Statistics of loop formation along double helix DNAs

Jie Yan, Ryo Kawamura, and John F. Marko
Department of Physics, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607-7059, USA
(Received 19 December 2004; published 13 June 2005)

We compute relative position distributions of distant sites along discretized semiflexible polymers, focusing on encounter statistics for pairs of sites along a double-stranded DNA molecule (dsDNA), using a transfer-matrix approach. We generalize the usual semiflexible polymer, considering nonlinear elasticity effects arising from inhomogeneities which either appear at any position via thermal fluctuation, or which occur at specific "quenched" locations. We apply our theory to two problems associated with dsDNA looping. First, we discuss how local flexible defects in double-helix structure facilitate cyclization of short dsDNA molecules. Flexible defects greatly enhance cyclization rate, and strongly modify its dependence on the closure orientational boundary condition. This effect is relevant to free-solution cyclization experiments, and to loop formation *in vivo*. Second, we present calculations of force dependence of the probability of formation of loops along single dsDNAs which show how the probability of loop formation is suppressed by tension.

DOI: 10.1103/PhysRevE.71.061905 PACS number(s): 87.14.Gg, 82.35.Pq, 87.15.Aa, 87.15.La

I. INTRODUCTION

The long double-stranded DNA (dsDNA) molecules in the chromosomes of living cells exist in an intermediate state of order interesting to consider using statistical mechanics. Chromosomal DNAs—often centimeters in length—are spatially ordered in the cell [1]. Some of this order is due to protein-mediated associations of distant loci along chromosomal DNAs, as occurs during, e.g., gene regulation [2,3], site-specific DNA "cut-and-paste" recombination [4], and general recombination [5]. These chromosomal events require formation of *crosslinks* or *synapses* of dsDNA molecules, generally mediated by proteins which bind the loop boundaries (Fig. 1).

If association of two sites along the same dsDNA molecule occurs, a dsDNA loop is formed, with the possibility that the intervening dsDNA is free to fluctuate thermally. A number of experiments suggest that at submicron scales, the dynamics of dsDNA in cells is dependent on conformational diffusion [6]. Formation of loops along a dsDNA via collisions of sites brought together by conformational diffusion is an important mechanism behind "higher-order" organization of dsDNA in vivo [7,8], and can be described using polymer statistical mechanics [9].

Looping dynamics of dsDNAs in the cell involve a wide range of proteins, for example DNA-bending proteins which facilitate dsDNA loops and other "higher-order" nucleoprotein structures [10–12]. Statistical-mechanical models for higher-order DNA organization *in vivo*, while valuable, therefore tend toward the qualitative, since there are so many unknowns. However, precise quantitative experimental study of the proteins which mediate higher-order organization of double helix DNA is commonplace in biochemistry laboratories. For example, recent single-molecule micromanipulation studies have revealed many new aspects of the mechanics and even basic biochemistry of DNA-bending proteins [13–16]. These precise biophysical characterizations of interactions of purified proteins with DNA motivate theoretical study of DNA organization by proteins, and are first steps toward building models of *in vivo* chromosome dynamics.

Two examples of DNA-looping proteins being studied using single-molecule approaches are the LacR and GalR transcriptional repressors, which bind to specific DNA sequences, and which can link two binding sites along a single dsDNA. Finzi and Gelles observed formation and dissolution of LacR-mediated-loops along single dsDNA molecules in a classic study [17]. More recent micromanipulation experiments [18] on GalR studied the role of DNA supercoiling and the DNA-bending protein HU on the formation of a loop between two GalR binding sites.

Another recent single-molecule experiment used the restriction enzyme BspMI, which binds to two copies of its specific recognition site 5′-ACCTGC, and therefore can form a loop between two sites along a single dsDNA [19]. Loop-opening events were observed when a BspMI-looped λ-DNA was put under large tension. Finally, we note that loop formation along single-stranded nucleic acids(ssNA) has also been studied, where the loops are stabilized by sequence-defined base-pairing interactions [20,21], thus not requiring protein. In those ssNA experiments, the kinetics of loop opening and closure as a function of applied tension could be rather precisely characterized. These studies show that loop closing and opening along individual biomolecules can be quantitatively studied using micromanipulation experiments.

Another motivation for theoretical study of loop formation comes from studies of circularization ("cyclization") of linear dsDNAs. Cyclization rate measurements have long been used to study bending of dsDNA [22]. This approach became widely used to study dsDNA flexibility [23], and is now being used as a tool to study spontaneously bent dsDNA [24,25], and chemical factors which affect DNA conformation, such as DNA-bending proteins [26].

A recent study by Cloutier and Widom [27] examined shorter dsDNA molecules than usually studied, with the result that anomalously large cyclization probabilities were observed, relative to those expected theoretically from the simple semiflexible polymer model [28–30]. Those experiments suggest that very tight dsDNA bends may occur more frequently than expected from the semiflexible polymer

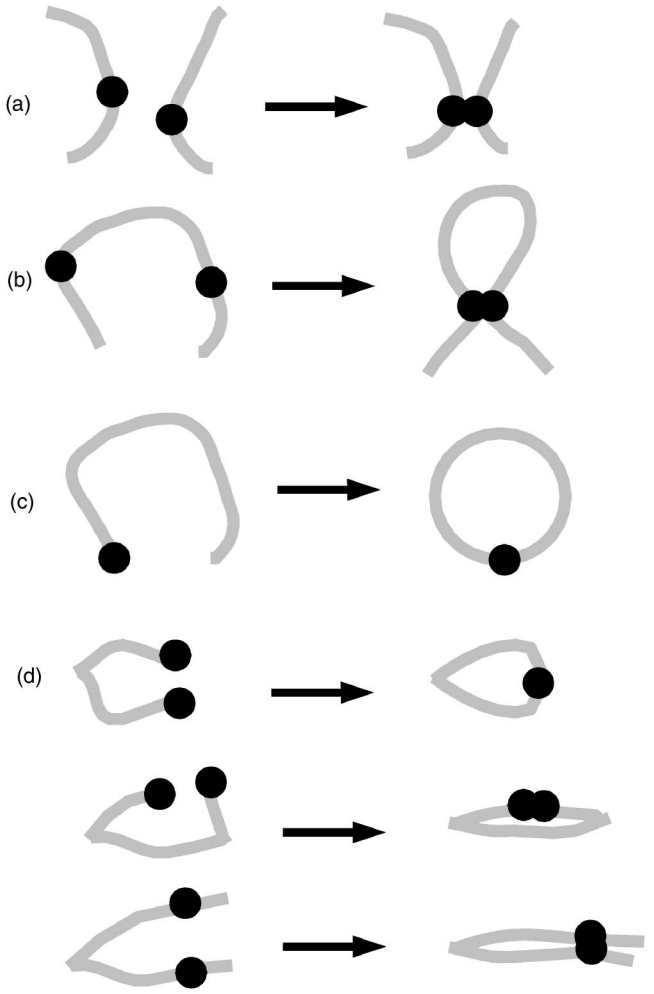

FIG. 1. dsDNA synapsis and loop formation. (a) Synapsis between two dsDNA segment (gray curves), mediated by protein-DNA complexes (black circles). (b) Loop formation mediated by proteins bound at two specific positions along a dsDNA molecule; the boundary condition in this sketch has $a \approx 90°$ synapsis angle between the interacting sites. (c) Cyclization of a linear dsDNA molecule, by the action of the enzyme DNA ligase (black circle), which is thought to require a "parallel-end" boundary condition. (d) Role of localized sharp bends in facilitating dsDNA looping for parallel-end and antiparallel-end boundary condition.

model. These experiments have triggered theoretical study of how nonlinear elasticity of the double helix can generate the experimental results, with particular attention on the possibility that localized, flexible defects in double helix structure may play an important role in loop formation along short dsDNAs [31,32].

In this paper, we study DNA loop formation theoretically. We calculate distribution functions for the difference in position of any two points along a discretized polymer, under zero or nonzero tension. Our transfer matrix approach is similar to that used in our previous paper studying DNA-bending proteins [33]. It describes the positional statistics for a broad class of polymer models including the discretized semiflexible polymer and variants on it with nonlinear elasticity. One type of nonlinear elasticity originates from thermally fluctuating localized defects, which might arise from the action of drugs or proteins which induce "permanent" (or "spontaneous") bends [33], or thermally-excited regions of modified flexibility [31]. Our calculations also allow us to introduce inhomogeneities in shape or flexibility at fixed positions along a polymer chain, describing, e.g., sequence variation of bending elasticity, or effects of proteins bound at specific sequence sites.

Section II introduces our calculation method. We compute the statistical distribution of the end-to-end vector, using transfer matrices analytically expressed in a basis of spherical harmonics. For zero force, we obtain analytical results for the tangent vector correlation function, and therefore the persistence length and mean squared end-to-end vector.

We then discuss two applications to specific experimental situations. Section III uses our theory to compute dsDNA cyclization probabilities (in the DNA cyclization literature, the "J factor") in the presence of thermally excited flexible defects. These spontaneously excited localized flexible defects may provide an explanation of the anomalously large short-dsDNA cyclization rates observed by Cloutier and Widom [27], or alternately are a model for proteins which facilitate double-helix bending by increasing its flexibility at points where they bind. We show how these defects can statistically dominate the cyclized states, and we show how changes in the cyclization boundary conditions drastically affect the J factor for short dsDNA segments.

If localized flexible defects are present, then for roughly 100 base-pair (bp) segments cyclized with a parallel-end boundary condition [Fig. 1(c)] *two* localized, hairpinlike defects dominate the cyclized states, essentially eliminating the need for any other bending [Fig. 1(d)]. We also show that if the synapsis boundary condition is changed to be *free* or *antiparallel* only one localized defect is necessary, enhancing the cyclization probability for very short dsDNA segments. For the antiparallel boundary conditions, the prominent peak in cyclization probability for dsDNA lengths near 500 bp seen in the classical harmonic-bending-energy theory [29] is entirely absent, with instead a monotonic increase in cyclization probability occuring as one goes to progressively shorter molecular lengths. This result has strong implications for *in vivo* dsDNA bending, where dsDNA loops or folds on short length scales are indeed often seen to occur over very short sequence distances.

We also compute the effect of permanent bends on dsDNA cyclization probability. Such permanent bends can be created by DNA-bending proteins that bind to specific sequences.

The second main application of our theory, to loop formation in dsDNA molecules under tension, is discussed in Sec. IV. We compute the end-to-end vector distribution of a dsDNA segment under force. The longitudinal and transverse distributions are nearly Gaussian for long (many kilobase) molecules. Non-Gaussian corrections become measurable for molecules in the kilobase range, the size often studied in single-molecule experiments on DNA transcription [34]. Our main result is a computation showing how the probability of dsDNA loops of roughly kilobase (kb) size is suppressed by applied force. This effect should be directly measurable in experiments on site-specific loop-forming proteins.

II. TRANSFER-MATRIX CALCULATION OF END-TO-END DISTRIBUTIONS

We now compute the end-to-end vector distribution, using a transfer matrix approach. We begin with the usual semiflexible polymer (Sec. II A), including various boundary conditions (Sec. II B), for free, unstretched chains (Sec. II C), and and chains subjected to applied tension (Sec. II D). We then generalize our model to include effects of localized "defects," e.g., flexible joints and spontaneous bends (Sec. II E). These localized defects are special cases of general, nonlinear elasticity; the general case is not difficult to analyze (Sec. II E 3). For zero applied force, the orientational correlation functions of all these models are of the simple exponential form of the simple semiflexible polymer (Sec. II E 4), characterized by a single "persistence length."

A. End-to-end vector distribution of discretized semiflexible polymer

The energy of a continuum semiflexible polymer of contour length L including the effect of an applied external force in the $\hat{\mathbf{z}}$ direction, $\mathbf{f}=f\hat{\mathbf{z}}$, is

$$\beta E = \int_0^L ds\left[\frac{A}{2}\left(\frac{d\hat{\mathbf{t}}}{ds}\right)^2 - f\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}\right], \tag{1}$$

where the polymer is described as a space curve: s is the arc length parametrizing the curve, and $\hat{\mathbf{t}}$ is the unit tangent vector describing the curve conformation. A is the persistence length, characterizing the bending rigidity of the polymer. For $f=0$, over contour distance A the tangent vector orientation becomes decorrelated. We consider the discretization of this model,

$$\beta E = \sum_{i=1}^{N-1}\left[\frac{a}{2}(\hat{\mathbf{t}}_{i+1}-\hat{\mathbf{t}}_i)^2 - bf\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}_i\right], \tag{2}$$

where the $\hat{\mathbf{t}}_i$ are tangent vectors describing orientations of the successive segments of length b. The force f is in k_BT units (dimensions 1/length). In the continuum limit $b\to 0$, the bending elastic constants of the continuum and discrete models are related by $a=A/b$.

The probability density $\rho(\mathbf{R})$ for the end-to-end distance $\mathbf{R}$ of this polymer is given by the average of $\xi(\hat{\mathbf{t}}_1,\hat{\mathbf{t}}_N)\delta^3(\mathbf{R}-b\Sigma_{j=1}^{N-1}\hat{\mathbf{t}}_j)$ over all the chain conformations, where ξ is a function that imposes a specific boundary condition. Decomposing the three-dimensional delta function into wave number components, and properly normalizing the average gives

$$\rho(\mathbf{R}) = \int \frac{d^3k}{(2\pi)^3} \times e^{-i\mathbf{k}\cdot\mathbf{R}} \frac{\int d^2\hat{t}_1\cdots d^2\hat{t}_N \xi \exp[-\beta E + ib\mathbf{k}\cdot\sum_{j=1}^{N-1}\hat{\mathbf{t}}_j]}{\int d^2\hat{t}_1\cdots d^2\hat{t}_N e^{-\beta E}}. \tag{3}$$

The numerator counts the number of configurations of the polymer subject to the constraint imposed by the delta function. The denominator provides normalization, and is just the total partition function for the linear chain with no constraint applied. Thus (3) is the normalized probability density for the end-to-end vector of the chain.

The expectation value (3) may be written using a $\mathbf{k}$-dependent transfer matrix $T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}')=e^{ib\mathbf{k}\cdot\hat{\mathbf{t}}}e^{-a(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2}e^{bf\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}}$ as

$$\rho(\mathbf{R}) = \int \frac{d^3k}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{R}} \frac{\int d^2\hat{t}_1 d^2\hat{t}_N \xi T_{\mathbf{k}}^{N-1}}{\int d^2\hat{t}_1 d^2\hat{t}_N T_{\mathbf{0}}^{N-1}}, \tag{4}$$

where the operation of matrix multiplication is taken to indicate integration over tangent vectors $\hat{\mathbf{t}}_i$. Denoting $Z_{\mathbf{k}}=\int d^2\hat{t}_1 d^2\hat{t}_N \xi T_{\mathbf{k}}^{N-1}$ and $Z=\int d^2\hat{t}_1 d^2\hat{t}_N T_{\mathbf{0}}^{N-1}$, (4) becomes

$$\rho(\mathbf{R}) = \frac{Z(\mathbf{R})}{Z}, \tag{5}$$

where $Z(\mathbf{R})=\int d^3k e^{-i\mathbf{k}\cdot\mathbf{R}}Z_{\mathbf{k}}/(2\pi)^3$ is the partition function subject to the constraint of end-to-end distance of $\mathbf{R}$ and boundary condition ξ, and where Z is the total partition function without any constraint.

The matrix $T_{\mathbf{k}}$ can be computed in the basis of spherical harmonics as

$$\begin{aligned}\langle lm|T_{\mathbf{k}}|l'm'\rangle &= \int d^2\hat{t}d^2\hat{t}' Y_{lm}^*(\hat{\mathbf{t}})T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}')Y_{l'm'}(\hat{\mathbf{t}}')\\ &= (4\pi)^{3/2}(-1)^{m'-m}\sum_{l_1,l_2,l_3} i^{l_3}(2l_1+1)(2l_2+1)\\ &\times\sqrt{(2l+1)(2l'+1)(2l_3+1)}\begin{pmatrix} l & l_2 & l_1\\ m & 0 & -m\end{pmatrix}\\ &\times\begin{pmatrix} l & l_2 & l_1\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l' & l_3 & l_1\\ -m' & m'-m & m\end{pmatrix}\\ &\times\begin{pmatrix} l' & l_3 & l_1\\ 0 & 0 & 0\end{pmatrix}e^{-a}i_{l'}(a)i_{l_2}(bf)j_{l_3}(bk)\\ &\times Y_{l_3,m'-m}(\hat{\mathbf{k}}),\end{aligned} \tag{6}$$

using spherical harmonic expansions for the exponential functions in $T_{\mathbf{k}}$, and expressing all integrals of spherical harmonics in terms of Wigner-3J symbols [35]. Here j_l and i_l are spherical Bessel functions and modified spherical Bessel functions of the first kind, respectively.

The transfer matrix for $\mathbf{k}=\mathbf{0}$, $T_{\mathbf{0}}$, used to compute the partition function in the denominator, has a simpler form:

$$\begin{aligned}\langle lm|T_{\mathbf{0}}|l'm'\rangle &= 4\pi(-1)^m\delta_{mm'}\sum_{l_2}(2l_2+1)\sqrt{(2l+1)(2l'+1)}\\ &\times\begin{pmatrix} l_2 & l & l'\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l_2 & l & l'\\ 0 & m & -m\end{pmatrix}e^{-a}i_{l'}(a)i_{l_2}(bf).\end{aligned} \tag{7}$$

When the force f is zero, (6) is simplified:

$$\langle lm|T_{\mathbf{k}}|l'm'\rangle = 4\pi(-1)^m \delta_{mm'} \sum_{l_2} i^{l_2}(2l_2+1)\sqrt{(2l+1)(2l'+1)} \times \begin{pmatrix} l_2 & l & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_2 & l & l' \\ 0 & m & -m \end{pmatrix} e^{-a} i_{l'}(a) j_{l_2}(bk). \tag{8}$$

For $f=0$, $\langle lm|T_{\mathbf{0}}|l'm'\rangle = 4\pi e^{-a} i_l(a)\delta_{ll'}\delta_{mm'}$; the total partition function is $Z=[4\pi e^{-a} i_0(a)]^{N-1}$.

For our calculations, the matrix multiplications and the integral over **k** are done numerically. Cutoffs of matrix dimension and **k** must be chosen large enough that the calculations are convergent. It is straightforward to incorporate quenched disorder into the calculations, by multiplying a series of matrices that represents the sequence of disorder.

The same approach can be used to calculate the end-to-end distribution of an arbitrary segment inside a long polymer by considering $T_{\mathbf{k}}^{N-1}$ in (4) to be a site-dependent matrix multiplication. This is done by setting the k of $j_{l_3}(bk)$ in (6) to zero when outside the segment of interest.

B. Orientational boundary conditions

We will compute $Z(\mathbf{R})$ with a few different types of boundary conditions imposed on the ends. Different boundary conditions provide models for different types of biochemical reactions that stabilize dsDNA loops. Generally, dsDNA loops mediated by a certain protein or protein complex might require the two ends to meet at a certain angle. In this type of situation we wish to compute $Z_{\mathbf{k}}$ subject to a constraint function ξ which counts only end meetings with certain orientations.

A simple example discussed in Sec. III is cyclization of dsDNA by the enzyme T4 DNA ligase, which requires the two ends to meet one another aligned end-to-end [27]. We will refer to this as *parallel* boundary condition, corresponding to the choice $\xi=\delta^2(\hat{\mathbf{t}}_1,\hat{\mathbf{t}}_N)$ in (3) and (4).

For other enzymes, the two ends might be required to meet in *antiparallel* alignment, so as to form a "hairpin" structure. This boundary condition is expressed by $\xi=\delta^2(\hat{\mathbf{t}}_1,-\hat{\mathbf{t}}_N)$. More generally, we might have the requirement that the two ends meet at a *fixed relative angle* α, requiring $\xi=\delta(\hat{\mathbf{t}}_1\cdot\hat{\mathbf{t}}_N-\alpha)$.

These boundary conditions are handled by conversion of the constraint function ξ to a matrix acting on the end orientations:

$$Z_{\mathbf{k}} = \sum_{l_1,m_1}\sum_{l_N,m_N}\int d^2\hat{t}_1 d^2\hat{t}_N \xi\langle\hat{\mathbf{t}}_1|l_1m_1\rangle\langle l_1m_1|T_{\mathbf{k}}^{N-1}|l_Nm_N\rangle \times\langle l_Nm_N|\hat{\mathbf{t}}_N\rangle. \tag{9}$$

The **k**-dependent partition function $Z_{\mathbf{k}}$ reduces to the following for the cases listed above:

$$\text{parallel:}\quad \sum_{lm}\langle lm|T_{\mathbf{k}}^{N-1}|lm\rangle$$

$$\text{free:}\quad 4\pi\langle 00|T_{\mathbf{k}}^{N-1}|00\rangle$$

$$\text{antiparallel:}\quad \sum_{lm}(-1)^l\langle lm|T_{\mathbf{k}}^{N-1}|lm\rangle$$

$$\text{fixed relative angle:}\quad (2\pi)^{-1}\sum_{lm}\int_{-\infty}^{\infty} dq\, e^{-iq\alpha} j_l(q) 4\pi i^l \times\langle lm|T_{\mathbf{k}}^{N-1}|lm\rangle.$$

C. End-to-end distribution for zero applied force

For applications to "free" molecules in solution, e.g., equilibrium cyclization (Sec. III), the external tension f is zero. In this case $Z_{\mathbf{k}}$ is isotropic, depending only on $|\mathbf{k}|$. Expanding $e^{-i\mathbf{k}\cdot\mathbf{R}}$ in spherical coordinates, the end-to-end vector distribution becomes

$$\begin{aligned}\rho(\mathbf{R}) &= (2\pi)^{-3}Z^{-1}\int d^3k\, e^{-i\mathbf{k}\cdot\mathbf{R}} Z_k \\ &= 4\pi(2\pi)^{-3}Z^{-1}\sum_{l,m}(-1)^{l+m} i^l \\ &\quad\times\int_0^\infty dk\, k^2 j_l(kR) Z_k \int d^2\hat{k} Y_{l,m}^*(\hat{\mathbf{k}}) Y_{l,m}(\hat{\mathbf{R}}) \\ &= 4\pi(2\pi)^{-3}Z^{-1}\int_0^\infty dk\, k^2 j_0(kR) Z_k,\end{aligned} \tag{10}$$

This distribution is isotropic, and is a function of the radial coordinate only:

$$\rho(R) = \frac{2}{\pi}\frac{1}{Z}\int_0^\infty dk\, k^2 R^2 j_0(kR) Z_k. \tag{11}$$

These distributions will be used in Sec. III to compute equilibrium cyclization probabilities for free chains.

D. End-to-end distribution for polymer under force

In the case that **f** is nonzero, $Z_{\mathbf{k}}$ is axisymmetric around the force $\hat{\mathbf{z}}$) direction. The integral over the azimuthal angle of **k** can be done, leaving integrals over the magnitude and polar angle:

$$\begin{aligned}\rho(\mathbf{R}) &= (2\pi)^{-3}Z^{-1}\int d^3k\, e^{-i\mathbf{k}\cdot\mathbf{R}} Z_{\mathbf{k}} \\ &= (2\pi)^{-3}(4\pi)Z^{-1}\sum_{l,m}(-1)^{l+m} i^l \int_0^\infty dk\, k^2 j_l(kR) \\ &\quad\times\int d^2\hat{k}\, Z_{\mathbf{k}} Y_{l,m}^*(\hat{\mathbf{k}}) Y_{l,m}(\hat{\mathbf{R}}) \\ &= (2\pi)^{-2}Z^{-1}\sum_l (-1)^l i^l (2l+1) P_l(x_R) \\ &\quad\times\int_0^\infty dk\, k^2 j_l(kR)\int dx_k P_l(x_k) Z_{\mathbf{k}},\end{aligned} \tag{12}$$

where $x_k \equiv \cos(\hat{\mathbf{z}}\cdot\hat{\mathbf{k}})$. This distribution is a function of the

magnitude of **R**, and its angle relative to the **z** axis as specified by $x_R \equiv \cos(\hat{\mathbf{z}}\cdot\hat{\mathbf{R}})$.

In single-molecule magnetic tweezer micromanipulation of DNA experiments [36], the fluctuations of the end-to-end vector under constant-force conditions may be directly observed, and thus the complete end-to-end distribution measured. Thus, the full vector distribution (12) can be measured experimentally. However, the one-dimensional distributions of *components* of the end-to-end vector are usually measured, i.e., end-to-end vector components *longitudinal* and *transverse* to the force direction. The moments of these two one-dimensional distributions are related; one of these relations is widely used to measure (calibrate) forces in magnetic tweezer experiments (Appendix) [36].

The distribution of the projection λ of the end-to-end vector along any direction described by unit vector $\hat{\boldsymbol{\lambda}}$ (the distribution of the one-dimensional variable $\lambda \equiv \mathbf{R}\cdot\hat{\boldsymbol{\lambda}}$) is (12):

$$\begin{aligned}\rho(\lambda) &= \int d^3R\, \rho(\mathbf{R})\,\delta(\lambda - \mathbf{R}\cdot\hat{\boldsymbol{\lambda}}) \\ &= (2\pi Z)^{-1}\int_{-\infty}^{\infty} dq\, e^{-iq\lambda}\int d^3k\, Z_{\mathbf{k}}\,\delta^3(\mathbf{k} - q\hat{\boldsymbol{\lambda}}) \\ &= (2\pi Z)^{-1}\int_{-\infty}^{\infty} dq\, e^{-iq\lambda} Z_{\mathbf{k}=q\hat{\boldsymbol{\lambda}}}. \end{aligned} \tag{13}$$

The projected distribution requires only $Z_{\mathbf{k}}$ evaluated along the wave number axis $\mathbf{k} = q\hat{\boldsymbol{\lambda}}$.

E. Localized hinge and kink defects in thermal equilibrium

The above calculations have been applied to the simple semiflexible polymer, each segment of which was considered to have the same flexibility and geometrical properties. However, virtually the same calculations can be applied to more complex and inhomogeneous situations, such as localized defects in double helix structure which impart local changes in dsDNA flexibility. Such local defects might be transiently thermally activated ("annealed"), or alternately located at specific, fixed positions ("quenched"). In either case, including such defects amounts to replacement of the simple transfer matrices $T_{\mathbf{k}}$ discussed above. In this subsection we describe transfer matrices which generate highly flexible "hinge" sites, and local spontaneous bends characterized by preferred bend angles.

By summing out the defect variables, these models also can be considered to describe strongly nonlinear bending elasticity. Despite the complication of nonlinear bending elasticity, at zero force, the orientational correlations in these theories retain a simple exponential decay, allowing one to define a zero-force persistence length.

1. Thermally excited flexible "hinges"

Double-helix DNA is considered to be a stiff polymer, with a persistence length of about 50 nm under physiological conditions [23,43]. This stiffness is a result of double helix structure, which is made robust by base-pairing and base-stacking interactions. By comparison, the covalently bonded sugar-phosphate backbones are completely flexible, characterized by a persistence length of about 0.7 nm [37]. This suggests a mechanism for generation of localized regions of extreme flexibility along the double helix: local disruption of base interactions may give rise to regions where the double helix can be bent easily. Such disruptions might occur by thermal fluctuations which open bases in a localized region of the double helix [31], action of proteins which locally disrupt double helix structure so as to generate local "hinged" regions [33], and chemical modification of double helix DNA which permanently disrupts base pairing in a localized region, for example by removal of one or a few bases along one strand, or by removal of part of one strand altogether.

We focus in this section on "annealed" hinge defects, with the simplification that each site along the molecule has an equal defect-formation probability. Such defects are models for the first two situations listed above, with the condition in the protein case that binding-unbinding equilibrium occurs (this may be less common than generally assumed, see [16]). We suppose that this type of defect changes the bending stiffness of our chain at one of its vertices from a to a'. The energy of the molecule with these hinge excitations possible at each of its vertices is

$$\beta E = \sum_{i=1}^{N-1}\left[\frac{(\delta_{n_i,0}a + \delta_{n_i,1}a')}{2}(\hat{\mathbf{t}}_{i+1} - \hat{\mathbf{t}}_i)^2 + \mu\delta_{n_i,1} - bf\hat{\mathbf{t}}_i\cdot\hat{\mathbf{z}}\right], \tag{14}$$

where the n_i are two-state variables, indicating whether segment i is either in double helix form ($n_i=0$) or contains a hinge defect ($n_i=1$). The defect creation energy μ controls the probability that a defect appears at any particular location, and represents either the free energy cost of generating a local defect [31], or for applications to protein-generated hinges, the binding free energy [33,44]. Note μ here has opposite sign to μ of [33,44].

Including the effect of these hinges, (3) should be replaced with

$$\rho(\mathbf{R}) = \int \frac{d^3k}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{R}} \frac{\displaystyle\int d^2\hat{t}_1\cdots d^2\hat{t}_N \xi e^{ib\mathbf{k}\cdot\sum_{j=1}^{N}\hat{\mathbf{t}}_j} \sum_{n_1,\cdots,n_{N-1}} e^{-\beta E}}{\displaystyle\int d^2\hat{t}_1\cdots d^2\hat{t}_N \sum_{n_1,\cdots,n_{N-1}} e^{-\beta E}}. \tag{15}$$

After summing over $\{n_i\}$, the transfer matrix becomes

$$T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}') = e^{ib\mathbf{k}\cdot\hat{\mathbf{t}}}\left[e^{-a(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2} + e^{-a'(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2-\mu}\right]e^{bf\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}}, \tag{16}$$

or in spherical harmonic representation

$$\begin{aligned}\langle lm|T_{\mathbf{k}}|l'm'\rangle &= \int d^2\hat{t}\, d^2\hat{t}' Y^*_{lm}(\hat{\mathbf{t}}) T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}') Y_{l'm'}(\hat{\mathbf{t}}') \\ &= (4\pi)^{3/2}(-1)^{m'-m}\sum_{l_1,l_2,l_3} i^{l_3}(2l_1+1) \\ &\quad\times(2l_2+1)\sqrt{(2l+1)(2l'+1)(2l_3+1)}\end{aligned}$$

$$\times\begin{pmatrix} l & l_2 & l_1 \\ m & 0 & -m \end{pmatrix}\begin{pmatrix} l & l_2 & l_1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times\begin{pmatrix} l' & l_3 & l_1 \\ -m' & m'-m & m \end{pmatrix}\begin{pmatrix} l' & l_3 & l_1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times[e^{-a} i_{l'}(a) + e^{-\mu-a'} i_{l'}(a')]$$
$$\times i_{l_2}(bf) j_{l_3}(bk) Y_{l_3,m'-m}(\hat{\mathbf{k}}). \quad (17)$$

2. Spontaneous bends ("kinks")

Many proteins which bind to the DNA double helix generate local bends, with $>90°$ deflections generated over just a few base pairs. The net flexibility of the protein-DNA complex can be modified as well; for example, the DNA-bending proteins HMGB1 and NHP6A generate severe bends, but which appear to be more flexible than the bare double helix [16]. Here we discuss the appearance of such "kinks" in thermal equilibrium, which in the context of the proteins means equilibration of binding and unbinding processes. We consider for the moment nonspecific binding where each DNA site is equally likely to be bound by protein. The calculations of this section include analytical expressions for transfer matrices of the models introduced in [33].

The energy including thermally equilibrated kinks is

$$\beta E = \sum_{i=1}^{N-1}\left[\frac{\delta_{n_i,0} a}{2}(\hat{\mathbf{t}}_{i+1}-\hat{\mathbf{t}}_i)^2 + \frac{\delta_{n_i,1} a'}{2}(\hat{\mathbf{t}}_i\cdot\hat{\mathbf{t}}_{i+1}-\gamma)^2 + \mu\delta_{n_i,1} - bf\hat{\mathbf{t}}_i\cdot\hat{\mathbf{z}}\right], \quad (18)$$

where a' is the bending modulus of the kink, and γ is the cosine of the preferred (lowest-energy) kink angle [33].

Similar to the case of fluctuating hinges, the transfer matrix for (18) is

$$T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}') = e^{ib\mathbf{k}\cdot\hat{\mathbf{t}}}[e^{-a(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2} + e^{-(a'/2)(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'-\gamma)^2-\mu}]e^{bf\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}}. \quad (19)$$

The function $e^{-(a'/2)(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'-\gamma)^2}$ can be decomposed:

$$e^{-(a'/2)(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'-\gamma)^2} = \int d\tau\,\delta(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'-\tau)e^{-(a'/2)(\tau-\gamma)^2}$$
$$= \frac{1}{2\pi}\int dq\, e^{iq\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'}\int d\tau\, e^{(-a'/2)(\tau-\gamma)^2-iq\tau}$$
$$= 4\pi\sum_{l,m} i^l Y^*_{l,m}(\hat{\mathbf{t}})Y_{l,m}(\hat{\mathbf{t}}')$$
$$\times\left(\sqrt{\frac{1}{2\pi a'}}\int_{-\infty}^{\infty} dq\, j_l(q)e^{-q^2/2a'-iq\gamma}\right). \quad (20)$$

If we define

$$\eta_l(a',\gamma) = i^l\sqrt{\frac{1}{2\pi a'}}\int_{-\infty}^{\infty} dq\, j_l(q)e^{-q^2/2a'-iq\gamma}, \quad (21)$$

the spherical harmonic representation of $T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}')$ becomes

$$\langle lm|T_{\mathbf{k}}|l'm'\rangle = \int d^2\hat{t}\, d^2\hat{t}' Y^*_{lm}(\hat{\mathbf{t}})T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}')Y_{l'm'}(\hat{\mathbf{t}}')$$
$$= (4\pi)^{3/2}(-1)^{m'-m}\sum_{l_1,l_2,l_3} i^{l_3}(2l_1+1)(2l_2+1)\sqrt{(2l+1)(2l'+1)(2l_3+1)}$$
$$\times\begin{pmatrix} l & l_2 & l_1 \\ m & 0 & -m \end{pmatrix}\begin{pmatrix} l & l_2 & l_1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times\begin{pmatrix} l' & l_3 & l_1 \\ -m' & m'-m & m \end{pmatrix}\begin{pmatrix} l' & l_3 & l_1 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times[e^{-a} i_{l'}(a) + e^{-\mu}\eta_{l'}(a',\gamma)]$$
$$\times i_{l_2}(bf) j_{l_3}(bk) Y_{l_3,m'-m}(\hat{\mathbf{k}}). \quad (22)$$

3. Nonlinear bending elasticity in annealed-defect models

Above, we have considered thermally fluctuating ("annealed") local defects, using excitations which change double helix elasticity, or introduce "kinks" of preferred angle. These models are still homogeneous, since every site is described by the same transfer matrix. Furthermore, since we have not introduced any coupling ("cooperativity") between adjacent defect variables n_i, we proceeded above by summing over them *before* carrying out the matrix multiplications. Our calculations are easily seen to be equivalent to homogeneous semiflexible polymers with nonharmonic bending elasticity.

For example, for the "hinge" defects of Sec. II E 1, the net bending energy of one of the segment-segment joints of (16) is

$$\beta E_b(\hat{\mathbf{t}},\hat{\mathbf{t}}') = -\ln[e^{-(a/2)(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2} + e^{-\mu}e^{-(a'/2)(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2}]$$
$$= \frac{a}{2}(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2 - \ln[1+e^{-\mu}e^{[(a-a')/2](\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2}]. \quad (23)$$

Consider the case $a>a'$ and $\mu>0$, describing rare, flexible hinge defects; then, for sufficiently small bends the linear elastic energy $(a/2)(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2$ dominates. However, for larger bends, the term $(a'/2)(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2$, characterized by a smaller bending rigidity $(a'<a)$ takes over. The crossover from the small-bend to tight-bend behavior occurs when $(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2 = 1-\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}' \approx \mu/(a-a')$. Thus the fluctuating-hinge calculation can also be thought of as describing a homogeneous semiflexible polymer with strongly nonlinear bending elasticity.

The general case is where the segment-segment joint bending energy is described by a function $g(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}')$, which can be expanded in Legendre polynomials:

$$e^{-\beta E_b(\hat{\mathbf{t}},\hat{\mathbf{t}}')} \equiv g(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}') = \sum_l \frac{2l+1}{2} C_l P_l(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}')$$
$$= 2\pi \sum_{lm} C_l Y^*_{lm}(\hat{\mathbf{t}}) Y_{lm}(\hat{\mathbf{t}}'), \tag{24}$$

where the expansion coefficients are $C_l = \int_{-1}^{1} P_l(x) g(x) dx$. The generalization of the transfer matrix (6) is

$$\langle lm|T_{\mathbf{k}}|l'm'\rangle = \int d^2\hat{t}\, d^2\hat{t}' Y^*_{lm}(\hat{\mathbf{t}}) T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}') Y_{l'm'}(\hat{\mathbf{t}}')$$
$$= \frac{1}{2}(4\pi)^{3/2}(-1)^{m'-m} \sum_{l_1,l_2,l_3} i^{l_3}(2l_1+1)(2l_2+1)$$
$$\times \sqrt{(2l+1)(2l'+1)(2l_3+1)} \begin{pmatrix} l & l_2 & l_1 \\ m & 0 & -m \end{pmatrix}$$
$$\times \begin{pmatrix} l & l_2 & l_1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l' & l_3 & l_1 \\ -m' & m'-m & m \end{pmatrix}$$
$$\times \begin{pmatrix} l' & l_3 & l_1 \\ 0 & 0 & 0 \end{pmatrix} C_{l'} i_{l_2}(bf) j_{l_3}(bk) Y_{l_3,m'-m}(\hat{\mathbf{k}}). \tag{25}$$

4. Zero-force orientational correlations and persistence length

The low-force elastic response of the usual semiflexible polymer is characterized by its persistence length, defined to be the tangent-tangent correlation length measured along the chain. In the discrete models with annealed local defects discussed above, the $f=0$ tangent vector correlation function has exactly the same form as in the simple semiflexible polymer, i.e., $\langle \hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i+j}\rangle = e^{-j\Phi}$, where Φ depends on the parameters describing the defects. Thus a persistence length for the annealed-hinge and kink models can be defined via $A_{\text{eff}} = b/\Phi$, where b is the segment length.

To see this, use the transfer matrices (6) for the case $f=0$ and $k=0$ to compute the expectation value of two tangent vectors separated by j segments,

$$\langle \hat{\mathbf{t}}_0 \cdot \hat{\mathbf{t}}_j \rangle = \frac{\int d^2\hat{t}_0 d^2\hat{t}_j (\hat{\mathbf{t}}_0\cdot\hat{\mathbf{t}}_j) \langle \hat{\mathbf{t}}_0|T_0^j|\hat{\mathbf{t}}_j\rangle}{\int d^2\hat{t}_0 d^2\hat{t}_j \langle \hat{\mathbf{t}}_0|T_0^j|\hat{\mathbf{t}}_j\rangle} = \frac{\int d^2\hat{t}_0 (\hat{\mathbf{t}}_0\cdot\hat{\mathbf{z}}) \langle \hat{\mathbf{t}}_0|T_0^j|\hat{\mathbf{z}}\rangle}{\int d^2\hat{t}_0 \langle \hat{\mathbf{t}}_0|T_0^j|\hat{\mathbf{z}}\rangle}. \tag{26}$$

The dot product in the numerator of (26) is proportional to $Y_{10}(\hat{\mathbf{t}}_0)$, selecting the $Y_{10}(\hat{\mathbf{t}}_0)$ component of T_0. The integral in the denominator selects the $Y_{00}(\hat{\mathbf{t}}_0)$ component of T_0. For zero force and wave number, the annealed-defect transfer matrices (17) and (19) are diagonal (proportional to $\delta_{ll'}\delta_{mm'}$), since in the absence of the vectors $\mathbf{f}$ and $\mathbf{k}$, $T_0(\hat{\mathbf{t}},\hat{\mathbf{t}})$ is a function of $\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'$. The matrix product reduces to a product of $j\langle 10|T_0|10\rangle$ contributions in the numerator, and $j\langle 00|T_0|00\rangle$ factors in the denominator. The correlation function (26) is just a constant (the ratio of the 10 and 00 components) raised to the j power.

Φ is therefore just the log of that ratio:

$$\Phi = \ln\left(\frac{e^{-a} i_0(a) + e^{-\mu-a'} i_0(a')}{e^{-a} i_1(a) + e^{-\mu-a'} i_1(a')}\right), \tag{27}$$

for fluctuating flexible hinges. In (27) the persistence length is a simple, positive real number, as for the simple semiflexible polymer.

For the fluctuating kink (spontaneous bend) case the persistence length takes the form

$$\Phi = \ln\left(\frac{e^{-a} i_0(a) + e^{-\mu} \eta_0(a',\gamma)}{e^{-a} i_1(a) + e^{-\mu} \eta_1(a',\gamma)}\right), \tag{28}$$

where $\eta_l(a',\gamma)$ is as defined by (21). The situation is more complicated in this case: the persistence length is a function of both the polymer intrinsic persistence length a, and the kink parameters γ and a'. The interior of the log is no longer constrained to be positive: for sharp and frequent bends, the adjacent-neighbor tangent correlation [the argument of the logarithm in (28)] can become negative (see below). In this case, the tangent correlations are "antiferromagnetic" and the correlations "ring down" with oscillating sign. The persistence length can still be defined, but it will have an imaginary part πi.

The general case for bending elasticity described by a function $g(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}')$ as in (24) is

$$\Phi = \ln\left(\frac{\int_{-1}^{1} P_0(x) g(x) dx}{\int_{-1}^{1} P_1(x) g(x) dx}\right). \tag{29}$$

F. Permanent bends at fixed locations

Above, we have considered spontaneous bends generated in thermal equilibrium, for example by binding or unbinding of DNA-bending proteins. However, it is possible that an irreversibly-bound protein [16] (or a chemical defect in the double helix) might generate a bend at some specific sequence position. This situation can be handled by making the transfer matrix segment dependent [33]. Here, this amounts to replacing $T_{\mathbf{k}}^{N-1}$ (and $T_{\mathbf{0}}^{N-1}$) in (4) by the multiplication $\Pi_{i=1}^{N-1} T_{i,\mathbf{k}}$, where $T_{i,\mathbf{k}}$ is the site dependent matrix.

We illustrate this using the example of a permanent kink at a site q in a chain which is an otherwise homogeneous semiflexible polymer. The energy function is

$$\beta E = \sum_{i=1}^{N-1} \left[\frac{(1-\delta_{i,q})a}{2} (\hat{\mathbf{t}}_{i+1} - \hat{\mathbf{t}}_i)^2 + \frac{\delta_{i,q} a'}{2} (\hat{t}_i \cdot \hat{t}_{i+1} - \gamma)^2 - bf\hat{\mathbf{t}}_i \cdot \hat{\mathbf{z}} \right] \tag{30}$$

and therefore the site-dependent transfer matrix is

$$T_{i,\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}') = e^{ib\mathbf{k}\cdot\hat{\mathbf{t}}}[(1-\delta_{i,q})e^{-a(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2/2} + \delta_{i,q}e^{-(a'/2)(\hat{\mathbf{t}}\cdot\hat{\mathbf{t}}'-\gamma)^2}]e^{bf\hat{\mathbf{z}}\cdot\hat{\mathbf{t}}}. \quad (31)$$

At the nonkink sites, the transfer matrix takes the form of Eq. (6); in general the transfer matrix is:

$$\begin{aligned}\langle lm|T_{i,\mathbf{k}}|l'm'\rangle &= \int d^2\hat{t}\, d^2\hat{t}' Y^*_{lm}(\hat{\mathbf{t}})T_{\mathbf{k}}(\hat{\mathbf{t}},\hat{\mathbf{t}}')Y_{l'm'}(\hat{\mathbf{t}}') \\ &= (4\pi)^{3/2}(-1)^{m'-m}\sum_{l_1,l_2,l_3} i^{l_3}(2l_1+1)(2l_2+1) \\ &\times\sqrt{(2l+1)(2l'+1)(2l_3+1)}\begin{pmatrix} l & l_2 & l_1 \\ m & 0 & -m\end{pmatrix} \\ &\times\begin{pmatrix} l & l_2 & l_1 \\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} l' & l_3 & l_1 \\ -m' & m'-m & m\end{pmatrix} \\ &\times\begin{pmatrix} l' & l_3 & l_1 \\ 0 & 0 & 0\end{pmatrix}[(1-\delta_{i,q})e^{-a}i_{l'}a \\ &+ \delta_{i,q}\eta_{l'}(a',\gamma)]i_{l_2}(bf)j_{l_3}(bk)Y_{l_3,m'-m}(\hat{\mathbf{k}}),\end{aligned} \quad (32)$$

which is just Eq. (22) with the modification that no unkinked state occurs. Of course, a, a' and γ may also be made site-dependent.

The same approach may be used to compute properties of a simple semiflexible polymer with quenched inhomgeneity in its bending stiffness, by using a sequence of transfer matrices of the form (6), with different bending stiffnesses a_i. A recent paper by Ranjith *et al.* has carried out this type of calculation, in order to study the effect of sequence inhomogeneity of bending rigidity on end-to-end statistics [46]. Similarly, different spontaneous bends could be put at each segment, to study the effect of sequence-directed bends [24,47]; the related case of a single "kink" at fixed location will be discussed below.

III. FLEXIBILITY AND CYCLIZATION OF DOUBLE-STRANDED DNAS

We now use the calculations of the previous sections to study loop formation along molecules where no force is applied. One application is to DNA cyclization experiments, and in particular experiments on short ($\approx$100 bp) molecules as studied by Cloutier and Widom [27]. In our previous paper [31] we showed how fluctuating "flexible hinge" defects may explain the surprisingly high cyclization probability observed by Cloutier and Widom. In this section we present more details of those calculations.

In this paper, we use a dsDNA segment length of $b=1$ nm, or 3 bp. This is a both a convenient and realistic cutoff on the simple bending elasticity models that we are considering. This length scale also defines the approximate size of disordered double helix regions that are likely to impart strong bending flexibility. For segments shorter than 3 bp, generation of a highly flexible joint would require a series of disordered segments, requiring a model including cooperativity effects. For this segment length, the bending elastic constant $a=50$ provides the usual 50 nm large-scale bending persistence length (recall $A=ab$); we use this value in all calculations in this paper.

We do not include any change in length of disordered regions. Although single-stranded DNA(ssDNA) does have a longer contour length than dsDNA of the same number of bases [37], this simplification does not affect the results of our calculation since we will not have a substantial fraction of our molecules ever converted to disordered form. Segment length changes in this type of model can be included if needed [33,38].

A. Persistence length in the presence of fluctuating local defects

1. Effect of flexible "hinge" defects

We begin by computing the persistence length for our model, in the presence of thermally excited bending defects. We treat the hinge case first, with emphasis on its application to the description of spontaneously (thermally) excited disordered regions of the DNA double helix, motivated by the finding of Widom and Cloutier [27] that short, linear dsDNAs cyclize far more readily than expected on the basis of the usual semiflexible polymer with linear bending elasticity [29,30].

Following our previous work [31], we take the hinge to represent the bending energy of a 3 base region of disordered dsDNA. We use a bending constant for the hinge region of $a'=1$, corresponding to assuming a persistence length for the hinge approximately 3 bases long. This choice of $a' \ll a$ is motivated by mechanical measurements which indicate ssDNA to have a persistence length of about one base [37]; our choice $a'=1$ corresponds to about three times the rigidity inferred from those experiments. Our use of a larger persistence length takes into account the fact that our "hinge" region contains two side-by-site ssDNAs. We arrived at the 3 bp defect size by considering that shorter defects will be unable to form sharp bends.

Once a' is set, the remaining parameter in the fluctuating hinge model is μ, the free energy (in k_BT units) associated with creation of the defect region. Figure 2(a) shows how the persistence length (27) depends on this parameter. When this energy cost is very large, the persistence length is equal to $ab=50$ nm. In the opposite limit where the free energy μ is small or negative, the hinge defects dominate, and the persistence length is pushed down to a'. There is a rather broad transition between the stiff and highly flexible regimes stretching from about $\mu=+5$ to $+10$. It is straightforward to show that for an unconstrained linear chain, the hinge density $\langle n\rangle$, i.e., the average number of hinges per segment, is

$$\langle n\rangle = \frac{1}{1+\frac{i_0(a)}{i_0(a')}e^{\mu}e^{(a'-a)}}. \quad (33)$$

Figure 2(b) shows the hinge density as a function of μ with $a=50$ and $a'=1$. Below we will briefly review how the cyclization data of Widom and Cloutier [27] indicate $\mu \approx 11k_BT$. This defect energy is well into the rare-defect re-

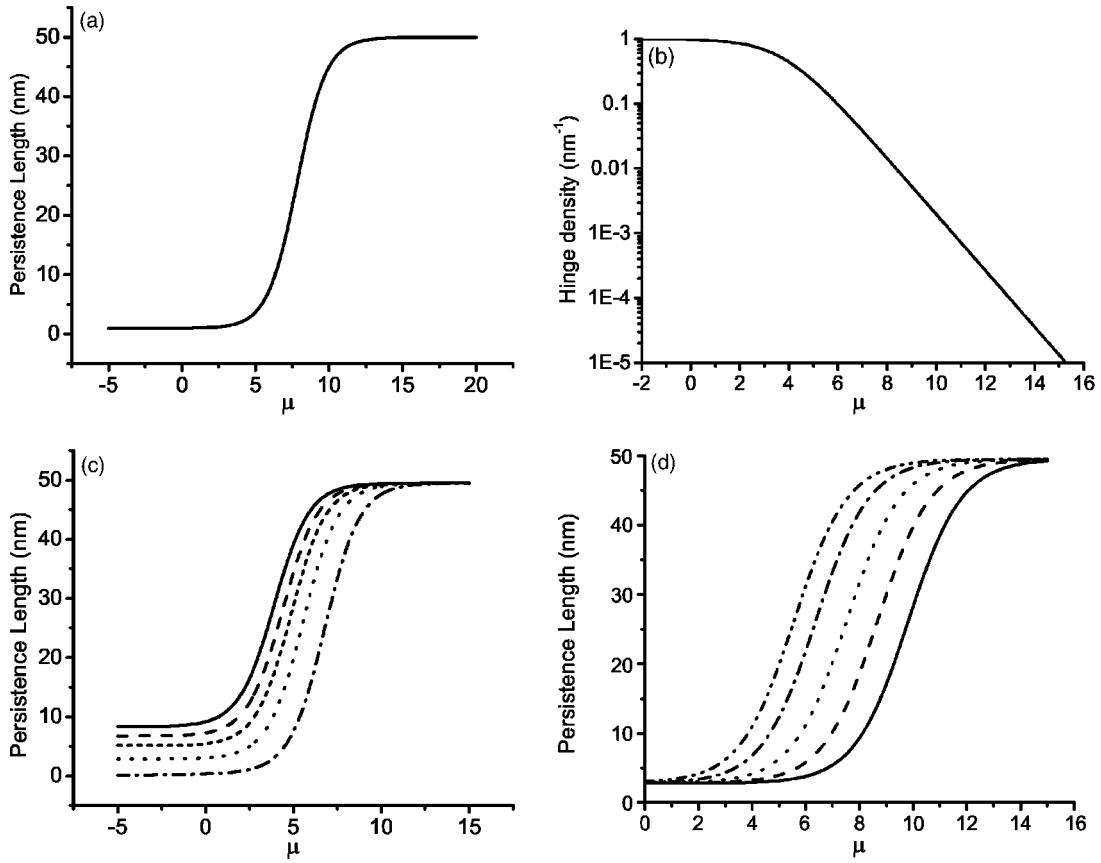

FIG. 2. Persistence lengths for fluctuating-defect models of dsDNA. In all cases, the segment length $b=1$ nm and the unperturbed double helix bending constant $a=50$, corresponding to 50 nm persistence length. (a) Flexible-hinge model persistence length as a function of defect free energy μ. The excited hinges have bending stiffness $a'=1$. For $\mu>10$, the defects are so rare that they do not significantly perturb the net flexibility, but for $\mu<10$, the hinges reduce the net persistence length. (b) Hinge density as a function of μ. For $\mu>10$, it is less than 1 hinge per 1000 nm of the molecule. (c) Spontaneous-bend "kink" model persistence length as a function of defect free energy μ, for various preferred bend angles. The kink bend constant is fixed at $a'=50$. In each case, large μ causes the persistence length to revert to the unperturbed 50 nm value; small μ greatly reduces the net persistence length. Results are shown for angles $\gamma=0$ (solid), $\pi/9$ (dash), $\pi/6$ (short dash), $\pi/4$ (dot), and $\pi/2$ (dash-dot). As the bending angle is increased, the net persistence length is gradually reduced. For a $\pi/2$ angle, the kink-dense model (low μ) has essentially zero persistence length as defined by tangent vector correlations. (d) Spontaneous-bend model persistence length as a function of defect free energy μ, at fixed preferred bend angle $\gamma=\pi/4$, for various kink flexibilities. Results are shown for $a'=50$ (dash-dot-dot), 10 (dash-dot), 1 (dot), 0.1 (dash), and 0.01 (solid). As the kinks are made more flexible, the effect of the defects on persistence length becomes more pronounced.

gime: there is about 1 hinge per 1000 segments (1000 nm or 3000 bp) along an unconstrained molecule.

2. Effect of spontaneous bend ('kink') defects

We now consider thermally excited spontaneous bend defects. While it is possible that local reorganizations of the double helix might give rise to thermally activated spontaneous bends, this type of local defect is most likely to find application to experiments on DNA-bending proteins or drugs which generate local bends [13,15,16], and which bind *non-specifically*, i.e., equally well to any site along a dsDNA. In this case, the defect free energy μ describes the binding free energy of the ligand or protein, as discussed in Ref. [33], via $\mu=-\ln c+\text{const}$, where c is the ligand or protein solution concentration.

Figure 2(c) shows how the effective persistence length of a dsDNA ($a=50$) with fluctuating kinks depends on the kink creation energy μ, and the kink angle $\gamma=0$ (solid), $\pi/9$ (dash), $\pi/6$ (short dash), $\pi/4$ (dot), and $\pi/2$ (dash-dot), respectively. The kink rigidity has been taken to be $a'=50$, comparable to the double helix itself in this calculation. In all cases, a kink energy $\mu\gg 1$ yields the unperturbed persistence length of 50 nm, with a gradual reduction in persistence length for smaller μ where the kink defects become more likely. In this "rigid kink" case, small bend angles can result in an appreciable reduction in the persistence length (solid), while larger bend angles push it down to near zero (dot). The case of $\gamma=\pi/2$ with frequent kinks ($\mu\leqslant 0$) has a persistence length of essentially zero, indicating that the tangent vector dot product has near-zero correlation from segment to segment, as a result of the nearly perpendicular kink. Larger bend angles generate a negative neighbor-tangent correlation, and therefore an imaginary component of the persistence length as defined by (28).

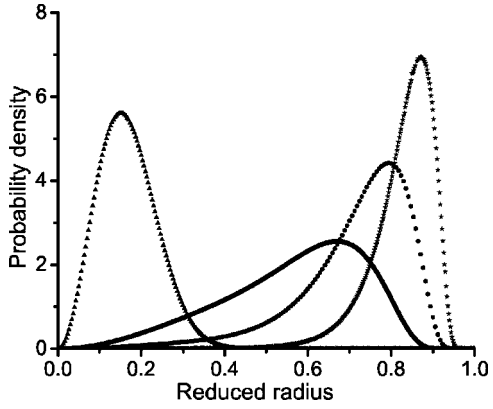

FIG. 3. Distribution of end-to-end radius for the simple semiflexible polymer model of dsDNA ($b=1$ nm, $a=50$, no hinge or kink defects), as a function of radius in units of molecule contour length (R/l_0)), for the parallel-end boundary condition. Results are shown for molecular lengths 100 (rightmost peak), 170, 280, and 3000 (leftmost peak) nm, corresponding to sequence lengths of 300, 510, 840, and 9000 bp.

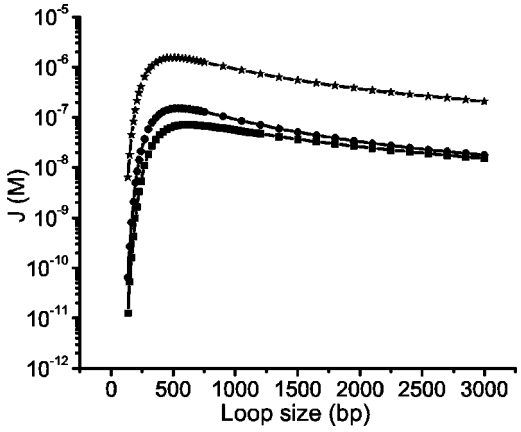

FIG. 4. J factors versus molecule length, for the simple semiflexible polymer with persistence length 50 nm. Results are shown for three different cyclization boundary conditions: parallel (lowest curve), antiparallel (middle curve), and free (highest curve).

Recent experiments [15,16] indicate that many DNA-bending proteins generate rather flexible kinks with persistence lengths in the few-nm range. Figure 2(d) shows how the effective persistence length of a dsDNA ($a=50$) with fluctuating kinks depends on the kink creation energy μ, and the kink rigidity $a'=0.01$ (solid), 0.1 (dash), 1 (dot), 10 (dash dot), and 50 (dash dot dot), respectively. The kink angle $\gamma=\pi/4$ in this calculation.

Note that there is an appreciable shift in the threshold μ for the persistence length to change from its high-μ value to its low-μ value; this is due to the varying joint bending entropy. For very small a', kinks are made more probable by their large conformational entropy; for very large a', the defect probability is reduced. This effect also occurs for the hinge defects.

B. Zero force end-to-end distributions and cyclization

We can compute the complete end-to-end distribution $\rho(\mathbf{R})$, the probability of the two ends of the polymer being a certain distance from one another and having a specific orientation (boundary condition). In the case that no external force is present, this distribution takes the form (11) which is a function only of the magnitude of the end-to-end distance R.

As a simple example, Fig. 3 shows the end-to-end radius distribution for the parallel boundary condition, for a semiflexible polymer (no fluctuating defects) of persistence length 50 nm. The figure plots ρ as a function of radius in units of contour length, i.e., as a function of R/L. Results are shown for molecular lengths 100, 170, 280, and 3000 nm dsDNAs (300, 510, 840, and 9000 bp). For very short molecules, the peak of the distribution is near $R/L=1$ since significant bends are rare. For larger molecules, the peak of the distribution moves to lower values of R/L; the typical end-to-end distance for a very long molecule scales as $R \propto (2AL)^{1/2}$ [39], so $R/L \propto (2A/L)^{1/2}$. The shape of the distribution changes from being very asymmetric for nearly straight molecules (right distribution) to being more symmetric (left distribution), as expected for long chains [39].

Experiments can be carried out so that the rate of cyclization of a dsDNA is proportional to the equilibrium probability of the end-to-end vector being zero [22,29]. For a free polymer ($f=0$) the quantity measured experimentally in such experiments is the "J factor" [24,30], which is essentially the end-to-end vector distribution (10) for $\mathbf{R}=\mathbf{0}$:

$$J = \frac{4\pi}{N_A}\rho(\mathbf{0}) = \frac{2}{\pi}\frac{1}{N_A Z}\int_0^\infty dk\, k^2 Z_k, \tag{34}$$

where N_A is Avogadro's number. This quantity, expressed in units of mol/litre (M) is a measure of the equilibrium concentration of one end of the polymer at the other. Our calculations are for closure of dsDNAs without torsional constraint [29,30]. We note that the 4π factor was omitted in Eq. (4) and Eq.(5) of Ref. [31] (although not in the numerical calculations).

Figure 4 shows J factors calculated for the simple semiflexible polymer with $b=1$ and $a=50$, with no thermally excited defects (persistence length 50 nm). The three curves show J versus molecule length, for three different cyclization boundary conditions: parallel (lowest curve), antiparallel (middle curve), and free (highest curve). The free boundary condition has the highest cyclization rate of the three, since it has the least severe conformational constraint. The antiparallel (hairpin-shaped) closure requires less bending than the parallel (circular) case, and therefore has the higher probability. These three results are in excellent agreement with previous cyclization calculations for the semiflexible polymer [24,28–30]. In each of the boundary condition cases of Fig. 4 the J factor has a peak near 500 bp, which is a compromise between the entropic suppression of J for long molecules, and the energetic (bending) suppression for short molecules [29]. For ligation by DNA ligase, the usual way that cyclization of dsDNA is done, the appropriate boundary condition is the parallel one.

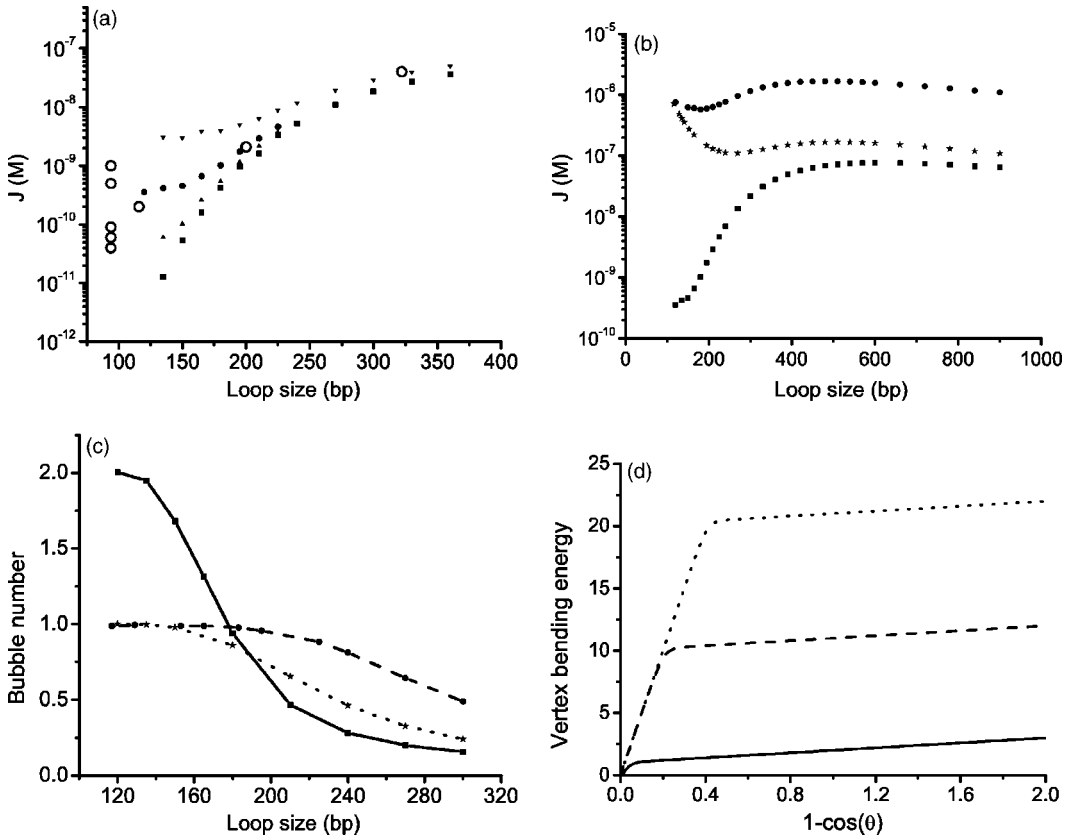

FIG. 5. Effect of thermally activated flexible hinges on cyclization of short dsDNAs. Hinge flexibility is $a'=1$ (3 base persistence length); unperturbed double helix flexibility is $a=50$ (50 nm persistence length). (a) J factor versus molecule length for parallel-end boundary condition. Results (from bottom to top) for $\mu=\infty$ (no hinge), 12, 11, and 10. Open circles show data of Widom and Cloutier [27]. (b) Influence of closure boundary condition on cyclization in the hinge model ($a=50$, $a'=1$, and $\mu=11$). J-factor versus molecule length is plotted for parallel (squares), antiparallel (stars), and free (filled circles) closure boundary conditions. (c) Number of hinge defects on molecule in cyclized state, versus molecule length, for $\mu=11$. For molecules shorter than 300 bp, cyclization becomes affected by molecules with hinge excitations. Results are shown for the parallel boundary condition (solid+squares), and also for the antiparallel (dashed +filled circles) and free (dotted+stars) boundary conditions. (d) Nonlinear bending elasticity versus angle for hinge model, for defect energies $\mu=1$ (solid), 10 (dashed), and 20 (dotted). The elasticity shows a "crossover" behavior from the small-angle, large-scale persistence length ($a=50$) of 50 nm, to a much smaller rigidity ($a'=1$), at a critical bend angle controlled by μ.

1. Role of excited hinge defects in cyclization of short dsDNAs

For dsDNA, cyclization experiments have produced results in good agreement with Fig. 4 for molecule lengths larger than 200 bp, but shorter molecules were not precisely studied before the recent work by Cloutier and Widom [27]. Remarkably, Cloutier and Widom found that for molecules of length near 100 bp, their measured J factors were in excess of 10^4 times larger than those expected from classical semiflexible polymer theory (i.e., Fig. 4). Their experimental data indicate that double-helix looping at the 100 bp and 500 bp scales cannot be described by the simple semiflexible polymer model with persistence length 50 nm.

In our previous paper [31] we suggested a simple resolution to the paradox presented by Widom and Cloutier's data, based on the thermally-activated "flexible hinge" model described above. In short, the idea is that at room temperature, thermally excited local defects in double helix structure occur, which will be highly flexible relative to the unperturbed, base-paired and stacked double helix. These defects are rare enough to not strongly perturb the large-scale persistence-length behavior of the double helix, but frequent enough to facilitate bending of short dsDNAs.

We considered "melting," or strand-separation, as a model of such defects since the free energy of strand separation is reasonably well known [40]. Using the "standard model" of DNA melting [40,41], we estimated that the free energy cost of formation of a 3 bp region of strand-separated dsDNA should be between 7 and 16 k_BT, depending on sequence [31]. Use of such models in straightforward calculations of opening fluctuations including cooperativity effects inside a 48.5 kb λ-DNA indicates that at zero force and room temperature, about 0.3% of the bases along a λ-DNA are unpaired at any given moment [42]. Depending on sequence, the probability of a 3 bp opening event might be as large as 0.1% per base, corresponding to a spacing of about 1 kb between thermally activated "hinges."

Figure 5(a) shows J-factors calculated using our "flexible hinge" theory for a range of defect energies μ, along with the experimental data of Widom and Cloutier [27]. For this calculation, $b=1$ nm, $a=50$, and $a'=1$. Results are shown for a range of μ values; we find that J factors which are in accord

with the short-molecule data occur for $\mu=11k_BT$. This is costly enough that for longer molecules >150 bp, the probability of hinge excitations is low enough that the polymer reverts to being well-described as a semiflexible chain of persistence length 50 nm (Fig. 2). For molecules $\geqslant 200$ bp, the calculated J factors closely approach those shown for the simple semiflexible polymer (Fig. 4).

In the short-molecule regime, boundary conditions have a large effect on the J factor. Figure 5(b) compares the parallel-end J factors calculated for the parameters $a=50$, $a'=1$, and $\mu=11$ (the choice which generates short-segment J factors consistent with the Widom-Cloutier experiment) with results for antiparallel and free boundary conditions. The antiparallel and free boundary conditions allow "hairpin" configurations to dominate the J factor. Our model predicts that for the antiparallel and free closures, the classical peak at 500 bp will be washed out by the strong J-factor enhancement; the peak J-factor is shifted to below 100 bp.

Figure 5(c) shows the average number of hinge defects occuring in the cyclized configurations of the chain, minus the average occuring in all configurations, as a function of molecular length. This is $\Delta n_{\mathrm{cyc}}=-\partial \ln J/\partial\mu+\langle n\rangle N$, where $\langle n\rangle$ is the hinge density for the unconstrained chain defined in Eq. (33) and N is the number of segments, since J is the ratio of the partition functions of cyclized and linear molecules. Under parallel boundary condition, for long chains >200 bp, Δn_{cyc} approaches zero, since cyclization occurs easily without any severe bending of the chain. However, for chains <150 bp, Δn_{cyc} rises from zero, and approaches 2 by about 100 bp. The cyclized configurations dominating for short chains with parallel boundary condition have two excited hinges. This cyclized state becomes favorable for short chains since large-scale bending may be entirely eliminated by forming a "squashed" configuration with two hinges.

The hinge number for the parallel boundary condition case may be rationalized by considering the relative energy costs of forming smoothly bent, one-hinge, and two-hinge cyclized states. A smoothly bent cyclized chain with parallel ends has bending energy of at least that of a circle, which is $2\pi^2A/L$ where L is the molecule length, and A is the 50 nm persistence length. Creation of one hinge allows part of the bending energy to be eliminated; the optimal energy becomes $\mu+14A/L$ (the numerical factor of 14 can be found in Ref. [29]); this drops below that of the circle state for $L/A<(2\pi^2-14)/\mu$. Plugging in $A=150$ bp and $\mu=11$ indicates that the one-hinge state becomes favorable for $L/A<80$ bp. However, the two-hinge state, with excitation energy 2μ, has *zero* bending energy, and therefore becomes lower in energy than the zero-hinge state when $L/A<\pi^2/\mu$. Taking $A=150$ bp and $\mu=11$ shows that the two-hinge state becomes preferable to the circle when $L<135$ bp. This estimate indicates that the double helix is sufficiently stiff that short chains should cyclize via creation of two hinge regions.

The antiparallel and free boundary conditions have much larger J factors than the parallel boundary condition case [see Fig. 5(b)]. Figure 5(c) shows that this is because the dominant cyclization states for short molecules require only one hinge excitation. With the antiparallel (or free) boundary condition, closure can occur with essentially zero bending energy by making a "hairpin" configuration. This is a critical prediction of the thermally-excited hinge-defect explanation of the Widom-Cloutier experiments. Ranjith *et al.* [46] have recently emphasized this for the free boundary conditions case.

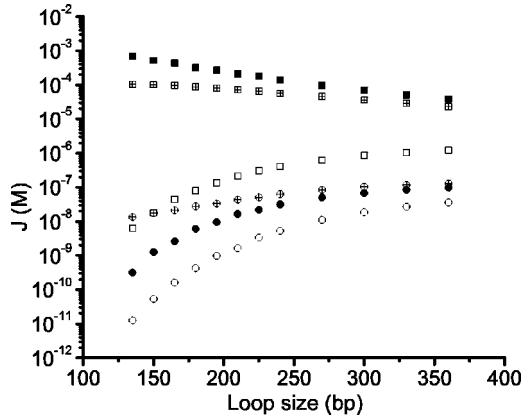

FIG. 6. Enhancement of J factor of dsDNA ($a=50$) by a single preferred-angle kink located at the midpoint of the molecule. The kink angle is $\gamma=\pi/2$ and is stiff ($a'=50$). J factors are shown for parallel (filled circles) and free (filled squares) boundary conditions, in the absence of any other defects (i.e., no thermally excited hinges). J factors are further affected by combining the fluctuating hinges and the permanent kink together (cross-in-circles for parallel boundary condition, and cross-in-squares for free boundary condition). J factors of defect-free dsDNA are also shown for parallel (open circles) and free (open squares) boundary conditions.

Finally we note that the fluctuating-hinge model can be thought of as a semiflexible polymer, with a nonlinear bending elasticity (Sec. II E 3). Figure 5(d) shows the effective bending elasticity of the hinge model given by Eq. (23), as a function of $(1/2)(\hat{\mathbf{t}}-\hat{\mathbf{t}}')^2=1-\cos\theta$ where θ is the angle between two adjacent tangent vectors. Results are shown for $\mu=1$, 10, and 20. In the $\mu=10$ case, for very small bends (small $1-\cos\theta$), the bending energy increases at a rate determined by the large-scale persistence length of 50 nm ($a=50$). However, at $1-\cos\theta\approx 0.2$, there is a crossover in the bending elasticity to a lower slope corresponding to the 50-fold higher flexibility of the excited hinge ($a'=1$). The experimentally observed cyclization can therefore be thought of as a consequence of the nonlinear elasticity of the double helix which makes high-curvature bends occur via "focusing" of bending into a localized defect. The point at which this effect occurs is determined by the parameter μ (Fig. 5).

2. Influence of a single preferred-angle kink on J-factor

Figure 6 shows how the dsDNA J factor ($a=50$) is enhanced by a single preferred-angle kink located at the midpoint of a linear dsDNA. The kink is considered to have an angle of $\gamma=\pi/2$ and to be stiff ($a'=50$). J factors are shown for parallel (filled circles) and free (filled squares) boundary conditions, in the absence of any other defects (i.e., no thermally excited hinges).

The enhancement of J factor is similar to that generated by the fluctuating-hinge model (Fig. 5). Treatment of more complex situations involving combinations of thermally ex-

cited and fixed-position defects is straightforward: results for combining the kink with fluctuating hinges ($\mu=11$) are shown here (cross-in-circles for parallel boundary condition, and cross-in-squares for free boundary condition). It is interesting to note that J factor for a kink plus fluctuating hinges is less than J factor for a kink only.

IV. END-TO-END DISTRIBUTIONS FOR DOUBLE-STRANDED DNA SEGMENTS UNDER TENSION

We now examine a few consequences of the results of the previous section, for dsDNA molecules under tension. We focus first on the end-to-end vector statistics, with some emphasis on how the distributions parallel to and perpendicular to the force direction are related. We then show how loop formation along a stretched dsDNA is suppressed by applied force.

A. End-to-end vector distribution

We consider a force applied to a dsDNA in the z direction. Figure 7(a) plots extension $\langle z\rangle$ for the conventional semiflexible polymer model of the double helix (no fluctuating hinges, 50 nm persistence length, open circles), and also the extension for the "fluctuating hinge" model used to fit the cyclization data of Cloutier and Widom [27] ($b=1$ nm, $a=50$, $a'=1$, $\mu=11$, closed circles) in the preceeding section. The force-extension curves are essentially indistinguishable, i.e., the fluctuating hinges do not generate enough flexibility to cause an experimentally distinguishable signal in the force-extension behavior, in agreement with a similar calculation by Wiggins and co-workers [32].

Figure 7(b) and 7(c) show these two cases, with extension plotted against $1/\sqrt{f}$, as is often done in order to extract the persistence length of dsDNA from experimental force-extension data using the semiflexible polymer model [43]. The simple semiflexible polymer and the polymer including fluctuating hinges generate essentially the same behaviors. Linear fits of extension to $1/\sqrt{f}$ over the 0.3 to 1 pN range usually used to extract persistence lengths result in 48.9 nm for the semiflexible polymer without any fluctuating hinges, and 48.5 nm for the polymer including the hinges. Again, this shows that effects of the fluctuating hinges that we propose to be responsible for the cyclization enhancement of Cloutier and Widom cannot be easily observed in force-extension data, consistent with Ref. [32].

The fluctuation of the end-to-end vector around its average extension is routinely measured in single-molecule manipulation of DNA experiments. We show distributions calculated for the semiflexible polymer model, along the force direction [longitudinal, Fig. 8(a)], and transverse to the force direction [transverse, Fig. 8(a)], for the semiflexible polymer of total length 10 microns (30 kb) and persistence length 50 nm ($a=50$, $b=1$ nm, no hinge fluctuations), at a force of 0.1 pN. At this force the polymer is extended to about 53% of its maximum length, as reflected by the peak near 5.3 microns. The longitudinal distribution is asymmetric around its peak, reflecting the symmetry-breaking effect of the applied force.

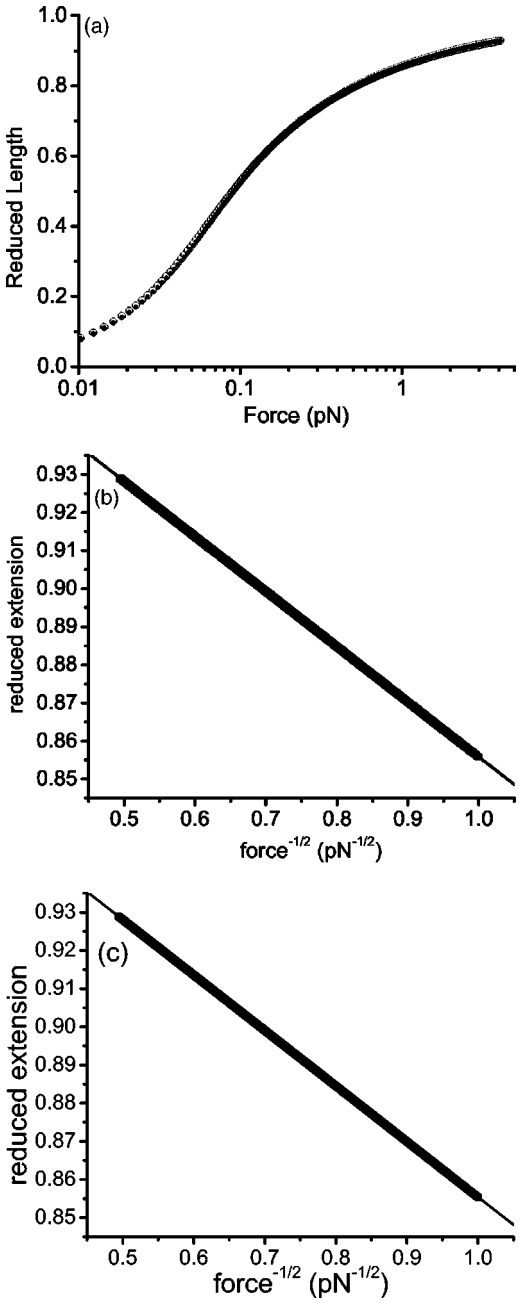

FIG. 7. Effect of thermally activated flexible hinges for $\mu=11$ on force-extension curve is negligible. (a) Force-extension curves for defect-free dsDNA (open circles) and for dsDNA with hinge fluctuation (filled circles); the two results coincide. (b) and (c) Persistence lengths for unperturbed dsDNA and dsDNA with hinge fluctuation are extracted from the force-extension curve at high force: 48.9 nm for unperturbed dsDNA, and 48.5 nm for dsDNA with hinge fluctuation.

By contrast, the transverse distribution is symmetric, and nearly a pure Gaussian.

The moments of the two distributions shown in Figs. 8(a) and 8(b) are related (see Appendix). Figure 8(c) shows a check of one of these relations, between the first moment of the longitudinal distribution (the average extension) and the second moment of the transverse distribution. The ratio of these two quantities is expected to be equal to the force di-

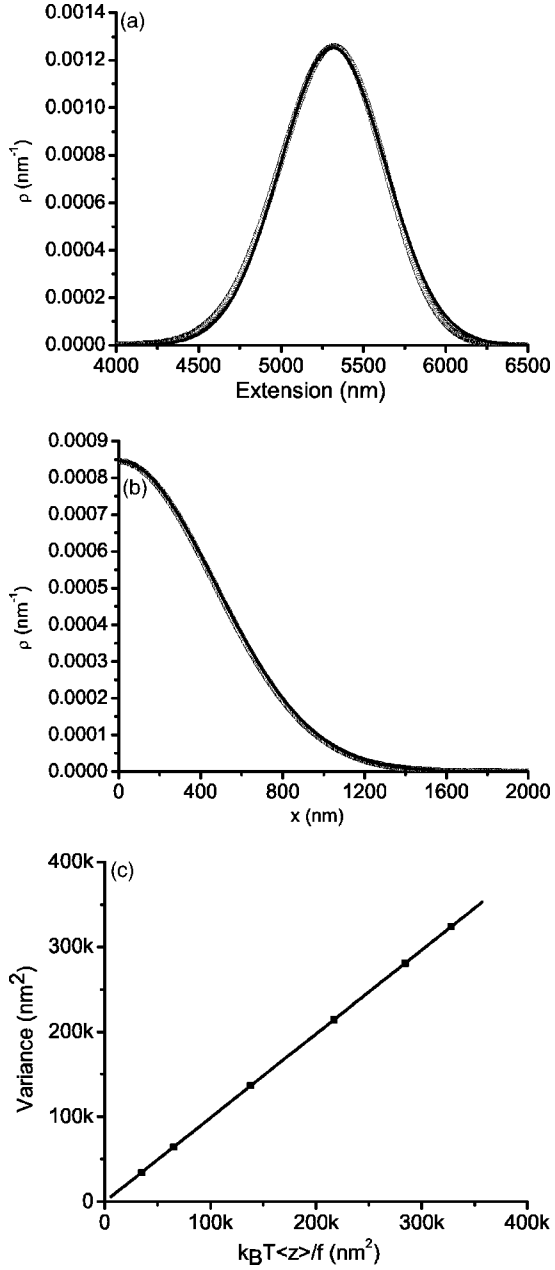

FIG. 8. End-to-end distributions of a 10 μm (30 kb) dsDNA, in longitudinal (force) and transverse directions. (a) Open circles are the longitudinal distribution, $\rho(z)$, at $f=0.1$ pN. Solid line is a Gaussian distribution with its peak matched to that of $\rho(z)$. (b) Open circles are transverse distribution, $\rho(x)$, at $f=0.1$ pN. It can be fit closely by a Gaussian distribution (solid line). (c) The variance of the transverse fluctuation is equal to the end-to-end distance divided by force in k_BT units ($(k_BT\langle z\rangle/f)$, for all forces.

vided by k_BT, independent of any details of the polymer model; this relation is used to *measure* forces in single-molecule manipulation of DNA experiments using magnetic tweezer. Figure 8(c) shows the moment ratio derived from distributions at a series of forces [filled squares, Fig. 8(c)]; this matches the expected ratio [straight line, Fig. 8(c)].

Figure 9 shows the form of the general distribution $\rho(x,z)$, where z is the coordinate of the force direction, and x is the coordinate of a direction transverse to the force direction, for a 334 nm (1000 bp) dsDNA at a force of 0.04 pN. The results shown are for the semiflexible polymer with 50 nm persistence length without fluctuating hinge defects ($a=50$, $b=1$ nm). This general distribution has symmetry under reflection of $x\rightarrow -x$, but no such symmetry in the z direction. For this short molecule and low force, the widths of the distribution are comparable to the average extension.

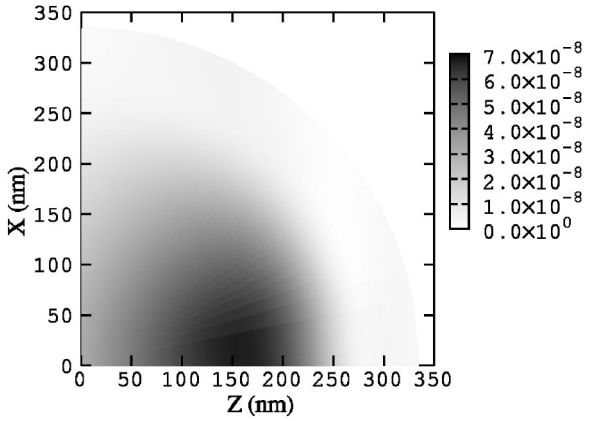

FIG. 9. The general end-to-end distance distribution of a 334 nm (1000 bp) dsDNA at a force $f=0.04$ pN. Here z is the coordinate of the force direction, and x is the coordinates of a direction transverse to the force direction. The probability density is in units of nm^{-3}.

B. Force-dependent loop formation probability

The end-to-end distributions of the previous section can be evaluated at zero distance, to predict loop-formation probabilities analogous to those discussed in Sec. III, but at nonzero force. One expects a force f which appreciably extends a polymer to strongly quench formation of loops along its length, since a fluctuation which does mechanical work $\approx f\ell$ must occur to spontaneously cyclize a segment of length ℓ under tension which stretches a chain [44]. This implies that loop formation probability should be suppressed as $J(f) \propto e^{-\beta f\ell}$. For forces less than $\approx k_BT/A \approx 0.1$ pN, a short dsDNA segment will be only weakly extended, allowing

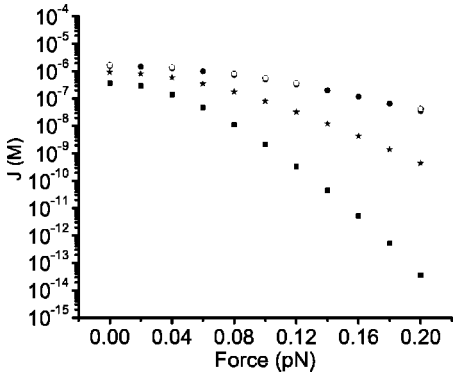

FIG. 10. Effect of force on J factors of dsDNA with different lengths. Results are shown for a 2 kb unperturbed dsDNA (filled squares), an 1 kb unperturbed dsDNA (stars), 0.5 kb without fluctuating hinge defects (filled circles), and 0.5 kb with fluctuating hinge defects ($\mu=11$, $a'=1$, open circles).

loop formation to occur against the applied tension via thermal fluctuation.

Our approach can be used to directly calculate tension effects of loop formation probability, which is dependent on both tension and loop length. Figure 10 shows the J factor (defined in the same way as in Sec. III, only for a stretched chain) as a function of force, for dsDNA segments of 500 bp (circles), 1000 bp (stars) and 2000 bp (squares) long. A "free" boundary condition was used in these calculations. For forces in excess of 0.1 pN, we observe the expected linear dependence of $\ln J$ on force. At 0.1 pN there is still appreciable suppression of J by force for the 2 kb loop (J is reduced by more than tenfold by a force of 0.1 pN), but relatively little for the 500 bp loop. Our results are in good agreement with recent analytical calculations by Sankararaman *et al.* [45].

Fluctuating-hinge excitations along the double helix might be expected to facilitate looping under tension. However, for the case ($\mu=11$, $a'=1$) discussed above, the effect is negligible for the 1 kb and 2 kb loops of Fig. 10, so we do not include them in the figure. The open circles in Fig. 10 show the fluctuating-hinge result ($\mu=11$, $a'=1$) relative to the usual semiflexible polymer model (filled circles) for the 500 bp case. The effect of the fluctuating hinges for the 500 bp case amounts to a small increase in J, by about 10%, similar to the zero-force enhancement. For shorter loops, the fluctuating hinge effect will lead to a stronger enhancement of $J(f)$.

V. CONCLUSION

We have generalized our previous work on use of transfer matrix calculations for semiflexible polymers [33], to compute statistical distributions of the vector between two points along a dsDNA. We have computed the transfer matrix elements analytically for a wide variety of polymer models which are variants on the usual semiflexible polymer model. We have paid particular attention to the semiflexible polymer including localized "flexible hinge" and "spontaneously bent" excitations. The former provides a model for thermal fluctuations of double helix structure which provide enhanced flexibility, while the latter provides a model for DNA-bending proteins binding and unbinding in thermal equilibrium.

Our calculations have emphasized calculation of the end-to-end distributions for finite-length dsDNAs. For molecules under zero force, the probability of zero length of the end-to-end vector gives the "J factor" commonly studied in cyclization experiments. Cloutier and Widom [27] experimentally found a J factor for $\approx$100 bp molecules in excess of 10^4 times larger than that expected from the semiflexible polymer model. This suggests that effects beyond that model facilitate cyclization of short molecules. We have proposed [31] that thermally activated localized flexible "hinges," generated by helix opening, may provide an explanation for those experiments. We find that including $11k_BT$ excitations (depending on DNA sequence, $11k_BT$ can generate a flexible hinge of size from 2 bp to 8 bp long), locally reducing the double helix persistence length to 3 bp, provide a way to generate the large J factors observed experimentally, while retaining the well-established simple semiflexible polymer behavior at larger scales. We have shown that our model produces J factors which converge to the predictions of the semiflexible polymer model beyond 200 bp, and also provide the dsDNA entropic elasticity observed in single-molecule experiments.

We previously noted that the J factor for short molecules was extremely sensitive to small changes in the hinge defect excitation energy, with a $1k_BT$ change in μ resulting in a roughly ninefold change in J for 135 bp [31]. Recently, Ranjith and colleagues [46] have introduced sequence disorder into our model, and have found strong dependence of J factor on sequence reminiscent of the experimental results. However, it remains to be seen if a sequence-dependent theory can explain the sequence-dependence observed by Widom and Cloutier [27].

In this paper we have also considered the effect of the boundary conditions for cyclization on the J factor. We have found an extremely strong dependence for short chains, depending on whether cyclization is forced to occur with ends "parallel" (as is thought to be the case for cyclization using DNA ligase), or ends "antiparallel."

The antiparallel closure could apply for a protein which binds two sequences so as to form a "hairpin" out of the double helix, e.g., Gal repressor [10]. The latter, antiparallel, boundary condition hugely enhances cyclization of short chains, since it reduces the number of hinge defects necessary to close the chain without bending energy from two, to only one. In calculations of the number of excited hinge defects we have shown that our model for parallel end closure indeed predicts that two hinges occur for $<$150 bp molecules, while for antiparallel closure, only one hinge is generated for short molecules. Results similar to the antiparallel case occur when a "free" closure boundary condition is used, in accord with Ranjith *et al.* [46] who further showed that in this case the single defect appears near the midpoint of the modecule; such a boundary condition might be realizable experimentally in a cyclization experiment that uses flexible, reactive end labels on dsDNA segments.

To account for DNA looping facilitated by DNA bending protein that binds to specific DNA sequences, we have calculated J factors for short dsDNA segments with a single bend located at the middle of the molecule. The effect of combination of a single bend, thermally excited flexible hinges, and end orientation boundary conditions on J factor were also computed.

The calculations of the present paper do not account for the torsional rigidity of dsDNA, which for short ($<$200 bp) segments introduces appreciable helix "phasing" effects in cyclization using DNA ligase. The twist rigidity of the double helix is sufficiently large that oscillations of J with segment length, with a period of about 10 bases, occur [29]. Our calculations are effectively for dsDNA with no twist rigidity, where we require only a tangent vector alignment boundary condition for cyclization; we give J factors intermediate between the "in phase" and "out of phase" limits [29,30]. Including the torsional rigidity and linking number constraint necessary to close a dsDNA is possible; however, for dsDNA segments shorter than 200 bp, it is plausible that

the same defects which facilitate dsDNA bending will facilitate dsDNA twisting, reducing the effect of the linking number constraint on J. Therefore a complete theory will require not just linking number constraint but also possible softening of the twist rigidity due to the hinge defects.

We have also carried out calculations for how J factor is suppressed by applied force. For loop sizes in the 500 bp to 2 kb range there is an enormous suppression by even small forces (0.2 pN). A wide variety of experimental possibilities exist for studying force dependence of looping, including loop-forming transcription factors [17,18] and loop-forming restriction enzymes [19]. Although some experiments of this type [18] have been done, the strong loop-length, force and boundary-condition dependences of the cyclization probability remain to be explored.

Finally, we want to draw reader's attention to other recently reported methods to calculate the end-to-end distance distribution of a chain molecule: using infinite continued fractions, Ref. [48] provides methods for exact calculation of the partition function of the stretched semiflexible polymer. Those results can be used to calculate zero-force end-to-end distance distributions of the molecule with "free" (unconstrained) boundary condition [32]. Another related recent paper [49] calculates the polymer end-to-end distribution in the force direction, based on the knowledge of the force-extension curve which can be experimentally obtained. The distribution function calculated from the force-extension curve might be used to analyze the underlying elasticity of the polymer.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation through Grant No. DMR-0203963. The authors thank J. Widom for communicating his experimental data which motivated the work on the influence of excited hinge defects on dsDNA cyclization, and R. Owczarzy for his advice regarding DNA strand separation. The authors also thank P. Wiggins for pointing out that the results of Ref. [31] imply that two hinges dominate cyclization of short dsDNAs in our excited-defect model, in accord with his independent calculation [32].

APPENDIX: END-TO-END VECTOR MOMENT RELATIONS FOR STRETCHED POLYMERS

A simple relation between moments of the end-to-end vector of a stretched polymer and the magnitude of the external force holds when the energy function of the polymer is of the form

$$E = E_{\text{polymer}} - \mathbf{f}\cdot\mathbf{R}, \tag{A1}$$

where $\mathbf{f}$ is the applied force, $\mathbf{R}$ is the end-to-end vector of the polymer, and E_{polymer} is the energy associated with the polymer itself, i.e., elastic energy, self-interactions, or interactions with other molecules free in solution.

The cumulants of end-to-end vector components are generated using derivatives on force components, e.g.,

$$\langle R_i\rangle = k_BT\frac{\partial \ln Z}{\partial f_i},$$

$$\langle R_i^2\rangle - \langle R_i\rangle^2 = (k_BT)^2\frac{\partial \ln Z}{\partial (f_i)^2}. \tag{A2}$$

In the context of this paper, these are cumulant moments of the direction-projected distribution functions (13).

We will assume that E_{polymer} is invariant under space rotation, which is true for a broad class of polymer models, including those describing dsDNA under torsional stress [43], and with either continuum or discretized degrees of freedom [33]. In this case, the partition function must be a function only of the magnitude of the force, since this is the only preferred direction in the problem, i.e., $Z=Z(f)$. For the case where the applied force is near to the $\mathbf{z}$ direction, i.e., $|\mathbf{f}_\perp| \ll f_z$ where $\mathbf{f}_\perp \equiv (f_x, f_y)$, writing the expansion $f=[f_z^2 + f_\perp^2]^{1/2} = f_z + f_\perp^2/(2f_z) + \mathcal{O}(f_\perp^4)$ allows us to expand $\ln Z(f)$ in force components:

$$\ln Z(f) = \ln Z(f_z) + \frac{f_\perp^2}{2f_z}\frac{d\ln Z(f_z)}{df_z} + \mathcal{O}(f_\perp^4). \tag{A3}$$

Using (A2), the first moment of the end-to-end vector in the force direction, i.e., the extension, is

$$\langle R_z\rangle = k_BT \left.\frac{\partial \ln Z(f_z)}{\partial f_z}\right|_{f_\perp=0, f_z=f} = k_BT\frac{d\ln Z(f)}{df}. \tag{A4}$$

The second moment of the end-to-end vector in any direction transverse to the applied force is

$$\langle R_x^2\rangle = (k_BT)^2 \left.\frac{\partial^2 \ln Z}{\partial f_x^2}\right|_{f_\perp=0, f_z=f} = \frac{(k_BT)^2}{f}\frac{d\ln Z(f)}{df}. \tag{A5}$$

Therefore, for a polymer stretched by a force f in the z direction, we have

$$f = k_BT\frac{\langle R_z\rangle}{\langle R_x^2\rangle}. \tag{A6}$$

This is easily generalized to show that the $2n$th cumulant of the transverse distribution is determined by the cumulants of the longitudinal distribution:

$$\langle R_x^{2n}\rangle_c = \sum_{\ell=1}^{n}\frac{(k_BT/f)^{2n-\ell}}{\ell!}\left(\frac{d^{2n}\sqrt{1+x^2}-1}{dx^{2n}}\right)_{x=1}\langle R_z^\ell\rangle_c. \tag{A7}$$

Accordingly the longitudinal and transverse distributions, $\rho(z)$ and $\rho(x)$, are linearly related:

$$\rho(z) = \int_{-\infty}^{\infty} dx\ \rho(x)\int_{-\infty}^{\infty}\frac{dk}{2\pi}e^{-ikz}e^{ix\sqrt{k^2-2ikf/(k_BT)}}. \tag{A8}$$

These exact relations hold for finite-length or self-interacting

polymers where the end-to-end distributions may be far from Gaussian, and for situations such as low forces where the transverse fluctuations are large compared to the extension. An experimental situation where these relations need not hold is for dsDNAs tethered to a surface, with sufficiently low forces applied so that the surface substantially affects the polymer fluctuations, thus breaking the rotational invariance of Z.

[1] J. Dekker, Trends Biochem. Sci. **28**, 277 (2003).
[2] K. Rippe, P. H. von Hippel, and J. Langowski, Trends Biochem. Sci. **20**, 500 (1995).
[3] S. Oehler, M. Amouyal, P. Kolkhof, B. Wilcken-Bergmann, and B. Muller-Hill, EMBO J. **13**, 3348 (1994).
[4] Y. Chen, and P. A. Rice, Annu. Rev. Biophys. Biomol. Struct. **32**, 135 (2003).
[5] B. M. Weiner, and N. Kleckner, Cell **77**, 977 (1994).
[6] W. F. Marshall, A. Straight, J. F. Marko, J. Swedlow, A. Dernburg, A. Belmont, A. W. Murray, D. A. Agard, and J. W. Sedat, Curr. Biol. **7**, 930 (1997).
[7] R. Scleif, Annu. Rev. Biochem. **61**, 199 (1992).
[8] S. E. Halford, A. J. Welsh, and M. D. Szczelkun, Annu. Rev. Biophys. Biomol. Struct. **33**, 1 (2004).
[9] J. M. Vilar and S. Leibler, J. Mol. Biol. **331**, 981 (2003).
[10] M. Geanacopoulos, G. Vasmatzis, V. B. Zhurkin, and S. Adhya, Nat. Struct. Biol. **8**, 432 (2001); S. Semsey, M. Y. Tolstorukov, K. Virnik, V. B. Zhurkin, and S. Adhya, Genes Dev. **18** 1898 (2004).
[11] V. L. Brandt and D. B. Roth, Curr. Opin. Immunol. **14**, 224 (2002).
[12] T. J. Richmond and C. A. Davey, Nature (London) **423**, 145 (2003).
[13] B. M. Ali, R. Amit, I. Braslavsky, A. B. Oppenhein, O. Gileadi, and J. Stavans, Proc. Natl. Acad. Sci. U.S.A. **98**, 10658 (2001).
[14] R. Amit, A. B. Oppenheim, and J. Stavans, Biophys. J. **84**, 2467 (2003).
[15] J. van Noort, S. Vebrugge, N. Goosen, C. Dekker, and R. T. Dame, Proc. Natl. Acad. Sci. U.S.A. **101**, 6969 (2004).
[16] D. Skoko, B. Wong, R. C. Johnson, and J. F. Marko, Biochemistry **43**, 13867 (2004).
[17] L. Finzi and J. Gelles, Science **267**, 378 (1995).
[18] G. Lia, D. Bensimon, V. Croquette, J. F. Allemand, D. Dunlap, D. E. Lewis, S. Adhya, and L. Finzi, Proc. Natl. Acad. Sci. U.S.A. **100**, 11373 (2003).
[19] J. Yan, D. Skoko, and J. F. Marko, Phys. Rev. E **70**, 011905 (2004).
[20] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struct. Biol. **10**, 279 (2000).
[21] J. Liphardt, B. Onoa, S. B. Smith, I. J. R. Tinoco, and C. Bustamante, Science **292**, 733 (2002).
[22] D. Shore, J. Langowski, and R. L. Baldwin, Proc. Natl. Acad. Sci. U.S.A. **79**, 4833 (1981).
[23] P. J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
[24] Y. Zhang and D. M. Crothers, Biophys. J. **84**, 136 (2003).
[25] Y. Zhang and D. M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **100**, 3161 (2003).
[26] H. Arthanari, K. Wojtuszewski, I. Mukerji, and P. H. Bolton, Biophys. J. **86**, 1625 (2004).
[27] T. E. Cloutier and J. Widom, Mol. Cell **14**, 355 (2004).
[28] H. Yamakawa and W. H. Stockmayer, J. Chem. Phys. **57**, 2843 (1972).
[29] J. Shimada and H. Yamakawa, Macromolecules **17**, 689 (1984).
[30] A. Podtelezhnikov and A. Vologodskii, Macromolecules **30**, 6668 (1997).
[31] J. Yan and J. F. Marko, Phys. Rev. Lett. **93**, 108108 (2004).
[32] P. A. Wiggins, R. Phillips, and P. C. Nelson Phys. Rev. E **71**, 021909 (2005).
[33] J. Yan and J. F. Marko, Phys. Rev. E **68**, 011905 (2003).
[34] M. D. Wang, M. J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S. M. Block, Science **282**, 902 (1998).
[35] Eric W. Weisstein, "Wigner 3j-Symbol," from MathWorld–A Wolfram Web Resource, http://mathworld.wolfram.com/Wigner3j-Symbol.html. The usage of spherical harmonics and Wigner-3J symbols is important for the computation in this paper. It allows us to get analytical expressions of the matrix elements, which is key to our semianalytical calculation of the problem. The analytical matrix expression also contributes to speeding up the computation.
[36] T. Strick, J. Allemand, V. Croquette, and D. Bensimon, Prog. Biophys. Mol. Biol. **74**, 115 (2000).
[37] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[38] C. Storm and P. C. Nelson, Phys. Rev. E **67**, 051906 (2003).
[39] M. Doi and S. F. Edwards, *Theory of Polymer Dynamics* (New York, Oxford, 1985).
[40] J. Santalucia, Proc. Natl. Acad. Sci. U.S.A. **95**, 1460 (1998).
[41] D. H. Matthews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999), see Table 17.
[42] S. Cocco, J. Yan, J.-F. Léger, D. Chatenay, and J. F. Marko, Phys. Rev. E **70**, 011910 (2004).
[43] J. F. Marko and E. D. Siggia, Phys. Rev. E **52**, 2912 (1995).
[44] J. F. Marko and E. D. Siggia, Biophys. J. **73**, 2173 (1997).
[45] S. Sankararaman and J. F. Marko, Phys. Rev. E **71**, 021911 (2005).
[46] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon, Phys. Rev. Lett. **94**, 138102 (2005).
[47] D. M. Crothers, T. E. Haran, and J. G. Nadeau, J. Biol. Chem. **265**, 7093 (1990).
[48] A. J. Spakowitz and Z. G. Wang, Macromolecules **37**, 5814 (2004).
[49] L. Dai and F. Liu, J. Chem. Phys. **119**, 8124 (2003).