

Generalized theory of semiflexible polymers

Paul A. Wiggins*

Division of Physics, Mathematics, & Astronomy, California Institute of Technology, Pasadena, California 91125, USA

Philip C. Nelson[†]

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

(Received 5 August 2005; published 7 March 2006)

DNA bending on length scales shorter than a persistence length plays an integral role in the translation of genetic information from DNA to cellular function. Quantitative experimental studies of these biological systems have led to a renewed interest in the polymer mechanics relevant for describing the conformational free energy of DNA bending induced by protein-DNA complexes. Recent experimental results from DNA cyclization studies have cast doubt on the applicability of the canonical semiflexible polymer theory, the wormlike chain (WLC) model, to DNA bending on biologically relevant length scales. This paper develops a theory of the chain statistics of a class of generalized semiflexible polymer models. Our focus is on the theoretical development of these models and the calculation of experimental observables. To illustrate our methods, we focus on a specific, illustrative model of DNA bending. We show that the WLC model generically describes the long-length-scale chain statistics of semiflexible polymers, as predicted by renormalization group arguments. In particular, we show that either the WLC or our present model adequately describes force-extension, solution scattering, and long-contour-length cyclization experiments, regardless of the details of DNA bend elasticity. In contrast, experiments sensitive to short-length-scale chain behavior can in principle reveal dramatic departures from the linear elastic behavior assumed in the WLC model. We demonstrate this explicitly by showing that our toy model can reproduce the anomalously large short-contour-length cyclization J factors recently measured by Cloutier and Widom. Finally, we discuss the applicability of these models to DNA chain statistics in the context of future experiments.

DOI: [10.1103/PhysRevE.73.031906](https://doi.org/10.1103/PhysRevE.73.031906)

PACS number(s): 87.14.Gg, 87.15.La, 82.35.Pq, 36.20.Hb

I. INTRODUCTION

The statistical mechanics of linear polymers has long attracted the attention of physicists and chemists alike. The mechanics of DNA is of considerable biological relevance to describing the free energy landscape controlling protein-induced DNA bending. These protein-DNA interactions are of central importance to cellular function on a microscopic scale, from chromosomal DNA packaging, to transcription and gene regulation, to viral packaging [1]. Protein-DNA interactions typically induce short-length-scale DNA bending which couples the chemical and physical properties of DNA [2–4].

A particularly important and successful application of polymer statistics has been in the description of double stranded DNA (dsDNA) by the wormlike chain model (WLC). In the WLC model, DNA is modeled as a fluctuating, linearly elastic rod. This simple model has been remarkably successful in describing many aspects of DNA mechanics and the statistics of semiflexible polymers generally. In particular, the WLC describes the extension of a single dsDNA molecule under an external force with impressive precision [5,6].

Despite the notable theoretical and experimental success of the wormlike chain model, recent DNA mechanics studies probing the high-curvature bending of DNA have cast doubt on its applicability. The high-curvature regime is experimentally accessible in cyclization studies, which measure the propensity of DNA to form cyclized molecules relative to bimolecular association. The experiments of Cloutier and Widom [7] suggest that the probability of highly bent configurations may be three to four orders of magnitude larger than predicted by the WLC model. In still more recent cyclization studies, Vologodskii and co-workers claim that the WLC model does accurately describe the cyclization of short DNA sequences [8] but for a persistence length shorter than the value commonly accepted. Shroff *et al.* [9] have also performed a measurement that directly probes the force required to bend short sequences of DNA. The measured forces appear to be significantly less than predicted by the WLC model (see Sec. IV A below).

With the current experimental situation still in flux, it seems imperative to reevaluate the WLC model *theoretically*. We wish to answer the following questions. (i) How could such a simple theory hope to describe a complex molecule like DNA? (ii) More precisely, which classes of experiments would we expect to be successfully described by the WLC model, and which might require a different theory? Do these experiments correspond to the known successes or the recently reported failures of the theory? In other words, we are asking how much room do the classic tests of WLC model leave for generalization of this model, and how completely do these experiments test the WLC model? Finally, we must

*Present address: Whitehead Institute, Cambridge, MA 02142, USA. Electronic address: wiggins@wi.mit.edu; URL: <http://jura.wi.mit.edu/~wiggins/>

[†]Electronic address: nelson@physics.upenn.edu

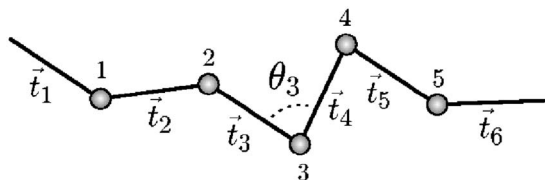


FIG. 1. Link and vertex numbering. The total bending energy is a function of the deflection angles. The deflection angle between links i and $i+1$ is θ_i .

ask (iii) would a breakdown of the WLC model have any biological significance?

The focus of this paper will be the theoretical analysis of these questions and the development and discussion of more general semiflexible polymer models. Although these ideas are widely applicable to polymer statistics in general, the focus of this paper will be exclusively the mechanics of DNA. We shall attempt a synthesis of the existing experimental knowledge to determine which aspects of DNA bending are probed by existing experiments. In particular, we determine which experiments are most sensitive to the DNA mechanics relevant for understanding biological systems.

II. DEFINING DISCRETE LINK THEORIES

To define generalized semiflexible chain theories, we introduce a description of the polymer as a chain of rigid links, length ℓ , connected by torsional-spring vertices. (See Fig. 1.) The link length ℓ has no physical significance, but it is a device necessary for defining an explicit theory. Let θ_i denote the bending angle between links i and $i+1$. As in the WLC model, we will assume that these bending angles are uncorrelated. The statistical mechanics of the chain is therefore completely characterized by a fundamental tangent distribution function, the distribution of bending angles for a single vertex. This fundamental tangent distribution function is the conditional probability density for the final tangent \vec{t}_{i+1} given an initial tangent \vec{t}_i :

$$g(\vec{t}_{i+1}, \vec{t}_i) \equiv q^{-1} \exp[-E(\theta_i)], \quad (1)$$

where we interpret this distribution as arising from an effective bending free energy function $E(\theta)$ (expressed in units of the thermal energy $k_B T$).

The independence of successive links [or equivalently the absence of derivatives in the effective bending free energy function $E(\theta)$] is a hypothesis that must be checked experimentally. The omitted higher-derivative terms would correspond physically to cooperative conformational changes along the polymer, and although there are hints of long-range cooperativity in DNA [10], still the phenomena addressed in this paper do not seem to require such behavior. In addition, our models as formulated omit any effect of DNA sequence on elastic behavior. Although such effects are documented, experimentally random-sequence (and hence natural) DNA is known to behave similarly to DNA without intrinsic bends, even in cyclization [11]. In short, our restricted class of theories is intended as an analytically tractable starting point for the study of the effects of nonlinear-elastic behavior in an

entropy-dominated chain; other kinds of physics can be added to this model as needed.

A. Wormlike chain

The WLC model is one particular realization of Eq. (1), where the bending energy function is harmonic:

$$E_{\text{WLC}}(\theta) = \frac{\kappa}{2\ell} \theta^2, \quad (2)$$

and we take the limit as the link length ℓ goes to zero. The WLC model is characterized by a single constant κ , the effective elastic bend stiffness of the chain. In this limit, E_{WLC} becomes a functional of the rod curvature θ/ℓ . Power-counting arguments then lead us to expect that anharmonic elasticity terms (higher than second order in curvature) will be irrelevant in the continuum limit (or indeed whenever we consider length scales much larger than ℓ), because the curvature has units of $(\text{length})^{-1}$. On length scales that are *not* much larger than ℓ , however, we may expect anharmonic terms to matter. In fact, early atomic force microscopy AFM imaging experiments did see hints of a deviation from WLC expectations on short scales [12]. Other experiments, discussed in Sec. IV, hold the prospect of greater sensitivity to the high-curvature regime.

B. Subelastic chain model

We now introduce an explicit illustrative model for DNA bending that seemingly differs dramatically from the WLC. The results below can be applied to the analysis of any local-elasticity model [that is, any model characterized by Eq. (1)]. We will illustrate the method by using a concrete example in this class of models, and show how to compute experimental observables like the force-extension relation, the cyclization J factor, etc. The model we will study has two key properties: (i) Its bending energy function is softer for high curvature than the WLC; and (ii) nevertheless, it reproduces the successful long-length-scale predictions of the WLC.

We have already described one such model in an earlier paper [13]; similar models were also formulated and solved by Yan and Marko [14,15], Popov and Tkachenko [16], and Ranjith *et al.* [17] (see also [18]). In these models, the high-curvature softening was introduced by allowing “kinking,” or curvature localization: Beyond a critical strain, the DNA’s resistance to bending falls suddenly to zero, or some small value. Although these kinking models reproduced the two desired features mentioned above, they predicted that highly curved DNA should be generically kinked [13]. However, atomic force microscopy imaging of small minicircles usually shows them as round (although kinking can be induced in unusual ionic conditions [19,20], or when DNA binds to proteins [21,22]). Moreover, tightly looped DNA shows enhanced sensitivity to DNase digestion that is not concentrated on a single kink point, but rather is spread throughout the loop [23]. Finally, recent molecular-dynamics simulations of DNA minicircles show the spontaneous formation of bends [24], but without the strand separation envisioned by Yan and Marko. For these reasons, this paper will explore a

class of models with nonlinear DNA elasticity but without the catastrophic breakdown characteristic of kinking.

The effective bending free energy function we will study comes from the observation, well known in continuum mechanics, that a rod bending energy density that is nonconvex in curvature induces kinking when the rod is strongly bent [25]. To avoid kinking, we must therefore require that our effective bending energy density be everywhere a convex function of curvature, at least on length scales observable via electron microscopy (EM) or AFM imaging.

A simple choice of bending energy function that meets the criteria above, and is radically different from the WLC model, is

$$E \text{ SEC}(\theta) = B|\theta|, \quad (3)$$

which defines a family of models parametrized by B and ℓ that we call “subelastich chain” (SEC) models. We will show that taking $B=5.3$ and $\ell=5$ nm gives rise to a model with the persistence length $\xi=53$ nm needed to describe the long-scale behavior of DNA in moderate salt solution [26]. As a final motivation, AFM studies of the tangent-tangent correlation of DNA adsorbed to mica appear to fit a bending energy of roughly this functional form [27].

We emphasize that the SEC defined by Eq. (3) is an illustrative model, and not intended as a realistic representation of DNA at the single-link length scale. In particular, the nonanalytic behavior of Eq. (3) at $\theta=0$ is not meant to be taken literally. (For instance, we could introduce a small quadratic regime to the bending energy without significantly affecting the predictions of the model or the fit to experiment. This particular functional form summarizes the theory concisely with a single constant B .)

C. The propagator and composition

The locality assumption in the definition of the bending energy implies that each vertex bends independently. Computing the tangent distribution functions for chains of several links is therefore straightforward. These conditional probabilities are simply the product of conditional probabilities for single vertices, summed over the orientations of the intermediate tangents. These configurational sums can be concisely written by introducing the propagation operator (or transfer matrix [15])

$$\mathcal{G} \equiv \int dt dt' |\vec{t}\rangle g(\vec{t}, \vec{t}') \langle \vec{t}'|, \quad (4)$$

where $\langle |$ and $| \rangle$ is the canonical bra ket notation of statistical mechanics (or quantum mechanics) [28]. The propagation operator \mathcal{G} applied on a state gives the state (probability distribution) after one additional link. The N -link tangent distribution function is

$$G(\vec{t}; \vec{t}'; N\ell) = \langle \vec{t} | \mathcal{G}^N | \vec{t}' \rangle, \quad (5)$$

where the weighted sum, or path integral, over all intermediate configurations is implicit.

Due to the rotational invariance of the polymer, the propagation operator is diagonal in the angular momentum basis

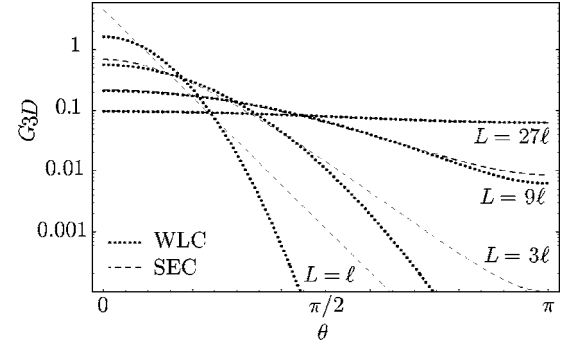


FIG. 2. Evolution of the three-dimensional tangent distribution function $G_{3D}(\vec{t}; \vec{t}'; L)$ with increasing separation L . The WLC and SEC tangent distribution functions are plotted as functions of the deflection angle θ for several contour lengths. The linear dependence of the SEC bending energy on the deflection angle, visible in the fundamental distribution function ($L=\ell=5$ nm), is lost at longer contour length. For $L \gg \ell$, the tangent distribution approaches the WLC distribution function with a persistence length of 53 nm despite dramatically different behavior at short contour length. This loss of the short-length structure of the tangent distribution function is universal and explains the success of the WLC model in describing many semiflexible polymer phenomena.

$$\mathcal{G} = \sum_{l, \mathbf{m}} g_l |l\mathbf{m}\rangle \langle l\mathbf{m}|. \quad (6)$$

In three dimensions, the spherical harmonics relate the tangent and angular momentum representations [28].

Figure 2 shows an explicit calculation of the N -link tangent distribution functions for the SEC and the WLC models. This figure explicitly illustrates the scale dependence of statistical mechanics theories. For short-contour-length chains, the WLC and SEC theories make dramatically different predictions, but as the contour length of the chain increases, the differences between the distribution functions of the two theories decrease until at long contour length, the theories are essentially indistinguishable. This is the essence of the renormalization group: at short length scales, the mechanics of the chain can be extremely complicated but the thermal fluctuations sum over many intermediate configurations and hide the underlying complexity on longer length scales. We shall show this is a general feature of semiflexible polymer models in Sec. II E.

D. Contour length continuation

Since we will frequently be interested in the properties of the polymer on length scales much longer than the fundamental link length ℓ , it is useful to introduce a “Hamiltonian” operator defined by

$$\mathcal{H} \equiv -\ell^{-1} \ln \mathcal{G} = \sum_{l, \mathbf{m}} h_l |l\mathbf{m}\rangle \langle l\mathbf{m}|. \quad (7)$$

\mathcal{H} is also diagonal in the angular momentum representation with eigenvalues $h_l = -\ell^{-1} \ln g_l$. The Hamiltonian spectrum for the the SEC and WLC models is compared in Fig. 3.

The advantage of this reformulation of the distribution function is that it introduces a natural extension to fractional

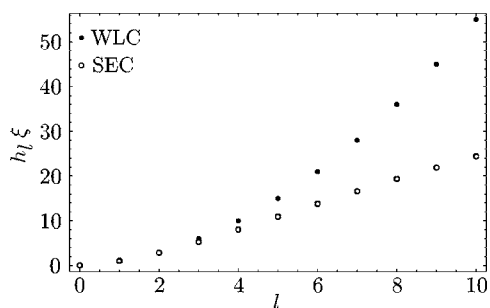


FIG. 3. The eigenspectrum of \mathcal{H} for the SEC and WLC models. The eigenvalues $\{h_l\}$ of the Hamiltonian operator for the WLC and SEC theories are compared as a function of the angular quantum number l . Both theories have an identical persistence length $\xi = h_1^{-1} = 53$ nm. The eigenvalues of the Hamiltonian are coincident for small l but diverge as l increases. The l th moment of the distribution function decays at large contour separation L as $\exp(-h_l L)$. The larger eigenvalues of the Hamiltonian, for which the two theories differ, are therefore relevant only for small L , implying that the SEC and WLC chain statistics are identical for long-contour-length chains.

numbers of links by replacing $N\ell$ by the contour length L defined for all positive real numbers:

$$\mathcal{G}(L) \equiv \exp(-\mathcal{H}L), \quad (8)$$

although rigorously, it is understood that this function is only defined for contour lengths equal to an integral number of links.

E. The meaning of persistence length

What is the meaning of persistence length in general models like the ones we have described? Persistence length describes the length scale on which the polymer maintains its tangent orientation. For the WLC model in D dimensions, the correlation of the tangents decays exponentially,

$$\langle \vec{t}(0) \cdot \vec{t}(L) \rangle = \exp(-L/\xi_D), \quad (9)$$

and the decay length

$$\xi_D = 2\kappa/(D-1) \quad (10)$$

is called the persistence length where κ is the bending modulus that appears in the WLC bending energy [Eq. (2)]. In general models, the tangent persistence [Eq. (9)] has the same functional form and therefore defines a persistence length ξ , but the persistence length is no longer related to the stiffness by Eq. (10).

The tangent persistence corresponds to the $l=1$ mode of the propagator. (In the quantum mechanical correspondence, \vec{t} is a vector and creates a state of spin 1.) Comparing Eqs. (8) and (9), the persistence length is

$$\xi_D \equiv h_1^{-1}, \quad (11)$$

where h_1 is the $l=1$ eigenvalue of the Hamiltonian. Note that since Eq. (9) can be integrated to get both the mean-squared end distance and the radius of gyration of the polymer chain, these expressions have the same form in general models as they do in the WLC model.

We can immediately exploit Eq. (11), with the help of Eq. (A2), to analyze the SEC model. The persistence length, computed for the SEC model in three dimensions with $B = 5.3$ and $\ell = 5$ nm, is 53 nm, which matches solution measurements.

F. All models flow to WLC

We now examine the tangent distribution function for a stiff polymer and show that the WLC model is universal at long contour length, as predicted above. Our explicit computations of the SEC tangent distribution function in Fig. 2 have already provided one explicit example of this behavior, but we address this question generally in this section.

By definition the stiff polymer limit implies that the fundamental tangent distribution function g is narrowly distributed around zero deflection. We exploit this fact by Taylor-expanding the basis functions in the deflection angle and computing the Hamiltonian operator Eq. (7). The details of this calculation are described in Appendix B. To lowest non-trivial order in the small quantity ℓ/ξ , the Hamiltonian operator for generalized theories is

$$\mathcal{H} = \mathcal{H}_{\text{WLC}} + \mathcal{O}[(\ell/\xi)^2 \mathcal{L}^4], \quad (12)$$

identical to the Hamiltonian operator of the WLC when the eigenvalue of the total angular momentum operator \mathcal{L}^2 is small. Thus, like the spectrum of the SEC theory (Fig. 3), the spectrum of a general model of stiff polymers coincides with the WLC at low angular quantum number l . These modes generally have the lowest eigenvalues, and therefore are the most relevant at long contour length [see Eq. (8)]. For example, the eigenmodes of the tangent distribution function caused by the sharp cusp in the fundamental tangent distribution Eq. (3) (see Fig. 2) decay quickly with contour length, resulting in a distribution that is well approximated by the WLC model.

III. THE SPATIAL DISTRIBUTION

For most applications, it is the *spatial* distribution of the polymer, rather than the tangent distribution function, that is of phenomenological interest. For example, in light scattering, the force-extension relation, cyclization, and looping, it is the spatial distribution function that is directly observable. In this section, we shall develop a formalism for computing the spatial distribution function. Our focus will be exclusively on three dimensions, but computations in other dimensions are a simple extension of the methods discussed here. Other methods for calculating such distributions are given in [29,30].

The tangent-spatial distribution function is the probability density of end displacement \vec{X} and final tangent \vec{t}_f given an initial tangent \vec{t}_i for an arclength L chain. It is convenient to write the tangent-spatial distribution in terms of the spatial δ function [31]

$$G(\vec{X}; \vec{t}_f, \vec{t}_i; L) = \langle \vec{t}_f | e^{-\mathcal{H}L} \delta^3[\vec{X} - \vec{X}(L)] | \vec{t}_i \rangle, \quad (13)$$

where the δ function enforces the spatial constraint. To compute the tangent-spatial distribution function, we introduce

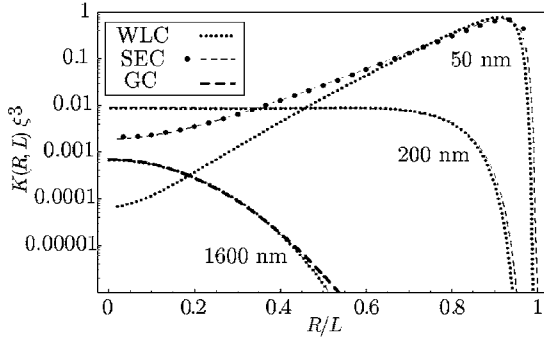


FIG. 4. The spatial distribution $K(R;L)$ for the WLC and SEC theories [see Eq. (16)]. All curves except the black dotted curve have been computed using the inverse transform technique. To check the validity of the technique, the black dots show a direct Monte Carlo integration for the shortest-contour-length SEC curve (thin dashed). We have chosen the contour lengths of the chains to illustrate two types of renormalization. At 50 nm for large deflection ($R/L \sim 0$), the SEC (thin dashed) and WLC (dotted) theories differ by two orders of magnitude. For a 200 nm contour length, SEC and WLC predict nearly identical distributions, but this distribution is clearly not Gaussian. For very long contour lengths, however, these theories do renormalize to the Gaussian chain model (fat dashed), as expected on general grounds.

an operator-valued spatial distribution function [32]

$$\mathcal{G}(\vec{X};L) \equiv \int d\vec{r} d\vec{r}' |\vec{r}\rangle G(\vec{X};\vec{r},\vec{r}';L) \langle \vec{r}'|, \quad (14)$$

which again allows us to keep the tangents implicit in our notation. As before, it will be convenient to work in the angular momentum basis with the matrix elements

$$G_{lm'l'm'}(\vec{X};L) \equiv \langle l m | \mathcal{G}(\vec{X};L) | l' m' \rangle, \quad (15)$$

since this basis diagonalizes the Hamiltonian (although not the spatial propagator).

Finding the spatial propagator reduces to the ability to explicitly compute all the $G_{lm'l'm'}$. In Appendix C, we derive near-exact expressions for the Fourier-Laplace transformed matrix elements using a method analogous to that exploited in Ref. [32].

The chain-end distance distribution is the probability density of the end displacement vector \vec{X} :

$$K(\vec{X};L) \equiv G_{0000}(\vec{X};L). \quad (16)$$

The corresponding distribution for $R = \|\vec{X}\|$ is shown in Fig. 4 for WLC and SEC chains of several contour lengths. The numerical techniques applied in this computation are described in Appendix D. These results again display the renormalization group flow toward the WLC model at long contour length. Although the two theories make dramatically different predictions for short-contour-length chains, the predictions coincide at long contour length.

IV. COMPARISON TO EXPERIMENTS

In the theoretical development above, we argued that the WLC model generically describes semiflexible polymers at

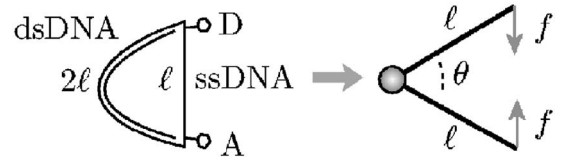


FIG. 5. Measuring the deflection force of highly bent short sequences of DNA using a FRET force sensor [9]. Cyclized sequences of single-stranded DNA are hybridized with shorter complementary sequences. The single-stranded region of DNA acts as a force sensor. The external force is measured by the FRET efficiency of FRET dyes (D and A) positioned at either end of the force sensor. For a rough estimate, we model this molecule as two links under a deflection force load f induced by the single-stranded DNA linker. The deflection angle θ is roughly $2\pi/3$ since the single-stranded DNA is roughly the same length as the link length.

long contour length regardless of the detailed form of the short-contour-length tangent distribution function (or equivalently the bending energy function). Many of the classic tests of the WLC description of DNA (for instance measurements of the force-extension relation) do not probe short-length-scale bending, and hence cannot distinguish between different models in the class that we study. In this section, we explicitly compute several experimental observables for the SEC model and compare them with the predictions of the WLC model and experimental measurements.

A. Measurements of the short-length-scale bending energy

The force required to tightly bend short sections of DNA has recently been directly measured by Liphardt and co-workers [9] via a fluorescence resonance energy transfer (FRET) force sensor. In this experiment, a sequence of DNA 9.18 nm in length is tightly bent by a linking sequence of single-stranded DNA as illustrated in Fig. 5. This contour length is represented in our theory by two links ($\ell = 5$ nm) and a single vertex. The deflection angle is roughly $2\pi/3$. It is straightforward to estimate the deflection force in both the discrete WLC and SEC models:

$$f_{\text{SEC}} \approx \frac{A}{\ell \cos \pi/6} \approx 5.5 \text{ pN}, \quad (17)$$

$$f_{\text{WLC}} \approx \frac{\xi \theta}{\ell^2 \cos \pi/6} \approx 25 \text{ pN}. \quad (18)$$

(In this estimate, we have used $\ell = 4.6$ nm, half the contour length of the dsDNA.) The experimentally measured force 6 ± 5 pN is approximately equal to that predicted by the SEC model but is more than a factor of 2 smaller than that predicted by the harmonic-elasticity model (WLC). These experiments again indicate that the WLC model fails to describe the high-curvature bending of short sequences of DNA. At least at this deflection angle, the SEC model approximately predicts the deflection force. Note that if the kinking model of Refs. [13,14] literally described short sequences of DNA, this force would be zero, contrary to the experiment—another motivation for our introduction of generalized elasticity models.

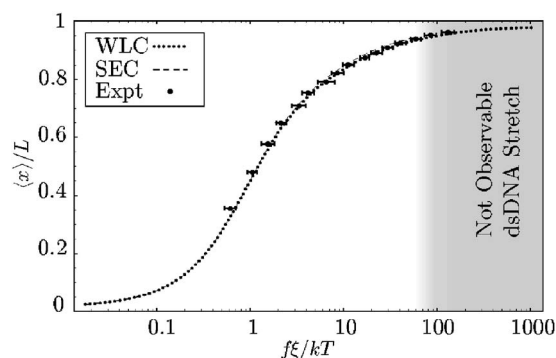


FIG. 6. Force extension for the WLC and SEC models compared with experimental measurements [26]. The WLC model was fitted to the experimental data to determine the contour length and persistence length ($\xi=53$ nm). Despite the dissimilar short-length-scale tangent distribution function, the behavior of the polymer under an external force is nearly identical. For forces greater than 10 pN, intrinsic stretching becomes important, obscuring the entropic part of the response.

B. Force extension

The force-extension of single polymer molecules has long been the subject of experimental interest [5,6]. The experimental observable in these experiments, the extension of the polymer under an external force, can be directly computed from the spatial distribution function. Typically this force is applied to a bead, tethered to the polymer, using an optical or magnetic trap [5,33,34]. The restoring force against extension is entropic in nature (for inextensible polymers). This entropic force is induced by the reduction in the number of microconfigurations available to the chain as the extension is increased.

The computation of the force-extension of the polymer is a straightforward extension of the results discussed in Sec. III. The partition function for a polymer under a constant external tension is related to the Fourier transform of the spatial distribution function via an analytic continuation of the wave number [35,36]. The numerical technique applied in this computation is described in Appendix D.

The successful comparison of the WLC to single-molecule force-extension data has been described as a strict test of the WLC model [5]. But how do other semiflexible polymer models compare? Can these models also reproduce the precise fit to experiment? The force-extensions for the SEC and WLC models are compared in Fig. 6. Despite the drastically different bending energy of the SEC model on short length scales, thermal fluctuations disguise these differences and give rise to an extension almost identical to that in the WLC model. In retrospect, these results are hardly surprising. The theories are identical at small force, because the extension is dominated by contributions from bending modes of wavelength longer than $k_B T/f$. Physically, the rare high-curvature bends, which could lead to differences between models, are even further suppressed by the application of tension. On the other hand, the extension at high force is already nearly 100%, so small differences between models in the entropic contribution are again hard to detect. Thus, force-extension measurements do probe the persistence

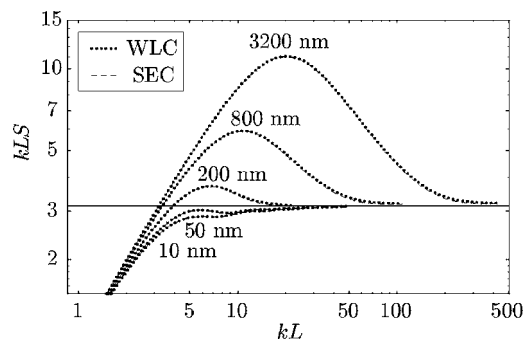


FIG. 7. The structure factor for the SEC and WLC models. In the figure above, the structure factor $S(k)$, scaled by the nondimensionalized wave number, is plotted for several contour lengths. The curves are nearly identical for the two theories since the structure factor is dominated by thermally accessible configurations. Although rare, high-curvature configurations are orders of magnitude more probable in the SEC than in the WLC theory, these configurations are still too rare to significantly affect the structure factor. At large k , both WLC and SEC are straight, which gives an asymptotic limit for large wave number $S(k) \rightarrow \pi/Lk$.

length and the inextensibility of DNA, but they do not effectively probe DNA elasticity on the length scales of interest for many biological processes.

C. Structure factor

Another experimental observable used to characterize polymers is the structure factor, measured by static light scattering, small-angle x-ray scattering, and neutron scattering experiments. Measurements of the structure factor can probe the polymer configuration on a wide range of length scales. Symbolically the structure factor is

$$S(\vec{k}) \equiv \frac{1}{L^2} \int_0^L ds ds' \langle e^{i\vec{k} \cdot [\vec{X}(s) - \vec{X}(s')]} \rangle, \quad (19)$$

where $\vec{X}(s)$ is the position of the polymer at arclength s and we have included an extra factor of the contour length in the denominator to make the structure factor dimensionless [36]. At high wave number, the structure factor is sensitive to short-length-scale physics, whereas the contour length and radius of gyration are probed by low wave number. The computation of the structure factor is also a straightforward extension of the results discussed in Sec. III [36]. The numerical technique applied in this computation is described in Appendix D.

The structure factors for several lengths of DNA are shown for both the SEC and the WLC models in Fig. 7. Again we find that the two theories make nearly identical predictions. The reasoning is again similar to that explained for force-extension. The two theories make dramatically different predictions for rare, highly bent configurations but the structure factor is dominated not by these rare high-curvature configurations but by typical thermal bending. We therefore find that the structure factor, like force-extension, does not effectively probe the high-curvature statistics of the polymer.

D. Cyclization

Cyclization experiments provide a bulk experimental method for probing the probability of rare, highly curved DNA configurations [7,8,37]. In these experiments, linear double-stranded sequences with complementary single-stranded ends are ligated into cyclized sequences [4,38–41]. The cyclization reaction proceeds via the capture of rare, thermally activated configurations and is thought to be very similar to the process by which looped DNA-protein complexes are formed.

The cyclization assay is performed under conditions such that the ligation reaction samples the equilibrium populations of unligated cyclized and oligomerized polymers [4]. The ratio of the cyclization equilibrium constant (K_C) to the dimerization equilibrium constant (K_D) is called the Jacobson-Stockmayer factor [38] (or J factor), and is proportional to the tangent-spatial distribution function of the polymer [4,40]

$$J \equiv K_C/K_D = 4\pi G(0; \vec{t}, \vec{t}; L) = \text{tr } \mathcal{G}(0; L), \quad (20)$$

where G is the tangent-spatial distribution function for end-to-end displacement 0 and aligned end tangents, for a contour length L polymer. The J factor can also be written as the trace of the spatial propagator. (The matrix elements of the spatial propagator are written explicitly in Appendix C.) Physically, the J factor is proportional to the concentration of one end at the other with the correct (aligned) orientation for hybridization.

Our analysis neglects the condition that DNA twist must also be aligned, which requires the use of models including the twist degree of freedom. This additional constraint modulates the J factor with a 10.5 base pair (bp) period equal to the helical repeat. Our interest here is in the value of the J factor averaged over a helical repeat, for which the effects of twist can be roughly ignored [31].

Figure 8 compares the cyclization J factor for the SEC and WLC theories. The numerical techniques applied in this computation are described in Appendix D. The J factors for sequences with contour lengths greater than two persistence lengths have long been known to match the predictions of the WLC model [4,39]. For sequences shorter than ≈ 200 bp, the figure illustrates the short-contour-length deviation of WLC and SEC predictions. For example, for contour lengths of roughly 0.6 persistence lengths, corresponding to loops with approximately the same radius of curvature as DNA bound to histones in nucleosome complexes, the SEC model J factor is three orders of magnitude larger than predicted by the WLC model, in rough agreement with cyclization measurements of Cloutier and Widom [7], as illustrated in Fig. 8.

The qualitative picture illustrated in Fig. 8 is the generic situation in general models: the WLC model describes long-contour-length chain statistics, but may nevertheless fail at sufficiently short contour length. These results were qualitatively predicted by the ideas we have discussed throughout the paper. In principle, cyclization assays are a powerful technique for probing the short-contour-length chain statistics of DNA, since (i) they probe the chain statistics of DNA in a way that is qualitatively similar to biological DNA loop-

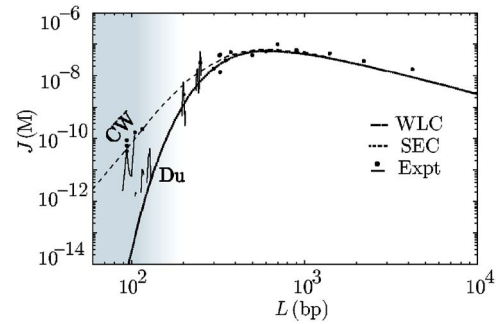


FIG. 8. (Color online) The cyclization J factor probes high-curvature chain statistics. In this figure, the cyclization J factor in units of molarity is plotted for the WLC (dotted curve) and SEC (thin dashed curve) models and compared with experimental measurements (circles) [4,7,8,11,37,39]. The theoretical curves do not include the twist-induced modulation visible in the continuous sets of experimental data (solid curves) [8,11,31,37,39]. The renormalization group predicts that the SEC model will be identical to the WLC for long-contour-length sequences. But, for sequences shorter than ≈ 200 bp, the short-contour-length chain statistics become important and the SEC J factor diverges from the WLC prediction. In fact, for 94 bp sequences, the SEC J factor is three orders of magnitude larger than that predicted by the WLC model, roughly matching the J factors measured by Cloutier and Widom [7,37] (CW). [Other measurements, by Du *et al.* (Du), appear to be commensurate with the predictions of the WLC model with an unusually short persistence length; see text.]

ing applications and (ii) they are extremely sensitive to the differences between models at short contour lengths. In practice, however, the continuing disagreements between existing experiments remind us that cyclization is a complex process (see Sec. V below). Other measurements of the distribution of one point of a chain in the neighborhood of another, for example using Förster resonance electron transfer, may prove to be easier to interpret.

E. Beyond the J factor

The J factor is not the only effective concentration of interest. DNA looping is integral to the function of many gene regulatory proteins. The affinity of these proteins for DNA, and therefore their function, depends sensitively on the looping free energy, or equivalently the effective concentration of the looped DNA. (For instance, see Refs. [42–44].) Once the geometry of the loop is known—the displacement of the binding sites (\vec{X}) and the orientation of the bound DNA (\vec{t} and \vec{t}')—both the SEC and WLC models make predictions for the effective concentration:

$$(\text{effective concentration}) = 4\pi G(\vec{X}; \vec{t}, \vec{t}'; L). \quad (21)$$

These statistical mechanics predictions can then be directly compared with quantitative measurements of gene expression [43,44] and *in vitro* experiments [45].

More general cyclization measurements may also be performed. For instance, cyclizing sequences with two short single-stranded gaps could be used to probe short-length-scale DNA mechanics. The short single-stranded sequences

are very flexible and can be approximated as free hinges. This technique could be exploited to directly measure the spatial distribution function shown in Fig. 4 for very short sequences of DNA.

V. DISCUSSION

A. The SEC

Sections II B–IV A introduced the SEC as a toy model for DNA bending, motivated by several physical measurements on DNA. We proceeded to show that this simple model exhibited the long-length-scale chain statistics of the WLC model, despite dramatically increasing the predicted probability of high-curvature configurations. In particular, we showed that the SEC model yields a cyclization J factor in agreement with the measurements of Cloutier and Widom [7]. More generally, we argued that this type of deviation from WLC behavior is generic in semiflexible chain models.

These potential deviations of DNA chain statistics from the wormlike chain model at short contour length are quite relevant for structural biology, where typical radius of curvature induced by DNA binding proteins is on order nanometers or tens of nanometers, not persistence lengths. For example, the radius of curvature of DNA bound in a nucleosome complex is roughly 6 nm. The structure of this complex shows sharp bends, but no sign of melting, consistent with our SEC model [22]. Similarly, DNA looped by a gene regulatory protein is typically bent on short length scales [42]. If DNA is described by the SEC model, these tightly bent DNA-protein complexes are much more stable than predicted by the WLC model. A quantitative understanding of biological DNA bending therefore awaits a consistent model of short-length-scale DNA bending.

Unfortunately, precise quantitative tests of short-length-scale DNA bending are still in the future. Vologodskii and co-workers recently made measurements questioning the results of Cloutier and Widom [8]. Their measurements suggest that the J factor agrees with that predicted by the WLC model, at least down to a contour length of 100 bp, but with a persistence length of just 47 nm. Widom and co-workers then repeated their own measurements, however, and still find an anomalously high cyclization rate [46]. Furthermore, Sec. IV A mentioned that Shroff *et al.* [9] also found that linear elasticity fails at high curvature. At the moment, it is difficult to reconcile these conflicting experiments. Instead our goal has been to show that existing experiments do not uniquely confirm the WLC; we have examined some of the options for theories compatible with those experiments that appear to be understood.

We have repeatedly emphasized that the SEC is more a proof of principle than a finished theory. It is a generalization of the WLC that is extremely compact to state and can be solved almost fully by analytic methods. It shows that the classic successes of WLC can be reconciled with more recent indications of elastic breakdown. It encodes locality at the mesoscopic length scale $\ell \approx 5$ nm, but assumes that linear elasticity does not hold at that scale. Indeed, we do expect that linear elasticity will break down at length scales corresponding to the curvature radius at which the DNA duplex is

not a minimum of free energy. We would not expect the usual duplex form to be stable when bent into a loop of radius 5 nm.

The SEC's other, less realistic features, such as the neglect of sequence dependence, can readily be addressed, albeit at the cost of explicit solutions. Its bending energy function, however, is not meant to be a literal depiction of DNA mechanics. In principle, the true bending energy function can eventually be deduced from statistical analysis of sufficiently accurate determinations of DNA contours, obtained either in solution via cryo EM, or when adsorbed to surfaces by AFM or EM. Alternatively, the short-length-scale bending energy might be calculated using molecular dynamics simulations. Direct all-atom molecular dynamics computations of the chain statistics for long-contour-length sequences of DNA are prohibitive computationally, but the generalized polymer model described in this paper is based upon the chain statistics of short-contour-length links which may be directly simulated.

B. Future directions

For many biological applications of DNA chain statistics, the twist degrees of freedom are also of great importance. For instance, for DNA looping, moving an operator (the DNA binding sequence) a few base pairs can change the looping probability by an order of magnitude [42]. This dramatic, short-contour-length dependence arises from the necessity of bring the DNA operator into twist registry with the binding site. The twist degree of freedom of DNA has also been described by a fluctuating elastic-rod model, the helical wormlike chain (HWLC) [31]. At long length scales, this modified WLC model has successfully described the twist dependence of DNA. Nevertheless, at high enough strain the HWLC model breaks down. For example, Bryant *et al.* have demonstrated that the restoring torque generated by twisted DNA saturates for high twist densities, implying that the linear elastic model breaks down when the undertwist $|\Delta\omega|$ exceeds 0.01 rad/base pair [47]. The twist density needed to join a misphased DNA loop of under 100 bp exceeds this threshold, and indeed Cloutier and Widom have also shown that the twist-induced modulation of the cyclization J factor is smaller for short sequences than predicted by the HLWC model [37].

Thus, although the bending of DNA for small twist densities may be adequately described by the HWLC model, a generalized model of DNA, including elastic breakdown of both bend and twist stiffnesses, may be necessary to describe the chain statistics of short sequences of looped DNA that are not naturally in twist registry when bound. Such generalized models are in principle a straightforward extension of the theory presented in this paper and exact results for the HWLC model recently derived by Spakowitz [48].

VI. CONCLUSION

In this paper, we have explored a class of generalized semiflexible polymer models in which the bending energy density is an arbitrary function of curvature. To analyze the

chain statistics of these models, we developed a formalism that is analogous to the techniques used for describing the WLC model. We demonstrated that the statistics of these general models coincide with those of the linear-elastic (WLC) model at long contour length, as predicted by the renormalization group. At short length scales, we show that the predictions of these models can be dramatically different from those of the WLC model. We computed expressions for the transformed spatial and tangent-spatial distribution functions with a method analogous to that recently exploited to find exact results for the WLC model. These generalized models provide an explicit example of a nonrenormalizable model which is nearly exactly solvable. We exploited these general theoretical results to compute several important experimental observables: force-extension, the structure factor, and the cyclization J factor. We explicitly performed these computations for a toy model of DNA bending, the subelastic chain model. The predictions of this model are essentially indistinguishable from those of the WLC model for force-extension, solution scattering, and long-contour-length cyclization measurements, despite dramatic differences between the bending energies of the two models on short length scales. For short-contour-length cyclization experiments, general models generically predict large deviations from WLC behavior. We expect these generalized models to be widely applicable for describing the high-curvature statistics of other semiflexible polymers.

ACKNOWLEDGMENTS

We thank N. R. Dan, Cees Dekker, M. Inamdar, Igor Kulić, Richard Lavery, J. Maddocks, John Marko, Fernando Moreno-Herrero, Rob Phillips, Ashok Prasad, Prashant Purohit, J. M. Schurr, A. Spakowitz, Thijn van der Heijden R. James, Z.-G. Wang, Jonathan Widom, and Yongli Zhang, for helpful discussions and correspondence. P.A.W. acknowledges financial support from the NSF and the Keck Foundation and from NSF Grant No. CMS-0301657, and the NSF-funded Center for Integrative Multiscale Modeling and Simulation. P.N. acknowledges NSF Grant No. DMR04-04674 and the NSF-funded NSEC on Molecular Function at the Nano/Bio Interface Grant No. DMR04-25780.

APPENDIX A: EXPLICIT EXPRESSIONS FOR g_l

It is straightforward to determine the g_l eigenvalues of any propagator using the orthonormal eigenbasis of the angular momentum representation. In two dimensions, the g_l are

$$g_l = \int_{-\pi}^{\pi} d\theta g(\vec{t}(\theta); \vec{e}_z) \exp(il\theta), \quad (\text{A1})$$

where θ is defined as the polar angle away from the z axis: $\vec{t}(0) = \vec{e}_z$. In three dimensions, the g_l are

$$g_l = \int d^2\vec{t} g(\vec{t}(\theta); \vec{e}_z) P_l(\vec{t} \cdot \vec{e}_z), \quad (\text{A2})$$

where the P_l are the Legendre polynomials and $\cos \theta = \vec{t} \cdot \vec{e}_z$.

APPENDIX B: STIFF POLYMER LIMIT

In this section, we show that a narrowly distributed fundamental tangent distribution function generically implies WLC statistics at long contour length. In dimension D , this calculation, though straightforward, requires some book-keeping, but these technical details are not important for the interpretation of the result.

We begin the derivation with the definition of the l th moment of the tangent distribution function expressed in terms of the propagator Eq. (4):

$$g_l = \langle l\mathbf{m} | \mathcal{G} | l\mathbf{m} \rangle, \quad (\text{B1})$$

where rigid-body-rotational invariance implies that g_l is independent of \mathbf{m} . We insert two complete sets of states into the tangent representation,

$$g_l = \int d\vec{t} d\vec{t}' \langle l\mathbf{m} | \vec{t} \rangle \langle \vec{t} | \mathcal{G} | \vec{t}' \rangle \langle \vec{t}' | l\mathbf{m} \rangle. \quad (\text{B2})$$

We can now replace the matrix element of the propagator with the fundamental tangent distribution function $g(\vec{t}; \vec{t}')$ [Eq. (4)]. Remember that this function depends only on the relative deflection angle of the tangents. We therefore replace the integral over the second tangent with an integral over rotation matrices \mathcal{R} , and make the substitution $\vec{t}' \equiv \mathcal{R}\vec{t}$:

$$g_l = \int d\vec{t} d\mathcal{R} \left| \frac{d\mathcal{R}}{d\mathcal{R}'} \right| \langle l\mathbf{m} | \vec{t} \rangle g(\vec{t}; \mathcal{R}\vec{t}) \langle \vec{t} | \mathcal{D}_{\mathcal{R}}^\dagger | l\mathbf{m} \rangle, \quad (\text{B3})$$

where we represent the change in measure symbolically and we have introduced the rotation operator [28]

$$\mathcal{D}_{\mathcal{R}} | \vec{t} \rangle \equiv | \mathcal{R}\vec{t} \rangle. \quad (\text{B4})$$

Our interest is in the case where the tangent distribution function is narrowly distributed. We shall therefore expand the rotation operator \mathcal{D} with respect to the rotation angles, which we shall assume are small. The rotation operator can be expanded in terms of these angles and the rotation generators [28]

$$\mathcal{D}_{\mathcal{R}} = \exp(-i\theta_{ij}\mathcal{L}_{ij}) \quad (\text{B5})$$

$$= 1 - i\theta_{ij}\mathcal{L}_{ij} - \frac{1}{2}\theta_{ij}\mathcal{L}_{ij}\theta_{mn}\mathcal{L}_{mn} + \dots, \quad (\text{B6})$$

where the $\theta_{ij} = -\theta_{ji}$ are the components of the rotation angle which multiply the generators of rotations in the ij plane.

To evaluate the integral over the rotation matrices, we must now choose a set of θ 's that give a single cover of the tangent space. Since $g(\vec{t}; \mathcal{R}\vec{t})$ is independent of \vec{t} , it is convenient to choose a coordinate system in which \vec{t} is in the direction of the D axis. (We shall return to the unrotated frame before performing the integral over \vec{t} .) In this new coordinate system, it is convenient to use the cover generated by the coordinates $\{\theta_{Dj}\}_{1, \dots, D-1}$, while setting all other θ 's to zero.

We denote the average taken with respect to the distribution function by $\langle \cdot \rangle$. Due to rigid-body-rotational invariance around the D axis,

$$\langle \theta_{Dj} \rangle = 0, \quad (\text{B7})$$

$$\langle \theta_{iD} \theta_{nD} \rangle = \langle \theta^2 \rangle \delta_{in} / (D-1), \quad (\text{B8})$$

where

$$\theta^2 \equiv \sum_{i=1}^{D-1} \theta_{iD}^2 \quad (\text{B9})$$

is the total deflection angle.

The nonzero matrix elements can be put in a coordinate-invariant form:

$$\langle l\mathbf{m} | \vec{e}_D \rangle \langle \vec{e}_D | \mathcal{L}_{D_i} \mathcal{L}_{D_i} | l\mathbf{m} \rangle = \langle l\mathbf{m} | \vec{e}_D \rangle \langle \vec{e}_D | \mathcal{L}^2 | l\mathbf{m} \rangle \quad (\text{B10})$$

since the added terms in the Casimir operator \mathcal{L}^2 are zero on $|\vec{e}_D\rangle$. We can now go back to the unrotated coordinate system by setting $\vec{e}_D = \vec{i}$.

After integrating over the complete set of tangent vectors, the resulting moment is

$$g_l = 1 - \frac{1}{2}(D-1)^{-1} \langle \theta^2 \rangle \langle l\mathbf{m} | \mathcal{L}^2 | l\mathbf{m} \rangle + \mathcal{O}(\mathcal{L}^4 \langle \theta^4 \rangle). \quad (\text{B11})$$

Since this expression is correct only to $\mathcal{O}(\theta^4)$, it is convenient to replace $\frac{1}{2}\theta^2$ with $1 - \cos \theta$. We can now use the definition of the persistence length given in Eq. (9) to eliminate the dependence on $\langle \cos \theta \rangle$:

$$g_l = 1 - \frac{\ell}{\xi(D-1)} \langle l\mathbf{m} | \mathcal{L}^2 | l\mathbf{m} \rangle + \mathcal{O}(\mathcal{L}^4 \ell^2 / \xi^2). \quad (\text{B12})$$

Finally, we reconstruct the propagator from its moments,

$$\mathcal{G} = \sum_{l,\mathbf{m}} g_l |l\mathbf{m}\rangle \langle l\mathbf{m}| = 1 - \frac{\ell}{\xi(D-1)} \mathcal{L}^2 + \mathcal{O}(\mathcal{L}^4 \ell^2 / \xi^2), \quad (\text{B13})$$

which completes the derivation. This result is discussed in Sec. II E.

APPENDIX C: THE TRANSFORMED SPATIAL PROPAGATOR

Section III introduced the operator $\mathcal{G}(\vec{X}; L)$ [Eq. (14)]. We shall call this operator the spatial propagator since it obeys the composition property of Green's functions:

$$\mathcal{G}(\vec{X}; L + L') = \mathcal{G}(\vec{X}; L) \otimes \mathcal{G}(\vec{X}; L'), \quad (\text{C1})$$

where \otimes denotes spatial convolution.

We now derive exact, analytic expressions for the Fourier-Laplace transform of the spatial propagator in the continuum theory in terms of the transformed matrix elements [Eq. (15)]. We adopt the Fourier transform convention

$$G(\vec{k}; L) \equiv \mathcal{F}_{X \rightarrow k} G(\vec{X}; L) \equiv \int d^3 X G(\vec{X}; L) \exp(-i\vec{k} \cdot \vec{X}) \quad (\text{C2})$$

and the Laplace transform convention

$$\tilde{G}(\vec{k}; p) \equiv \mathcal{L}_{L \rightarrow p} G(\vec{k}; L) \equiv \int_0^\infty dL G(\vec{k}; L) \exp(-pL). \quad (\text{C3})$$

The derivation of the transformed matrix elements exploits the same techniques used recently by Spakowitz and Wang [32,36,48] to derive exact results for the WLC model. The extension of these results to the generalized theories considered here is straightforward.

To derive closed form expressions for the spatial propagator, we Fourier transform the spatial propagator over the relative displacement \vec{X} . In particular, we consider the Fourier transform of Eq. (C1) since in Fourier space, the spatial convolutions are simply products:

$$\tilde{G}(\vec{k}; L + L') = \tilde{G}(\vec{k}; L) \tilde{G}(\vec{k}; L'). \quad (\text{C4})$$

We choose a coordinate system where \vec{k} is in the z direction.

We now wish to use this composition property of the spatial propagator to write a differential equation for \mathcal{G} . We therefore consider \mathcal{G} for a differential arc length dL and then expand the Fourier transform of Eq. (13) for arclength dL :

$$\tilde{\mathcal{G}}(\vec{k}; dL) = \mathcal{I} - \mathcal{A} dL, \quad (\text{C5})$$

where \mathcal{I} is the identity operator and $\mathcal{A} \equiv \mathcal{H} + ik \cos \theta$ where θ takes its canonical meaning in spherical coordinates, $\cos \theta = \vec{i} \cdot \hat{z}$. Substituting this expression into Eq. (C4), we can write a differential equation for $\tilde{\mathcal{G}}$:

$$\frac{d}{dL} \tilde{\mathcal{G}}(k; L) = -\mathcal{A} \tilde{\mathcal{G}}(k; L). \quad (\text{C6})$$

It is now convenient to make a Laplace transform from arc-length L to its conjugate variable p . After solving for the propagation operator, we have an operator equation for the Laplace-Fourier transform of the spatial propagator:

$$\tilde{\mathcal{G}}(k; p) = \{p\mathcal{I} + \mathcal{A}(k)\}^{-1} = \{p\mathcal{I} + \mathcal{H} + ik \cos \theta\}^{-1}, \quad (\text{C7})$$

but this expression is not explicit since it is written in terms of the inverse of an infinite-dimensional operator.

We can express $\cos \theta$ in the angular momentum basis. It is most convenient to define a set of ladder operators

$$\cos \theta = a_+ + a_-, \quad (\text{C8})$$

where

$$a_+ \equiv \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{l+1,l,m} |l+1\ m\rangle \langle l\ m|, \quad (\text{C9})$$

$$a_- \equiv \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{l,l+1,m} |l\ m\rangle \langle l+1\ m|. \quad (\text{C10})$$

The $A_{l,l+1,m}$ are

$$A_{l,l+1,m} = A_{l+1,l,m} = \sqrt{\frac{(l-m+1)(l+m+1)}{(2l+1)(2L+3)}}. \quad (\text{C11})$$

The ladder operators have the property that they increase (decrease) the total momentum quantum number of a state by plus (minus) 1.

Next, we obtain explicit expressions for the matrix elements of the transformed spatial propagator. The Hamiltonian is diagonal in the angular representation, so it is convenient to factor the spatial propagator [Eq. (C7)] into diagonal and nondiagonal factors:

$$\tilde{\mathcal{G}}(k;p) = [\mathcal{I} + \{p\mathcal{I} + \mathcal{H}\}^{-1} ik(a_+ + a_-)]^{-1} \{p\mathcal{I} + \mathcal{H}\}^{-1}, \quad (\text{C12})$$

and expand it in a power series

$$\tilde{\mathcal{G}}(k;p) = \sum_{n=0}^{\infty} [-ik\{p\mathcal{I} + \mathcal{H}\}^{-1}(a_+ + a_-)]^n \{p\mathcal{I} + \mathcal{H}\}^{-1}. \quad (\text{C13})$$

As a first step, we will compute a diagonal matrix element

$$\tilde{G}_{lm|lm} = \langle l m | \tilde{\mathcal{G}}(k;p) | l m \rangle. \quad (\text{C14})$$

Computing these matrix elements is achieved by grouping the infinite set of terms in Eq. (C13) into subsets which can be summed explicitly [36].

We introduce $\tilde{G}_{l'm|l'm}^+$, which is the matrix element of a subset of the terms in Eq. (C13), in which there are only transitions to states with total momentum $l=l'$ or greater [36]. This matrix element can be defined recursively since only transitions to adjacent states are possible. The matrix element is the sum over n of the matrix elements with n transitions to and from the $l \geq l' + 1$ states, which can be written in terms of $\tilde{G}_{l'+1,m,l'+1,m}^+$. The terms of this matrix element, a geometric series, give [32]

$$\tilde{G}_{l'm|l'm}^+ = \frac{1}{p + h_{l'n=0}} \sum_{n=0}^{\infty} \left(\frac{-k^2 A_{l,l+1,m}^2 \tilde{G}_{l+1,m,l+1,m}^+}{p + h_l} \right)^n \quad (\text{C15})$$

$$= (p + h_l + k^2 A_{l,l+1,m}^2 \tilde{G}_{l+1,m,l+1,m}^+)^{-1}. \quad (\text{C16})$$

This sum is shown schematically in Fig. 9.

Similarly, we define $\tilde{G}_{l'm|l'm}^-$ as the matrix element of the propagation operator that allows transitions to states with total momentum $l=l'$ or less:

$$\tilde{G}_{l'm|l'm}^- = (p + h_l + k^2 A_{l,l-1,m}^2 \tilde{G}_{l-1,m,l-1,m}^-)^{-1}. \quad (\text{C17})$$

In terms of G^\pm we can now define the matrix element without transition restrictions by grouping the transitions into sets that do not cross $l=l'$. These sets can be written in terms of the matrix elements of \mathcal{G}^\pm and then summed in a geometric series [32]:

$$\begin{aligned} l \text{---} &= p + h_l \\ l \text{---} \boxed{+} &= ik A_{l,l+1,m} \\ l \text{---} \vdots &= \tilde{G}_{lmlm}^+ \\ l m \text{---} \vdots &= l m \text{---} + \begin{array}{c} \vdots \\ l+1 m \text{---} \end{array} + \begin{array}{c} \vdots \\ l+1 m \text{---} \end{array} \begin{array}{c} \vdots \\ l+1 m \text{---} \end{array} + \dots \end{aligned}$$

FIG. 9. Diagrammatic rules for the propagator: diagrams and their algebraic representations. Connected diagrams represent the products of the corresponding algebraic representations. The matrix element of the spatial propagator $\tilde{G}_{l'm|l'm}$ is the sum of all diagrams which begin at state lm and end at state $l'm$ with an arbitrary number of intermediate transitions. (a) Horizontal lines represent propagation. Vertical lines represent transitions induced by the wave number. $\tilde{G}_{l'm|l'm}^+$ is the matrix element of the spatial propagator where transitions to states with total angular momentum $l-1$ or smaller are forbidden. This matrix element is represented by the line with ellipses, representing all transitions to states with higher l . (b) $G_{l'm|l'm}^+$ can be defined recursively in terms of $\tilde{G}_{l+1,m,l+1,m}^+$. The definition of $\tilde{G}_{l'm|l'm}^-$ is analogous, but it is the sum of all diagrams with transitions to states with total angular quantum number l and smaller.

$$\begin{aligned} \tilde{G}_{l'm|l'm} &= (p + h_l + k^2 A_{l,l+1,m}^2 \tilde{G}_{l+1,m,l+1,m}^+ \\ &\quad + k^2 A_{l,l-1,m}^2 \tilde{G}_{l-1,m,l-1,m}^-)^{-1}. \end{aligned} \quad (\text{C18})$$

The diagonal matrix element computed above is sufficient for describing many observables of phenomenological interest. Note that the only difference between this expression and the WLC expression [32] is that the eigenvalues of the Hamiltonian operator have changed.

For some applications we will want completely general matrix elements $\tilde{G}_{l'm|l'm}$. We can again define these general matrix elements in terms of the recursive definitions of G^\pm . Again, the trick to summing the terms is grouping them. In this general case, there are many equivalent ways of achieving this grouping. See Fig. 10 for an explanation of the set grouping. The matrix element can be written [32]

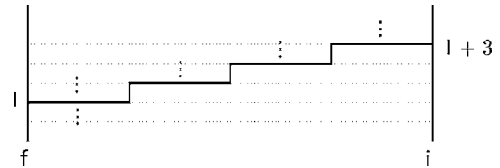


FIG. 10. General matrix elements. A diagram of the sum for the matrix element $\tilde{G}_{l'm|l'm} = \tilde{G}_{l+n|l+n}$. To compute the matrix element, we group the terms by the location of the first steps from $l+n$ to $l+n-1$ and from $l+n-1$ to $l+n-2$, etc. In the diagram, these steps are represented by the vertical lines. We use the G^+ operator to sum over all possible diagrams with upward transitions between these steps. These upward transitions are represented by the ellipses. We multiply by the transition matrix element for each of the vertical lines. After we reach l for the first time we allow all transitions up or down. This enumeration counts each contributing diagram once but this recipe is not unique.

$$\tilde{G}_{l+n,m,l,m'} = \tilde{G}_{l,m,l+n,m'} = \delta_{m-m'} \tilde{G}_{lmlm} \prod_{q=1}^n -ikA_{l+q-1,l+q,m} \tilde{G}_{l+q,m,l+q,m}^+ \quad (\text{C19})$$

We have now explicitly solved for the spatial propagator having written expressions for all the matrix elements.

APPENDIX D: THE COMPUTATION OF SPATIAL DISTRIBUTIONS

The previous section discussed expressions for the Fourier-Laplace transformed spatial and tangent-spatial distribution functions. We now invert the transforms numerically to compute the real-space distribution functions.

1. Force extension and the structure factor

The computations of force-extension and the structure factor require only a single numerical inverse Laplace transform. We cut off the continued fraction at $l=10$ and then used the `InverseLaplaceTransform` function in MATHEMATICA.

2. The spatial distribution and the J factor

For computations of the spatial distribution function and the J factor, we exploited two different numerical techniques: numerical transform inversion and Monte Carlo simulation. For contour lengths of a persistence length and above, it is convenient to directly invert the transforms numerically by truncating the continued fraction in the transformed propagator [Eq. (C18)]. Typically we used $l=15$ as the cutoff although in some cases higher l values were used for short contour lengths.

In the inverse transform technique, both numerical Laplace and Fourier transform inversions must be computed. We have used two different implementations for these computations. (i) In MATHEMATICA, we used the `InverseLaplaceTransform` function. We then integrated numerically (using an explicit sum) to invert the Fourier transform. We found that the built-in numerical integration in MATHEMATICA was too slow for practical use. (ii) In MATLAB, we used a code which explicitly computed the Laplace transform by computing the sum of the residues of the inverse Laplace transform contour integral. The Fourier transform inversion was again performed by numerical integration using an explicit sum. The MATLAB code was based on one shared with us by Andy Spakowitz.

For contour lengths of order a persistence length and shorter, inverting the transformed expressions is impractical. The continued fraction is an expansion around weak end-tangent correlation. For contour lengths shorter than a persistence length, a larger- l cutoff is required, significantly slowing the numerical inversions. In addition, the numerical integration over the wave number becomes impractical since the numerical integrations must be extended to a very large cutoff momentum. These convergence issues are not unique to the continued fraction approach. For example, the transfer matrix approach is plagued by similar shortcomings, requiring difficult numerical work at short contour length [14].

We therefore used a much simpler, although less elegant, solution in the form of direct Monte Carlo integrations. Monte Carlo integration in the short-contour-length regime (i) is numerically more efficient than direct inversion, (ii) requires very minimal implementation, and (iii) serves as a useful check of our theoretical results. The agreement between these two methods is excellent in the overlap of their domains. The theoretical curve for the cyclization J factor (Fig. 8) contain both inversion and Monte Carlo computations.

-
- [1] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson, *Molecular Biology of the Cell*, 3rd ed. (Garland Publishing, New York, 1994).
- [2] J. Widom, *Q. Rev. Biophys.* **34**, 269 (2001).
- [3] K. Rippe, P. R. von Hippel, and J. Langowski, *Trends Biochem. Sci.* **20**, 500 (1995).
- [4] D. Shore, J. Langowski, and R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **170**, 4833 (1981).
- [5] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, *Curr. Opin. Struct. Biol.* **10**, 279 (2000).
- [6] P. Nelson, *Biological Physics: Energy, Information, Life* (W. H. Freeman and Co., New York, 2004).
- [7] T. E. Cloutier and J. Widom, *Mol. Cell* **14**, 355 (2004).
- [8] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaja, and A. Vologodskii, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5397 (2005).
- [9] H. Shroff, B. M. Reinhard, M. Siu, H. Agarwal, A. Spakowitz, and J. Liphardt, *Nano Lett.* **5**, 1509 (2005).
- [10] J. M. Schurr, J. J. Delrow, B. S. Fujimoto, and A. S. Benight, *Biopolymers* **44**, 283 (1997).
- [11] M. Vologodskaja and A. Vologodskii, *J. Mol. Biol.* **317**, 205 (2002).
- [12] C. Rivetti, M. Guthold, and C. Bustamante, *J. Mol. Biol.* **264**, 919 (1996).
- [13] P. A. Wiggins, P. C. Nelson, and R. Phillips, *Phys. Rev. E* **71**, 021909 (2005).
- [14] J. Yan and J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
- [15] J. Yan, R. Kawamura, and J. F. Marko, *Phys. Rev. E* **71**, 061905 (2005).
- [16] Y. O. Popov and A. V. Tkachenko, *Phys. Rev. E* **71**, 051905 (2005).
- [17] P. Ranjith, P. B. S. Kumar, and G. I. Menon, *Phys. Rev. Lett.* **94**, 138102 (2005).
- [18] C. A. Sucato, D. P. Rangel, D. Aspleaf, B. S. Fujimoto, and J. M. Schurr, *Biophys. J.* **86**, 3079 (2004).
- [19] W. Han, M. Dlakic, Y. J. Zhu, S. M. Lindsay, and R. E. Harrington, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10565 (1997).
- [20] W. Han, M. Dlakic, Y. J. Zhu, S. M. Lindsay, and R. E. Harrington, *Nature (London)* **386**, 563 (1997).
- [21] M. Hogan, T. Rooney, and R. H. Austin, *Nature (London)* **328**, 554 (1987).

- [22] T. J. Richmond and C. A. Davey, *Nature (London)* **423**, 145 (2003).
- [23] A. Hochschild and M. Ptashne, *Cell* **44**, 681 (1986).
- [24] F. Lankas, R. Lavery, and J. H. Maddocks (private communication).
- [25] R. L. Fosdick and R. D. James, *J. Elast.* **11**, 165 (1981).
- [26] C. Bouchiat, M. D. Wang, J.-F. Allemand, T. Strick, S. M. Block, and V. Croquette, *Biophys. J.* **76**, 409 (1999).
- [27] J. van Noort, T. van der Heijden, M. de Jager, C. Wyman, R. Kanaar, and C. Dekker, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7581 (2003).
- [28] J. J. Sakurai, *Modern Quantum Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1994).
- [29] H. Kleinert, *Path Integrals*, 3rd ed. (World Scientific, Singapore, 2004).
- [30] B. Hamprecht and H. Kleinert, *Phys. Rev. E* **71**, 031803 (2005).
- [31] H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions* (Springer, Berlin, 1997).
- [32] A. J. Spakowitz and Z.-G. Wang, *Phys. Rev. E* **72**, 041802 (2005).
- [33] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith, *Science* **265**, 1599 (1994).
- [34] T. R. Strick, V. Croquette, and D. Bensimon, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10579 (1998).
- [35] S. Stepanow and G. M. Schutz, *Europhys. Lett.* **60**, 546 (2002).
- [36] A. J. Spakowitz and Z.-G. Wang, *Macromolecules* **37**, 5814 (2004).
- [37] T. E. Cloutier and J. Widom, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3634 (2005).
- [38] H. Jacobson and W. H. Stockmayer, *J. Chem. Phys.* **18**, 1600 (1950).
- [39] D. Shore and R. L. Baldwin, *J. Mol. Biol.* **170**, 957 (1983).
- [40] J. Shimada and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- [41] P. J. Hagerman, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).
- [42] J. Muller, S. Oehler, and B. Muller-Hill, *J. Mol. Biol.* **257**, 21 (1996).
- [43] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 125 (2005).
- [44] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [45] L. Finzi and J. Gelles, *Science* **68**, 378 (1995).
- [46] J. Widom (private communication).
- [47] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, *Nature (London)* **424**, 338 (2003).
- [48] A. J. Spakowitz, (in press).