the N vertices, as derived from the energy functions discussed above, $(\phi_i^{t+\Delta t} - \phi_i^t)$ is the rotation about the ith rotational degree of freedom, and τ_i^t is the torque acting on the ith segment based on the systematic energy function. Hydrodynamic interactions with the solvent are specified by the rotational friction coefficients $\{\xi_i\}$ and the configuration-dependent diffusion tensor $\mathbf{D}^t$ [defined in Eq. (42) below]. The BD algorithm also includes two stochastic terms $\mathbf{R}^t$ and ω_i^t, which represent white noise random values used to model thermal interactions with the solvent. The correlation structure of those white noise terms is related to the hydrodynamic interactions by[55]

$$\langle \omega_i^t \omega_i^{t'} \rangle = 2 k_B T \xi_i \delta_{t,t'}, \tag{40}$$

$$\langle (\mathbf{R}^t)(\mathbf{R}^{t'})^T \rangle = 2 \Delta t \mathbf{D}^t \delta_{t,t'}, \tag{41}$$

where $\delta_{t,t'}$ is the Kronecker delta function.

In practice, there are two different approaches available for calculating $\mathbf{R}^t$, the correlated random displacement vector. The traditional approach depends on the Cholesky decomposition of $\mathbf{D}^t$, which requires $\mathcal{O}(N^3)$ operations.[56] An alternative method proposed by Fixman,[57] utilizing Chebyshev polynomial representations, has a lower scaling term of $\mathcal{O}(N^{2.25})$ but a larger prefactor due to the computational complexity. The Chebyshev alternative has been recently applied for various polymer systems.[58–60] We tested computational performances of both approaches and chose the Cholesky approach due to a slight computational advantage in our system. The more efficient Chebyshev algorithm will likely be more important for larger system sizes.

G. Modeling DNA segments bound to proteins: Hydrodynamics

We use a nonidentical Oseen diffusion tensor to specify the hydrodynamic interaction of the DNA chain bound to proteins.[61,62] We position beads of radius ρ_i at the ith vertex of the chain. These beads are used only to define the hydrodynamic interaction and thus do not affect equilibrium properties of the model chain.

The diffusion tensor used in the Brownian dynamics (BD) algorithm [Eq. (38)] is a $3N \times 3N$ tensor $\mathbf{D}$ for a N-bead system based on a set of $\mathbf{T}_{ij}$:

$$\mathbf{D} = k_B T \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1N} \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2N} \\ \vdots & \vdots & & \vdots \\ \mathbf{T}_{N1} & \mathbf{T}_{N2} & \cdots & \mathbf{T}_{NN} \end{bmatrix}, \tag{42}$$

where each $\mathbf{T}_{ij}$ is a 3×3 matrix representing the interaction between the ith and jth beads. Each $\mathbf{T}_{ij}$ can be calculated as follows:

$$\mathbf{T}_{ij} = \frac{1}{6\pi\eta_0\rho_i}\mathbf{I} \quad \text{for} \quad i=j \quad (\text{same bead}) \tag{43}$$

$$= \frac{1}{8\pi\eta_0 r_{ij}}\left[\left(1+\frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2}\right)+\frac{\rho_i^2+\rho_j^2}{r_{ij}^2}\left(\frac{1}{3}\mathbf{I}-\frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2}\right)\right]$$

$$\text{for} \quad i\neq j \quad (\text{different beads}), \tag{44}$$

where $\mathbf{I}$ is the 3×3 identity matrix, ρ_i is the hydrodynamic bead radius, η_0 is the viscosity of the solvent, $\mathbf{r}_{ij}$ is the displacement between ith and jth bead, and $r_{ij}=|\mathbf{r}_{ij}|$.

The DNA/protein complexes are treated as spheres for hydrodynamic purposes. The rotational frictional coefficients [ξ_i in Eq. (39)] can be expressed as

$$\xi_i = 8\pi\eta_0\rho_i^3, \tag{45}$$

where ρ_i is the effective hydrodynamic radius of the DNA/protein complex and η_0 is the solvent viscosity. For beads representing free DNA segments, the rotational friction coefficient can be expressed as

$$\xi_i = 4\pi\eta_0\rho_i^2 l_0, \tag{46}$$

in which $\rho_i=\rho_0=2.24$ nm is the hydrodynamic radius of DNA. This value is chosen to provide the experimentally measured values of sedimentation coefficients of circular DNA.[18,63,64] Since rotation occurs only about the DNA axis $\mathbf{a}_i$, the ξ_i values are effectively infinite for the rotations along $\mathbf{b}_i$ and $\mathbf{c}_i$.

For the DNA/protein complex, we can directly calculate the effective hydrodynamic radius ρ_i if the translational diffusion coefficient of the DNA/protein complex has been experimentally determined. We can derive the value of ρ_i from the experimental value of the translational diffusion coefficient D_t based on the following relationship:

$$\rho_i = \frac{k_B T}{6\pi\eta_0 D_t}. \tag{47}$$

However, for most of DNA/protein complexes, experimental values of translational diffusion coefficients are not yet available. We can thus use theoretical modeling to first estimate the value of the translational diffusion coefficient. We use the package HYDROPRO[65] to build detailed models for Hin-dimer/DNA and Fis-dimer/DNA complexes based on atomic-level structures. The algorithm of HYDROPRO involves building a shell model[66] and calculating of the global hydrodynamic properties such as the translational diffusion coefficient.

The shell method[66] builds the surface of a macromolecule as a shell derived from many small spheres. When extrapolating the sphere size to zero, the shell modeling method has proven accurate for calculating hydrodynamic properties of macromolecules with irregular shapes.[67,68] The macroscopic properties are calculated based on the shell model with small spheres and the calculation is repeated iteratively with decrease of the sphere size to derive final results at the limit of sphere size zero. This method has been tested by Garcia de la Torre $et\ al.$[65] on 13 different atomic-level protein structures, including BPTI and lysozyme by comparing the calculated hydrodynamic properties with ex-

TABLE IV. Computed values of the translational diffusion coefficients for the Hin-dimer/DNA and Fis-dimer/DNA complexes based on hydrodynamic shell models of HYDROPRO. The second column lists the maximum number of the spheres used in the shell models. Various number of spheres are used for each complex, and D_t values are based on the extrapolation of the sphere size to zero. The calculations are performed at the temperature T of 298 K and the solvent viscosity for water η_0 of 0.089 poise.

Complex	Maximum number of spheres	D_t (cm²/s)
Hin-dimer/34 bp	4407	6.14×10^{-7}
Fis-dimer/29 bp	2764	7.91×10^{-7}

perimental values. The percentage difference between calculated and experimental values for the translational diffusion coefficient varies from 0.0 to 4.9 % for those 13 complexes.

The computation involved in the shell modeling requires minutes to hours of computer time to calculate macroscopic properties based on a shell model with the maximum number of spheres for the model in the order of thousands (Table IV). However, such a calculation needs to be done only once before the BD simulations to determine the effective ρ_i for the Hin-dimer/DNA and Fis-dimer/DNA complexes. Thus, the computational cost involved is acceptable. Our final values of the translation diffusion coefficients resulting from HYDROPRO are listed in Table IV. The corresponding effective hydrodynamic radii ρ_i can thus be directly calculated based on Eq. (47).

III. RESULTS AND DISCUSSIONS

A. Validation: comparison of D_t and R_g to experimental data and Monte Carlo

To validate our methodology and computational procedure, we first compare the conformational properties obtained by BD versus Monte Carlo (MC) simulations, as well as experimental data. Sufficiently long BD trajectories should reproduce thermodynamic ensembles of the equilibrium conformations generated in MC.

We first compare the translational diffusion coefficient D_t of the supercoiled DNA chain obtained by BD with data available from light-scattering experiments.[18,64] We calculate D_t from BD simulations from the displacement of the center of mass vector, $\mathbf{m}(t)$, according to the Einstein–Stokes equation:

$$6tD = \langle|\mathbf{m}(t)-\mathbf{m}(0)|^2\rangle, \tag{48}$$

where t denotes time. We analyze multiple trajectories to obtain reasonable statistics. All BD simulations are performed for the 5.28 kbp supercoiled DNA molecule (with superhelical density $\sigma=-0.06$) under 0.2 M monovalent salt. Light-scattering techniques have been applied to measure the translational diffusion coefficient of supercoiled DNA molecules in the laboratory. We plot in Fig. 3 published values of D_t by the light scattering for supercoiled DNA sizes ranging from 1.9 to 12 kbp. Experimental data are also available for 0.2 M monovalent salt concentration, with σ close to -0.06. Figure 3 shows a good agreement between

8582   J. Chem. Phys., Vol. 117, No. 18, 8 November 2002
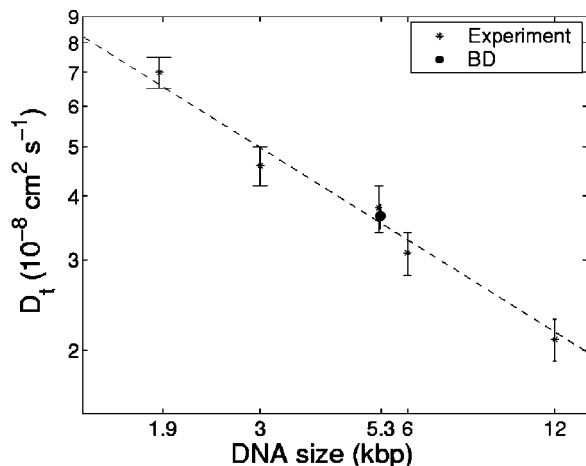
J. Huang and T. Schlick

FIG. 3. Comparison of translation diffusion coefficients D_t: BD simulations vs experimental data. The stars are the experimental data available from light scattering for supercoiled DNA molecules (without proteins) of sizes from 1.9 kbp to 12 kbp (Refs. 18 and 64). The filled circle is the value calculated by our BD simulations for 5.28 kbp DNA. The superhelical density of the DNA is -0.06 in our models, close to the experimental value. Both the BD and light-scattering data are obtained under 0.2 M monovalent ion concentration.

the BD and light-scattering data. The error bar indicates good statistics in our results (six trajectories of length 20 ms were used).

We next compare the radius of gyration R_g of the supercoiled DNA molecules obtained by BD, MC, and static light-scattering experiments under various superhelical densities (σ). Static light-scattering provides the average of the radius of gyration, R_g, of molecules in solution. Figure 4 shows the published experimental data[5,69] of R_g for the 5.2 kbp SV40 DNA plasmid with $\sigma \approx -0.05$ under 0.179 M NaCl solution. The error bar of this experimental result is largely due to the equipment limitation (limited resolution of convenient light source). MC ensembles were used to systematically measure R_g as a function of σ for a 5.2 kbp supercoiled DNA chain
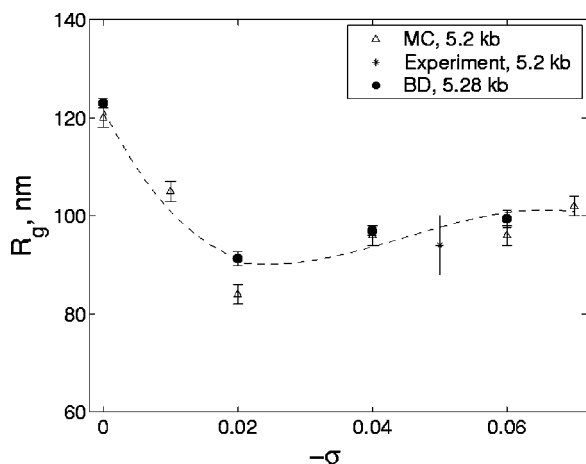

FIG. 4. Variation of the radius of gyration (R_g) of supercoiled DNA with the superhelical density (σ) as computed by Brownian dynamics (BD), Monte Carlo (MC) simulations, and light-scattering experimental data for supercoiled DNA molecules (without proteins) (Ref. 69). Both BD and MC data are obtained under 0.2 M monovalent ion concentration while the light-scattering data is obtained under 0.179 M monovalent salt.

under 0.2 M monovalent salt,[5] plotted in Fig. 4 as triangles. We measure R_g directly from the conformations in our BD trajectories for 5.28 kbp supercoiled DNA molecules under 0.2 M monovalent salt using the standard formula

$$R_g = \sqrt{\frac{\sum_{i=1}^{N} \mathbf{r}_i - \left(\sum_{i=1}^{N} \mathbf{r}_i / N\right)^2}{N}}, \quad (49)$$

where $\mathbf{r}_i, i = 1, \ldots, N$, is each particle's position. The BD data are plotted in Fig. 4 as filled circles. We see from the figure that all data agree reasonably well with each other.

We also calculated the persistence length of DNA as $p = 2l_0/\langle \theta_i^2 \rangle$. The result is 50 nm as expected.

B. Protein-induced directional bending compacts global DNA conformations as a function of σ

Protein binding directly induces inhomogeneous, local changes in the DNA substrate such as directional bends and altered flexibilities (e.g., Ref. 70). Proteins also change the mass distribution and the electrostatic potential near the binding sites. However, effects on the global properties of DNA are complex and not easily determined. Our modeling and simulation methods can help us systematically examine those effects.

Our system of a 5.28 kbp DNA with two *hix* sites separated by 990 bp of DNA provides a good model because of the availability of relevant protein/DNA atomic models based on crystal structure and other experiments.[23,26,42,43] We examine the conformational properties of the free DNA (before two *Hin* dimers bind on the *hix* sites), versus those after the binding. In the atomic structure for the complex of the *hix* site bound to *Hin* dimer from Johnson and co-workers,[43] the DNA segment is bent by 18° to widen the minor groove [Fig. 1(B)]. The binding of the *Hin* dimer also changes the electrostatic charge distribution near the *hix* site by making the site more bulky and less negatively charged due to charge shielding.

We first measure the radius of gyration R_g, which describes the overall size of the molecule, as a function of the DNA superhelicity before and after the binding of *Hin* dimers. Figure 5 shows results for free supercoiled DNA (filled circles) and DNA bound to *Hin*-dimers (squares). We note only a small but statistically significant effect on DNA's global conformations due to *Hin*-dimer binding. The radius of gyration slightly decreases after the binding of *Hin* dimers for relaxed DNA molecules or DNA molecules with a low level of negative supercoiling ($\sigma = 0 \sim -0.04$), while the radius gyration of the fully negative supercoiled DNA ($\sigma = -0.06$) is not notably affected. We can explain these trends by the shapes of DNA conformations themselves as a function of σ: at low superhelical density, protein effects can be larger since the DNA is more open, floppy, and random in shape, whereas at large σ the DNA is more tightly supercoiled (e.g., Refs. 16 and 17) and global distributions are overall smaller. The local charge screening of the polyelectrolyte backbone by the proteins also permits sites to move close to each other (Fig. 9), contributing to the decrease of R_g upon protein binding.
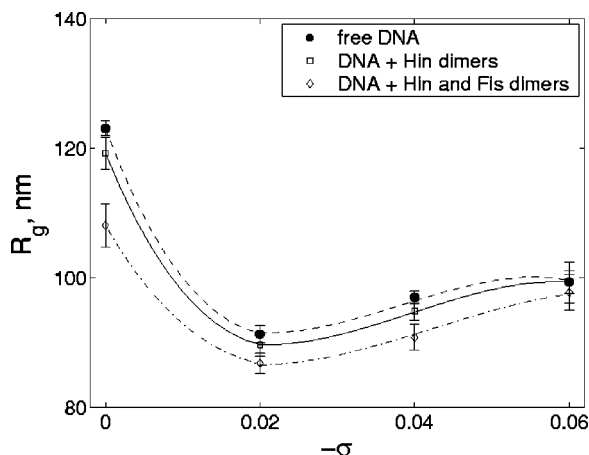
FIG. 5. Radius of gyration R_g as a function of DNA supercoiling (σ). Filled circles show the radius of gyration based on simulations of free, supercoiled DNA molecules. Squares show the same DNA molecule with two *hix* sites bound to *Hin* dimers. Diamonds show the same molecule with two *hix* sites bound to *Hin* dimers and two *Fis* binding sites in the enhancer sequence bound with *Fis* dimers. The values for N, i_{hix}^1, i_{hix}^2, i_{fis}^1, i_{fis}^2, and c_S are 176, 1, 34, 5, 6, 0.2 M, respectively.

To test this interpretation, we further study our *Hin*-bound DNA system when two *Fis* dimers are also added (as in Fig. 1). The positions of the enhancer sequence are i_{fis}^1 and i_{fis}^2 (see Ref. 20). A larger decrease in R_g can now be noted from Fig. 4. This effect is also illustrated in Fig. 1(B) where the compaction effect is evident. Namely, a single *Fis*-dimer bends the DNA segment by $\sim 60°$ to compress the minor groove at the center of the *Fis* binding site. Since the two *Fis* binding sites are located on the same enhancer sequence, the directions of these two bends are strongly correlated.[24] We thus expect a further decrease of R_g following *Fis*-dimer binding, with the decrease more pronounced at low $|\sigma|$. Figure 5 plots these data for *Hin*- and *Fis*-bound DNA as diamonds. At $\sigma = 0$, the value of R_g decreases 12% for the $Hin/Fis/$DNA system compared with free DNA; the *Hin*-dimer binding alone only caused 3% decrease. As $|\sigma|$ increases R_g decreases by a smaller relative value: at $\sigma = -0.06$, R_g of $Hin/Fis/$DNA only decreases by 2% compared to the DNA system; the effect of the *Hin*-dimer binding was negligible. These observations are consistent with our previous interpretations: protein-induced directional bending compacts global DNA conformations, with the effect more pronounced for more open conformations at low $|\sigma|$ than in more tightly supercoiled forms at higher $|\sigma|$.

C. Protein binding profoundly affects the DNA dynamics and juxtaposition processes

To quantitatively study the dynamic process of two *hix* sites moving into spatial vicinity, we define τ_{c1} as the average time for two sites in a randomly selected, equilibrated DNA conformation to move until the distance between two *hix* sites is smaller than the juxtaposition distance criterion d_0:

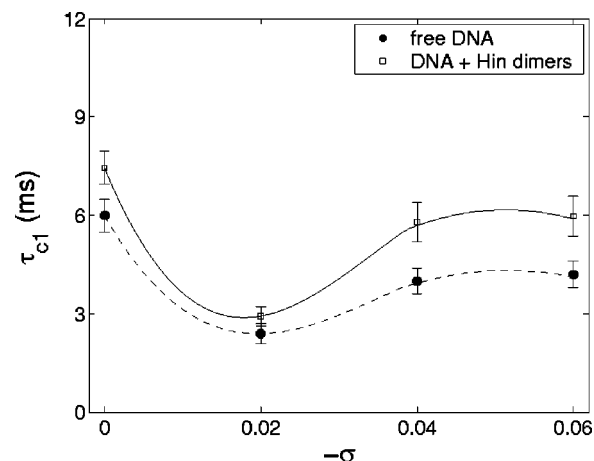$$|\mathbf{r}_{i_{hin}^1} - \mathbf{r}_{i_{hin}^2}| < d_0, \tag{50}$$

FIG. 6. Dependence of site juxtaposition times τ_{c1} for two *hix* sites on the level of DNA supercoiling, with or without protein binding. The values for N, i_{hix}^1, i_{hix}^2, and c_S are 176, 1, 34, 0.2 M, respectively.

where $\mathbf{r}_{i_{hin}^1}$ and $\mathbf{r}_{i_{hin}^2}$ are the coordinates of two *hix* sites [$i_{hin}^1 = 1$, $i_{hin}^2 = 34$ in our system, see Fig. 1(A)], and $d_0 = 100$ Å following the experimental system.[71] Such site-juxtaposition dynamics were previously studied in our work on protein-free DNA[13,14] to investigate reaction kinetics.

We perform simulations for both supercoiled DNA molecules and DNA molecules bound to two *Hin* dimers, based on the experimental system,[25] at various superhelicities to study effects of proteins on long-time processes like site juxtaposition. Figure 6 indicates that τ_{c1} slightly increases with the decrease of $|\sigma|$ for both free DNA and DNA molecules bound to *Hin* dimers. This is consistent with the experimental observation that low values of $|\sigma|$ slow down, but do not inhibit the pairing of *Hin* bound *hix* sites.[25]

Unlike the simpler effect on R_g, the DNA supercoiling level affects site juxtaposition kinetics in two opposing ways:[14] supercoiling increases the correlation between successive site juxtaposition events and thus decelerates the juxtaposition times, while at the same time higher $|\sigma|$ values compact the DNA, accelerate site juxtaposition reactions and thus produce lower times τ_{c1}. These two competing effects result in juxtaposition times that depend sensitively on the DNA system and conditions.

Comparing our simulations of DNA with or without bound *Hin* dimers, we find that in our system τ_{c1} is about 10%–25% smaller after *Hin*-dimers bind to the *hix* sites (Fig. 6). The juxtaposition process mainly involves two steps: First, two *Hin*-dimer bound *hix* sites come into proximity through large-scale conformational evolution such as slithering; second, two sites juxtapose (come within <100 Å) through a local diffusion process.

We next try to interpret this effect of *Hin*-dimer binding on the juxtaposition process of *hix* sites by inspecting these two steps systematically. The large-scale conformational changes include the slithering motions on the branch and the formation as well as deletion of branches.[14] We illustrate the juxtaposition evolution of a particular site, i_1, with other sites along the DNA as a time series in Fig. 7. A data point $\{\tau, i_2\}$ means that site i_1 is juxtaposed with site i_2 at time τ ($|\mathbf{r}_{i_1} - \mathbf{r}_{i_2}| < d_0$). The dependence of i_2 on time contains
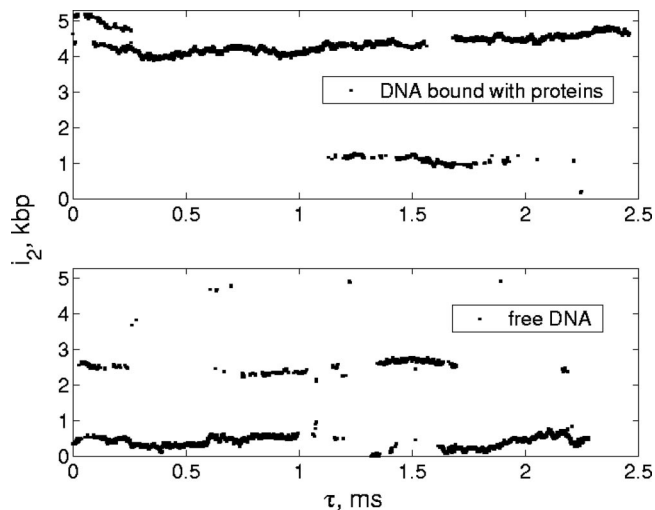
FIG. 7. Juxtaposition of one *hix* site (i_1) with sites of the DNA as a function of time. The coordinate of the second site, i_2, is recorded over one BD trajectory. Simulations for DNA molecules 5.28 kbp in length are performed both for DNA with bound *Hin* dimers (upper panel), and without proteins (bottom panel). Both trajectories employ $\sigma = -0.06$ and $c_S = 0.2$ M monovalent salt.

ample information on the dynamics of juxtaposition. In our case, we seek to understand how two *hix* sites approach one another and how this process is affected by protein binding. We thus choose for our analyses $i_1 = i_{hin}^1 = 1$ (i.e., the first *hix*

site) for the simulations of a free, 5.28 kbp supercoiled DNA and the corresponding system with both *hix* sites bound to *Hin* dimers (recall the parameters for superhelical density and the salt concentration are $\sigma = -0.06$, $c_S = 0.2$ M).

The plots in Fig. 7 show that for this highly supercoiled DNA (the average writhing number is ~ -22)[5] the time evolution of i_2 is strongly correlated to past values in both cases. The supercoiled DNA molecules under such conditions adopt tightly interwound conformations (see BD snapshots in Fig. 8). Thus, the slow slithering motions of the opposing DNA segments on the interwound conformations continuously change i_2 as shown in both Figs. 7 and 8. The site i_1 sometimes moves through a branch point, and this causes the jump from one continuous curve to another in the i_2 time evolution plots. These jumps are infrequent; the slithering motions on one branch continues on the millisecond time scale.

To quantitatively assess the effects of *Hin*-dimer binding on the *hix*–*hix* juxtaposition, we compute the autocorrelation function of $i_2(t)$. The normalized autocorrelation function, c_{i_2}, of $i_2(t)$ is

$$c_{i_2}(\Delta t) = \frac{\langle i_2(t+\Delta t) \cdot i_2(t)\rangle - \langle i_2(t)\rangle^2}{\sigma_{i_2}^2}, \quad (51)$$

where $\sigma_{i_2}^2$ is the standard variance of $i_2(t)$. We fit the autocorrelation function to an exponential form of $\exp(-t/\tau_{i_2})$
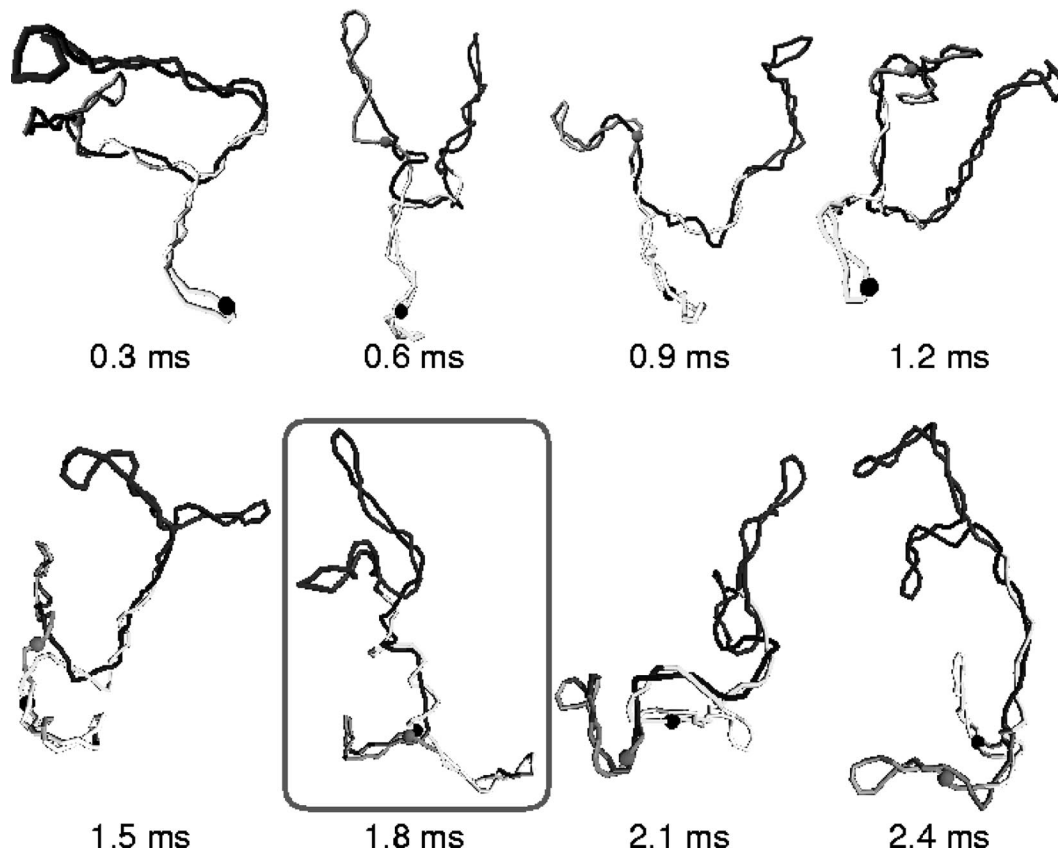

FIG. 8. Dynamics of the site juxtaposition in supercoiled DNA. BD snapshots of supercoiled DNA molecules bound with *Hin* dimers represent the part of the simulation trajectory shown in Fig. 7 (upper panel). The two *hix* sites ($i_{hin}^1 = 1$, $i_{hin}^2 = 34$) are shown by the gray and black spheres. The actual simulated orientation of the molecule is rotated to enhance the connection between successive conformations.
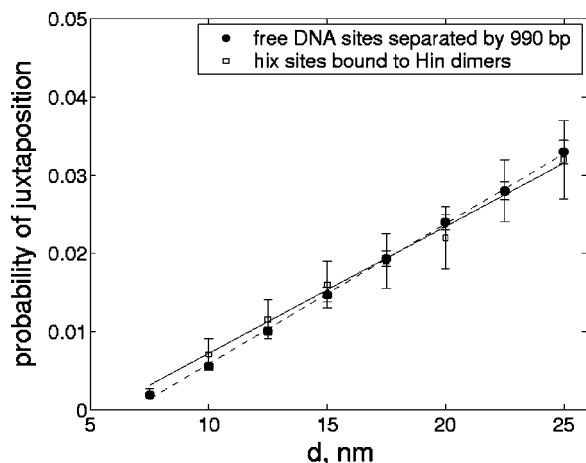
FIG. 9. Juxtaposition probability of two *hix* sites as a function of the distance d between two sites: before and after *Hin*-dimer binding. The juxtaposition probabilities are measured for all site pairs separated by 990 bp. The number of total sites is 176, the superhelical density is -0.06, and the monovalent salt concentration is set to 0.2 M.

and estimate τ_{i_2} based on two sets of trajectories for free and protein-bound DNA. For highly supercoiled DNA ($\sigma = -0.06, c_S = 0.2$ M), we obtain $\tau_{i_2} = 154 \pm 49$ μs for the free 5.28 kbp DNA system and $\tau_{i_2} = 222 \pm 78$ μs for the DNA bound to two *Hin* dimers. This rate characterizes the correlation between juxtaposition events, which reflect both the slithering motion within a branch and the intrahelix collisions. Our results therefore indicate that the binding of two *Hin* dimers to *hix* sites slightly increases the correlation between juxtaposition events.

When the distance between the two *Hin*-dimer bound *hix* sites is on the order of several Debye lengths or less, electrostatic effects by the *Hin* dimers can screen the charge and alter the dynamics. At the 0.2 M monovalent salt concentration, the effective Debye–Hückel charge of a 30 bp DNA segment is -99.4 e. In our case, the sum of effective charges for the *Hin*-dimer bound *hix* site is -18.0 e at 0.2 M monovalent salt [Fig. 1(B)]. Even though the electrostatic potential decreases sharply with the increase of distance under this high salt environment, we can expect less repulsion between two *Hin*-dimer bound *hix* sites and a lower free energy required to pair two *Hin*-dimer bound *hix* sites compared with the free *hix* sites.

We quantitatively measure this effect by computing the juxtaposition probability of the two *hix* sites as a function of the distance between them. For simulations of DNA bound to *Hin* dimers, recall that the juxtaposition probability is defined as the probability for finding two *hix* sites separated by less than or equal to some distance criterion d. Our study of the dependence of the juxtaposition probability as a function of d can be interpreted as measuring the relative local concentration of one *hix* sites with respect to another. For simulations of free DNA we can define the juxtaposition probability similarly, as the probability of finding any site pairs that are linearly separated by 990 bp (the length as between two *hix* sites) at a distance smaller than or equal to d. The results in Fig. 9 are plotted for d values ranging from 7.5 to 25 nm. The error bars for the free DNA data are much smaller be-

cause we can sample 176 data pairs (N pairs of sites with 990 bp separation, $N = 176$) from a free supercoiled DNA trajectory but only one data point from a protein-bound trajectory. Even though the statistical errors for the DNA bound to *Hin* dimers are significant, Fig. 9 shows that the juxtaposition probabilities are slightly higher after *Hin*-dimer binding when the distance between two *hix* sites, d, is in the range of 7.5–15 nm. When the two sites are further separated by $d > 15$ nm, the effect of *Hin*-dimer binding on the juxtaposition probability of *hix* sites becomes negligible.

Because the distance between the molecular surface to the mass center is in the range of 2 nm to 5 nm [Fig. 1(B)], the actual distance between a pair of effective charges located on two *Hin*-dimer complexes should be 0 to 17.5 nm if $d = 7.5$ nm or 5 to 25 nm if $d = 15$ nm. Given that the electrostatic potential decreases exponentially with the Debye length (which is 0.68 nm at 0.2 M monovalent salt condition), the electrostatic interaction is negligible if the distance between charges is larger than 10 times of the Debye length, i.e., 6.8 nm. That means that the effect is noticeable for $d = 7.5$ nm, where many pairs of effective charges on two *Hin*-dimer complexes might be located very near each other. This effect decreases with d because less pairs of effective charges tend to come into spatial proximity as the two complexes separate.

In sum, we find that in the *Hin*-dimer bound DNA system, the evolution of DNA global conformations is slightly slower with respect to free DNA but that the juxtaposition process becomes faster after two *hix* sites approach one another. Net effects of *Hin*-dimer binding (Fig. 6) result from the balance of these two major contributions.

IV. CONCLUSION

We have developed a polymer-level computational model and simulation protocols for supercoiled DNA molecules with protein bound sites. Our model incorporates the electrostatic and structural properties of DNA molecules with bound proteins based on the well established discrete wormlike chain model for supercoiled DNA. The model reflects the electrostatic force associated with both DNA and protein/DNA complexes and refines standard mechanical and hydrodynamic representations. The model thus efficiently integrates atomic-level details of macromolecules into a reasonable description on the polymer-level to treat a system too large to be modeled on the atomic scale (thousands of DNA basepairs). Based on atomic-level structures, we have carefully parameterized the model, to define positions and values of effective charges, directional bending angles and rigidities, effective hydrodynamic radii, and excluded-volume parameters. The model was validated through its reproduction of translational diffusion coefficients and radius of gyrations as measured by Monte Carlo and light-scattering experiments.

Our motivation for the development of these methods is to investigate dynamic aspects of the complex *Hin*-mediated inversion process. Here, we have analyzed the juxtaposition dynamics between two *hix* sites before and after the protein binding. We find that the local protein binding affects the global conformations of DNA but also the global dynamics.

In particular, the global shapes are more compact (Fig. 5). The binding of *Hin* dimers to the *hix* sites slows down the slithering process and increases the site juxtaposition probabilities (Fig. 9). We will present a more detailed study of the *Hin*-mediated inversion process in a future work.

The modeling and simulation methods developed here are general. They can thus be applied to a broad range of problems in which an efficient, polymer-level description for the long-time dynamics (millisecond to second) of DNA molecules with bound proteins in solution is desired, such as resolvase-DNA and integrase-DNA system.[2,24]

[1] A. Travers, Annu. Rev. Biochem. **58**, 427 (1989).
[2] *DNA-Protein: Structural Interactions*, edited by D. Lilley (Oxford University Press, New York, 1995).
[3] C. Bustamante and C. Rivetti, Annu. Rev. Biophys. Biomol. Struct. **25**, 395 (1996).
[4] J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, Proc. Natl. Acad. Sci. U.S.A. **53**, 1104 (1965).
[5] A. Vologodskii and N. Cozzarelli, Annu. Rev. Biophys. Biomol. Struct. **23**, 609 (1994).
[6] *DNA Topology*, edited by A. D. Bates and A. Maxwell (Oxford University Press, New York, 1993).
[7] R. Kannar and N. Cozzrelli, Curr. Opin. Struct. Biol. **2**, 369 (1992).
[8] G. Bellomy and M. T. Record, Jr., Prog. Nucleic Acid Res. Mol. Biol. **39**, 81 (1990).
[9] J. Gralla, Curr. Opin. Genet. Dev. **6**, 526 (1996).
[10] R. Tjian and T. Maniatis, Cell **77**, 5 (1994).
[11] M. Watson, M. Boocock, and W. Stark, J. Mol. Biol. **257**, 317 (1996).
[12] S. Allison, R. Austin, and M. Hogan, J. Chem. Phys. **90**, 3843 (1989).
[13] H. Jian, T. Schlick, and A. Vologodskii, J. Mol. Biol. **284**, 287 (1998).
[14] J. Huang, T. Schlick, and A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **98**, 968 (2001).
[15] A. Vologodskii, S. Levene, K. Klenin, M. Frank-Kamenetski, and N. Cozzarelli, J. Mol. Biol. **227**, 1224 (1992).
[16] T. Schlick and W. Olson, J. Mol. Biol. **223**, 1089 (1992).
[17] T. Schlick, B. Li, and W. Olson, Biophys. J. **67**, 2146 (1994).
[18] M. Hammermann, C. Steinmaier, H. Merlitz, U. Kapp, W. Waldeck, G. Chirico, and J. Langowski, Biophys. J. **73**, 2674 (1997).
[19] M. Silverman and M. Simon, Cell **19**, 845 (1980).
[20] R. Johnson and M. Simon, Cell **41**, 781 (1985).
[21] K. Heichman and R. Johnson, Science **249**, 511 (1990).
[22] H. Jian, Ph.D. thesis, New York University, 1997.
[23] J. Feng, R. Johnson, and R. Dickerson, Science **263**, 348 (1994).
[24] *Mobile DNA*, edited by D. Berg and M. Howe (American Society for Microbiology, Washington, D.C., 1989), pp. 637–659.
[25] H. Lim and M. Simon, J. Biol. Chem. **267**, 11176 (1992).
[26] M. Safo, W. Yang, L. Corselli, S. Cramton, H. Yuan, and R. Johnson, EMBO J. **16**, 6860 (1997).
[27] H. Huber, S. Iida, W. Arber, and T. Bickle, Proc. Natl. Acad. Sci. U.S.A. **82**, 3776 (1985).
[28] R. Johnson, A. Glasgow, and M. Simon, Nature (London) **329**, 462 (1987).
[29] *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (American Society for Microbiology, Washington, D.C., 2002), pp. 230–271.
[30] M. J. Haykinson and R. C. Johnson, EMBO J. **12**, 2503 (1993).
[31] J. Huang, Ph.D. thesis, New York University, 2002.
[32] B. Stocker, J. Hyg. (London) **47**, 398 (1949).
[33] G. Chirico and J. Langowski, Biopolymers **34**, 415 (1994).
[34] P. Hagerman, Annu. Rev. Biophys. Bioeng. **17**, 265 (1988).
[35] C. Baumann, S. Smith, V. Bloomfield, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **94**, 6185 (1997).
[36] D. Crothers, J. Drak, J. Kahn, and S. Levene, Methods Enzymol. **212**, 3 (1992).
[37] K. Klenin, A. Vologodskii, V. Anshelevich, V. Klishko, A. Dykhne, and M. Frankkamenetskii, J. Biomol. Struct. Dyn. **6**, 707 (1989).
[38] S. Allison, Macromolecules **19**, 118 (1986).
[39] R. Dickerson, J. Mol. Biol. **208**, 1797 (1989).
[40] M. Bruist, A. Glasgow, R. Johnson, and M. Simon, Genes Dev. **1**, 762 (1987).
[41] A. Glasgow, M. Bruist, and M. Simon, J. Biol. Chem. **264**, 10072 (1989).
[42] M. Haykinson, L. Johnson, J. Soong, and R. Johnson, Curr. Biol. **6**, 163 (1996).
[43] S. Merickel, M. Haykinson, and R. Johnson, Genes Dev. **12**, 2803 (1998).
[44] D. Perkins-Balding, D. Dias, and A. Glasgow, J. Bacteriol. **179**, 4747 (1997).
[45] C. Pan, S. Finkel, S. Cramton, J. Feng, D. Sigman, and R. Johnson, J. Mol. Biol. **264**, 675 (1996).
[46] T. Kerppola and T. Curran, Mol. Cell. Biol. **13**, 5479 (1993).
[47] D. Beard and T. Schlick, Biopolymers **58**, 106 (2001).
[48] D. Beard and T. Schlick, Structure (London) **9**, 105 (2001).
[49] W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, Jr., D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell, and P. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[50] K. Chin, K. A. Sharp, B. Honig, and A. M. Pyle, J. Struct. Biol. **6**, 1055 (1999).
[51] D. Xie and T. Schlick, ACM Trans. Math. Softw. **25**, 108 (1999).
[52] J. White, Am. J. Math. **91**, 693 (1969).
[53] A. Initesta and J. Garcia de la Torre, J. Chem. Phys. **92**, 2015 (1990).
[54] H. Jian, A. Vologodskii, and T. Schlick, J. Comput. Phys. **136**, 168 (1997).
[55] D. Ermak, J. Chem. Phys. **69**, 1352 (1978).
[56] G. Dahlquist and Å. Björck, *Numerical Methods* (Prentice Hall, Englewood Cliffs, NJ, 1974).
[57] M. Fixman, Macromolecules **19**, 1204 (1986).
[58] T. Schlick, D. Beard, J. Huang, D. Strahs, and X. Qian, IEEE Comput. Sci. Eng. **2**, 38 (2000) (special issue on Computational Chemistry).
[59] M. Kröger, A. Alba-Perez, M. Laso, and H. C. Öttinger, J. Chem. Phys. **113**, 4767 (2000).
[60] R. M. Jendrejack, M. D. Graham, and J. J. de Pablo, J. Chem. Phys. **113**, 2894 (2000).
[61] J. Rotne and S. Prager, J. Chem. Phys. **50**, 4831 (1969).
[62] J. Garcia de la Torre and V. Bloomfield, Biopolymers **16**, 1747 (1977).
[63] V. Rybenkov, A. Vologodskii, and N. Cozzarelli, J. Mol. Biol. **267**, 299 (1997).
[64] J. Langowski and U. Giesen, Biophys. Chem. **34**, 9 (1989).
[65] J. Garcia de la Torre, M. Huertas, and B. Carrasco, Biophys. J. **78**, 719 (2000).
[66] V. Bloomfield, W. Dalton, and K. van Holde, Biopolymers **5**, 135 (1967).
[67] E. Swanson, D. Teller, and C. De Haen, J. Chem. Phys. **68**, 5097 (1978).
[68] B. Carrasco and J. Garcia de la Torre, Biophys. J. **75**, 3044 (1999).
[69] A. Campbell and R. Eason, FEBS Lett. **55**, 212 (1975).
[70] D. Strahs, D. Barash, X. Qian, and T. Schlick, Biopolymers (submitted).
[71] M. Hammermann, N. Brun, K. Klenin, R. May, K. Toth, and J. Langowski, Biophys. J. **75**, 3057 (1998).