

Self-avoiding wormlike chain model for double-stranded-DNA loop formationYaroslav Pollak,^{*} Sarah Goldberg, and Roe Amit,[†]*Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa, Israel 32000*

(Received 24 August 2014; published 17 November 2014)

We compute the effects of excluded volume on the probability for double-stranded DNA to form a loop. We utilize a Monte Carlo algorithm for generation of large ensembles of self-avoiding wormlike chains, which are used to compute the J factor for varying length scales. In the entropic regime, we confirm the scaling-theory prediction of a power-law drop off of -1.92 , which is significantly stronger than the -1.5 power law predicted by the non-self-avoiding wormlike chain model. In the elastic regime, we find that the angle-independent end-to-end chain distribution is highly anisotropic. This anisotropy, combined with the excluded volume constraints, leads to an increase in the J factor of the self-avoiding wormlike chain by about half an order of magnitude relative to its non-self-avoiding counterpart. This increase could partially explain the anomalous results of recent cyclization experiments, in which short dsDNA molecules were found to have an increased propensity to form a loop.

DOI: [10.1103/PhysRevE.90.052602](https://doi.org/10.1103/PhysRevE.90.052602)

PACS number(s): 36.20.Ey, 87.10.Rt, 87.14.gk, 87.16.af

I. INTRODUCTION

Cyclization or looping of polymers has functioned as an important experimental paradigm for probing nanometric elastic characteristics of polymers for more than half a century [1–4]. In particular, experimental studies of polymer cyclization, in conjunction with modern single-molecule approaches [5,6], have been used extensively in recent years to study double-stranded DNA (dsDNA). While these studies have led to important advances in our understanding of the properties of dsDNA polymer physics, a detailed comparison of the data that has emerged from the different experimental approaches has generated surprising and unexpected puzzles such as dsDNA hyperbendability at short polymer lengths [4,6,7]. Since many important biological regulatory processes that occur at the level of the genome are strongly related to DNA looping or cyclization, it is crucial to develop a detailed theoretical understanding of dsDNA cyclization that can explain the seemingly contrasting force extension [5] and cyclization [4,6,7] experimental observations.

As a result of these experimental works, a consensus picture of dsDNA as a semiflexible polymer has emerged. At lengths that are smaller than the Kuhn length [2] DNA behaves like a rigid rod. At intermediate lengths it behaves like a Gaussian chain [8]. At large lengths, dsDNA becomes a swollen chain due to excluded volume effects that emerge from the inability of DNA to cross itself [2,8]. A recent set of theoretical studies [9,10] showed that the DNA length, at which the transition from the ideal to the swollen chain occurs is highly dependent on the effective aspect ratio of DNA, which we define as the ratio of the effective width or cross section of DNA w to the Kuhn length b . As $\frac{w}{b} \rightarrow 0$, the DNA transition length becomes larger, which implies that the polymer behaves like an ideal chain for an ever-increasing length. Alternatively, as $\frac{w}{b} \rightarrow 1$, the DNA transition length shrinks, and the DNA behaves increasingly like a swollen chain for all lengths. Since DNA is anisotropic with an effective aspect ratio that is estimated

to be in the range of 0.03–0.09 [11], the power-law exponent for the end-to-end distance of DNA was predicted [12] to not deviate substantially from the Gaussian chain prediction for lengths as high as ~ 100 kb. This prediction was recently confirmed experimentally, showing a radius of gyration power law of 0.52 ± 0.02 [10], that is, very close to the Gaussian chain prediction of 0.5.

The process of DNA cyclization or looping has not been modeled to date by algorithms that take into consideration the effective aspect ratio of the DNA. Historically, this process is quantified by the Jacobson-Stockmayer factor [1], which was defined as [3]

$$J = \frac{K_C}{K_D}, \quad (1)$$

namely, the ratio of the equilibrium constant of cyclization K_C to that of bimolecular association K_D , assuming the measurement is carried out on a dilute solution of dsDNA in standard saline conditions [3]. This definition led Flory, Shimada, and Yamakawa [13,14] to re-derive the J factor via equilibrium thermodynamic considerations as the probability for the two ends of the dsDNA to bond, normalized by the infinitesimal bonding volume and angular tolerance in order to avoid any measurable ambiguities that might depend on arbitrary choices of the bonding volume.

Experimentally, measuring the J factor for DNA has proven to be quite challenging, but improved technology has yielded a set of increasingly more precise *in vivo* [15] and *in vitro* measurements [4,7] of this quantity. As a result, an interesting if not controversial picture had emerged. While the behavior for longer lengths ($L > b$) has matched well with the predictions of the wormlike chain (WLC) model, for shorter chains where $L < b$ a deviation from the WLC model has been consistently observed [4,6,7,15]. In this regime, the dsDNA is expected to behave like a rigid rod whose propensity to form a loop depends strongly on the elastic or bending energy. Therefore, as the chain becomes shorter, the probability to form a loop is expected to decrease exponentially. However, the experimental data showed that this prediction is not observed to the extent predicted by the WLC model, and instead an increasing propensity for bending as compared with model predictions was detected as the length of the

^{*}Also at Russell Berrie Nanotechnology Institute, Technion.[†]roeamit@technion.ac.il; also at Russell Berrie Nanotechnology Institute, Technion.

DNA chain was decreased below the Kuhn length. Several theoretical hypotheses were raised to explain this discrepancy, including localized DNA kinking or formation of melting bubbles [16,17], and possible subharmonic contributions to the WLC which render the DNA more flexible at short lengths [18]. However, these modifications have not been proven conclusively to be the underlying cause for the experimental results, and the problem remains open.

In this work we study cyclization of DNA with finite width. Section II introduces the basic theory that underlies the definition of the self-avoiding J -factor integral. We utilize an advanced Monte Carlo algorithm, described in Sec. III, that is based on the weighted Rosenbluth and Rosenbluth method, similar to the algorithm used by Tree *et al.* [12], to generate ensembles of anisotropic DNA chains up to 5000 bp in length. Our algorithm is able to reproduce the theoretically predicted behavior of DNA at long lengths. In Sec. IV we use our algorithm to compute for the first time the J factor for self-avoiding polymers. In particular, for long chain lengths the numerically computed J -factor power-law exponent converges on the scaling law prediction [8,19] of -1.92 regardless of the width of the chain. For shorter lengths we find that the chosen definition of the end-to-end bonding strongly affects the J -factor behavior. We demonstrate that this function can be systematically changed by choosing various end-to-end bonding conditions, and leads to an apparent increased “bendability” effect in this regime as compared with the WLC model.

II. THEORY

A. The self-avoiding wormlike chain (SAWLC) model

DNA is typically modeled as a discrete semiflexible chain made of individual and irreducible links of length l , such that the deviation of one link from its neighboring link depends strongly on some elastic energy. This class of polymer models is based on the original work of Kratky and Porod [20] and is referred to as the class of wormlike chain (WLC) models. However, except for a few notable exceptions [9,12], the WLC models do not take into account energetic and entropic effects that emerge from the cross section or “thickness” of the DNA double helix.

In the following, we describe each chain by the locations of its elements, and a local coordinate system defined by three orthonormal vectors $\hat{u}, \hat{v}, \hat{t}$ at each element. The vector \hat{t} points along the direction of the chain. For the continuous WLC these vectors are defined continuously along the chain contour. For the discrete WLC, the joint locations \mathbf{r}_i and the local coordinate systems of the links fully define the chain. We number the links of a chain in the range $1..N$ for a chain of N links, and the endpoints of the links (chain joints) in the range $0..N$, where joint 0 is the beginning terminus of the chain [see Fig. 1(a)]. When mapping wormlike chains to actual dsDNA, the chain links correspond to DNA base pairs and chain joints correspond to midpoints between DNA base pairs.

The conventional bending energy for the WLC models is the elastic energy associated with bending link $i \in \{2, \dots, N\}$ relative to link $i - 1$ with angles θ_i, ϕ_i (zenith and azimuthal angles in local spherical coordinates of link $i - 1$), which can

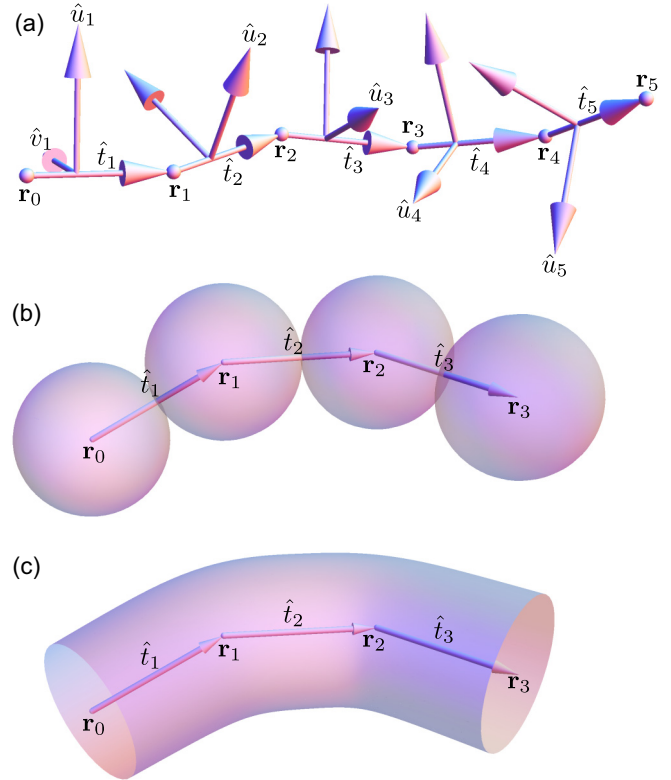


FIG. 1. (Color online) WLC and SAWLC chains. A discrete chain is described by the locations of its joints \mathbf{r}_i , and a local coordinate system defined by three orthonormal vectors $\hat{u}_i, \hat{v}_i, \hat{t}_i$ at each link. The vector \hat{t}_i points along the direction of the i th link. (a) A WLC chain with five links and link length $l = 1$. (b) A SAWLC chain with three links and $l = w = 1$. \hat{u}_i and \hat{v}_i are not shown. (c) Representation of an ideally modeled chain with finite width.

be written as

$$\beta E^{el}(\theta_i, \phi_i) = \frac{a}{2} |\hat{t}_i - \hat{t}_{i-1}|^2 = a(1 - \cos \theta_i), \quad (2)$$

where a corresponds to the bending rigidity of the DNA chain (assuming azimuthal symmetry), and $\beta = (k_B T)^{-1}$. It is important to note that the angles θ_i, ϕ_i are given in the *local coordinate system* of the $(i - 1)$ th link. For a specific configuration of the chain, we introduce the notation $\{\theta_n, \phi_n\}$ to denote the set of all the links' angles of the chain, from link 1 to link n .

To account for the finite thickness of the DNA we introduce a second energy contribution. We engulf each joint by a “hard-wall” spherical shell of diameter w . This allows us to model the final contribution to the elastic energy as a set of hard-wall potentials. For the simple case in which the chain link length is larger than the link diameter ($l \geq w$), and therefore no two neighboring hard-wall spheres overlap, the hard-wall potential energy for the i th chain link can be defined as

$$E_i^{hw}(\{\theta_i, \phi_i\}) = \begin{cases} \infty & \text{joint } i \text{ overlaps with one or more joints } 0..(i-1) \\ 0 & \text{otherwise} \end{cases}. \quad (3)$$

This allows us to write an expression for the total elastic energy associated with the chain of spheres as follows:

$$E(\{\theta_N, \phi_N\}) = \sum_{i=1}^N E^{el}(\theta_i, \phi_i) + \sum_{i=1}^N E_i^{hw}(\{\theta_i, \phi_i\}). \quad (4)$$

We emphasize that this approach [Fig. 1(b)] is an approximation for the more realistic uniformly thick chain [Fig. 1(c)]. We choose it for our model due to its simplicity and adequate approximation of the chain excluded volume. However, it may result in incorrect representation of chain excluded volume near the chain termini. In particular, if we compare Fig. 1(c) to Fig. 1(b), we note that the minimally possible end-to-end distance of the chain is 0 in the former and w in the latter representations, respectively.

The configurational partition function for the model DNA chain consisting of N links is defined as

$$Z_N = \int_{-1}^1 d \cos \theta_1 \int_0^{2\pi} d\phi_1 \cdots \times \int_{-1}^1 d \cos \theta_N \int_0^{2\pi} d\phi_N \exp[-\beta E(\{\theta_N, \phi_N\})], \quad (5)$$

where $\beta = \frac{1}{k_B T}$. Substituting E_i^{hw} from Eq. (3) and opening the sums yields

$$Z_N = \int_{-1}^1 d \cos \theta_1 \int_0^{2\pi} d\phi_1 \exp[-\beta E^{el}(\theta_1, \phi_1)] \times \Theta_1^{hw}(\{\theta_1, \phi_1\}) \cdots \int_{-1}^1 d \cos \theta_N \int_0^{2\pi} d\phi_N \times \exp[-\beta E^{el}(\theta_N, \phi_N)] \Theta_N^{hw}(\{\theta_N, \phi_N\}), \quad (6)$$

where

$$\Theta_i^{hw}(\{\theta_i, \phi_i\}) = \begin{cases} 0 & \text{joint } i \text{ overlaps with one or more joints } 0..(i-1) \\ 1 & \text{otherwise} \end{cases}. \quad (7)$$

We term the model described by Eq. (4) the self-avoiding wormlike chain model (SAWLC).

B. Generalizing the J factor for the SAWLC

The process of cyclization [1], defined as the joining of one end of the molecule upon itself, can be quantified by the equilibrium constant for the following process [1,13,21]:

$$M_{2N} \rightleftharpoons M_N + cM_N, \quad (8)$$

where M_N , M_{2N} , and cM_N correspond to the monomeric, dimeric (defined as the end-to-end joining of two separate molecules), and cyclized polymer with N links, respectively. This process can be equivalently expressed as two intermediate processes:

$$M_N + M_N \rightleftharpoons M_{2N}, \quad (9a)$$

$$M_N \rightleftharpoons cM_N. \quad (9b)$$

The Jacobson-Stockmayer factor (or J factor) is defined as the equilibrium constant for the entire process (8)

[1,21]:

$$J \equiv \frac{[M_N][cM_N]}{[M_{2N}]}, \quad (10)$$

and the equilibrium constants for the intermediate processes are similarly defined as

$$K_D \equiv \frac{[M_{2N}]}{[M_N]^2}, \quad (11)$$

and

$$K_C \equiv \frac{[cM_N]}{[M_N]}, \quad (12)$$

which implies that

$$J \equiv \frac{K_C}{K_D}. \quad (13)$$

Equation (13) is a useful expression for the J factor from an experimental perspective, but less convenient for numerical simulation.

In order to derive a more insightful expression for J , we must first set a geometrical definition for bond formation between the two termini in a chain of finite thickness. We denote the locations of the bonding termini by \mathbf{r}_0 and \mathbf{r}_N , and the directions of their bonds as \hat{t}_1 and \hat{t}_N (Fig. 1), regardless of whether these termini are part of the same chain (i.e., cyclization) or different chains (i.e., dimerization). We define the minimal possible separation between the two termini as d_{\min} . This leads to the following general bond formation conditions for the SAWLC model (Fig. 2).

\mathbf{r}_N is confined to a volume $\delta\mathbf{r}$ around \mathbf{r}_0 ,

which is defined by a thin shell between d_{\min} and $d_{\min} + \varepsilon$ and a solid angle $\delta\omega'$ around $-\hat{t}_1$. (14)

\hat{t}_N is collinear with \hat{t}_1 within the range $\delta\omega$. (15)

This definition is an extension of the corresponding bond-formation condition defined by Flory [2] for the WLC. Note that our definition reduces to the Flory criterion when we

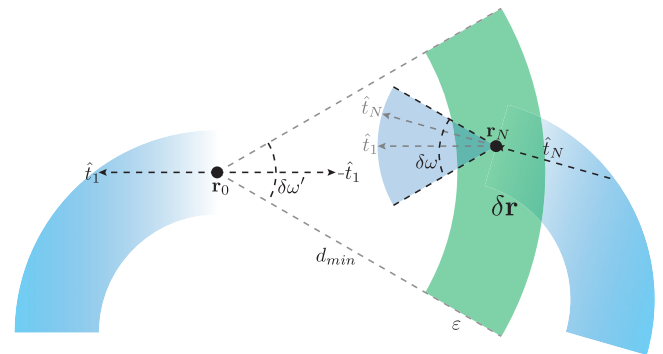


FIG. 2. (Color online) Illustration of bond formation criteria. The termini and their bond directions are denoted by $\mathbf{r}_0, \mathbf{r}_N$ and \hat{t}_1, \hat{t}_N respectively. The minimal possible separation between the termini is denoted by d_{\min} . The cross section of the volume $\delta\mathbf{r}$ in which \mathbf{r}_N must reside is shown as a green (dark gray) wedge. The possible directions for \hat{t}_N are shown as a blue (light gray) cone with its base at \mathbf{r}_N . The rest of the chain is shown schematically with blue (light gray) arcs.

set $d_{\min} = 0$ and $\delta\omega' = 4\pi$. The motivation for extending the Flory condition arises from our imperfect understanding of DNA bond formation. For example, if the transient bonds formed in (9a) and (9b) are identical to the bonds that comprise the bulk of the chain, then the correct representation of the chain would have to be the one presented in Fig. 1(c) with $d_{\min} = 0$. If, however, the termini have a repulsive interaction below some separation d_{\min} , then $d_{\min} > 0$, as shown for the hard-wall interaction in Fig. 1(b).

Using our geometrical definition for bonding, we can now consider the process in (9a). Before dissociation occurs, \mathbf{r}_N is confined to a volume $\delta\mathbf{r}$ around \mathbf{r}_0 . In that volume the orientation of the bond direction \hat{t}_N is further restricted by a solid angle of $\delta\omega$ around \hat{t}_1 . After the dissociation, it can be anywhere in volume v , and its bond can assume any angle within 4π . Denoting the symmetry number of a cyclic chain species σ_a , we obtain the following change in configurational entropy due to process (9a):

$$\Delta S_{(a)} = R \ln \left(\frac{4\pi v / \sigma_a}{\delta\mathbf{r} \delta\omega} \right) = R \ln \left(\frac{V}{N_A \sigma_a \delta\mathbf{r} \frac{\delta\omega}{4\pi}} \right), \quad (16)$$

where we denote the change in entropy as $\Delta S_{(a)}$, and the volume per mole as V resulting in $v = \frac{V}{N_A}$ as the volume per one molecule, where N_A is Avogadro's number. If, for simplicity, we assume that the species are present in standard state concentration of one mole per unit volume ($V = 1$), we obtain

$$\Delta S_{(a)} = R \ln \left(\frac{1}{N_A \sigma_a \delta\mathbf{r} \frac{\delta\omega}{4\pi}} \right). \quad (17)$$

The standard molar Gibbs free-energy change for the dimerization process (9a) can be written as

$$\Delta G_{(a)}^\circ = \Delta H_{(a)}^\circ - RT \ln \left(\frac{1}{N_A \sigma_a \delta\mathbf{r} \frac{\delta\omega}{4\pi}} \right), \quad (18)$$

where $\Delta H_{(a)}^\circ$ is the dissociation heat of the dimer.

In the cyclization process (9b) an intramolecular bond is formed which is equivalent to the one severed in the dimerization (9a). We assume here that $\Delta H_{(b)}^\circ = -\Delta H_{(a)}^\circ$ where $\Delta H_{(b)}^\circ$ is the dissociation heat of the ring. This assumption is valid unless the overall length of the chain is so small as to induce strain [2]. For the computation of the change in configurational entropy due to process (9b), $\Delta S_{(b)}$, we note that the termini of chain M_N must meet within the same ranges of $\delta\mathbf{r}$ and $\delta\omega$ as defined in Eqs. (14) and (15). This implies that the probability for the termini to meet is given by

$$\int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega, \quad (19)$$

where $\mathbf{C}(\mathbf{r}, \omega)$ is the function expressing the distribution of the end-to-end vector $\mathbf{r} \equiv \mathbf{r}_N - \mathbf{r}_0$ within a solid angle ω around \hat{t}_1 ($\frac{\omega}{2\pi} \equiv \hat{t}_1 \cdot \hat{t}_N$), per unit range in \mathbf{r} ($\int_v \int_{4\pi} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega = 1$). Thus, the change in configurational entropy in process (9b) is

$$\Delta S_{(b)} = R \ln \left(\int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega \frac{\sigma_a}{\sigma_{R_N}} \right), \quad (20)$$

where σ_{R_N} is the symmetry number of a ring of N links. Hence

$$\Delta G_{(b)}^\circ = \Delta H_{(b)}^\circ - RT \ln \left(\int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega \frac{\sigma_a}{\sigma_{R_N}} \right). \quad (21)$$

Adding the Gibbs free energy change for processes (9a) and (9b), we obtain the change in the Gibbs free energy for the entire process (8):

$$\begin{aligned} \Delta G &= -T(\Delta S_{(a)} + \Delta S_{(b)}) \\ &= -RT \ln \left(\frac{\int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega}{N_A \sigma_{R_N} \delta\mathbf{r} \frac{\delta\omega}{4\pi}} \right). \end{aligned} \quad (22)$$

This allows us to extract the J factor:

$$J = \frac{4\pi \int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega}{N_A \sigma_{R_N} \delta\mathbf{r} \delta\omega}. \quad (23)$$

Finally, when taking the infinitely thin or WLC limit for long chains ($L \gg b$), we can assume that around $\mathbf{r} = 0$, $\mathbf{C}(\mathbf{r}, \omega) \approx \frac{\mathbf{C}(0)}{4\pi}$ is approximately uniform and independent of ω . Defining $\delta\mathbf{r}$ by $d_{\min} = 0$ and $\delta\omega' = 4\pi$, while taking $\delta\mathbf{r} \rightarrow 0$, the expression above reduces to

$$J = \frac{4\pi \int_{\delta\mathbf{r}} \int_{\delta\omega} \mathbf{C}(\mathbf{r}, \omega) d\mathbf{r} d\omega}{N_A \sigma_{R_N} \delta\mathbf{r} \delta\omega} \approx \frac{\mathbf{C}(0) \delta\mathbf{r} \delta\omega}{N_A \sigma_{R_N} \delta\mathbf{r} \delta\omega} = \frac{\mathbf{C}(0)}{N_A \sigma_{R_N}}, \quad (24)$$

which is the original expression by Flory [2] for the WLC J factor. Note that for the case of DNA cyclization $\sigma_{R_N} = 1$ [22]. We use the expression in Eq. (23) to compute the J factor by numerical simulation.

III. NUMERICAL IMPLEMENTATION OF SAWLC

A. Numerical implementation overview

In order to evaluate the SAWLC looping probabilities, we developed a Monte Carlo algorithm that is capable of generating a large number of plausible self-avoiding DNA chains. The Monte Carlo simulation was written in CUDA C++, and was executed on two NVIDIA GeForce GTX TITAN cards. The length of the generated chains ranged from 50 to 5000 links. Each chain in our ensemble was grown one link at a time by selecting the link's orientation according to the distribution described in Sec. III B, taking into account disallowed directions due to the excluded volume of previous links. In addition, each completed chain was assigned a Rosenbluth weight as described in Sec. III C, resulting in a faithful representation of the full configurational space. Typical sizes of generated ensembles were $N_c \approx 10^{10}$ chains.

We computed the mean-square end-to-end distance $\langle R^2 \rangle$ and the polymer end-to-end separation power-law exponent ν ($\sqrt{\langle R^2 \rangle} \propto N^\nu$) in order to compare our simulation to previous results [9, 23]. For the computation of the J factor, we computed the function $\mathbf{C}(\mathbf{r}, \omega)$ describing the distribution of the end-to-end vector \mathbf{r} with angle ω from \hat{t}_1 [see Eq. (23)]. The procedures used in the computation of these expressions are detailed in Sec. III C.

In the following, we denote the number of links in the simulated chain by N , the length of the links by l , the length of the chain by $L \equiv Nl$, the diameter (or width) of the chain by w , and the Kuhn length of the corresponding WLC as b , where b is given by [12]

$$\frac{b}{l} = \left(\frac{a - 1 + a \coth a}{a + 1 - a \coth a} \right), \quad (25)$$

and where a is the bending constant of the chain.

B. Sampling of chain angles

Our goal was to sample angles θ_i, ϕ_i that satisfy the probability distribution,

$$\begin{aligned} P_i^{\theta, \phi}(\cos \theta_i \in [-1, 1]; \phi_i \in [0, 2\pi)) \\ \propto \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}) \\ = \exp[a(1 - \cos \theta_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}). \end{aligned} \quad (26)$$

Note that we sampled $\cos \theta_i$ and not θ_i according to this distribution, since the quantity that is distributed uniformly over the unit sphere is $\cos \theta_i$ and not θ_i .

We chose to sample the i th link's orientation angles $(\cos \theta_i, \phi_i)$ using inversion sampling [24]. Here the inversion sampling of a single random variable X from a probability distribution function [PDF(X)] is carried out by, first, integrating the PDF(X) over the entire range of X , resulting in a cumulative distribution function [CDF(X)]. Then, a random number y in the range $(0, 1)$ is generated from a continuous uniform distribution. Finally, the desired variable value x is extracted from the inverse function of the CDF(X) at y , i.e., $x = \text{CDF}^{-1}(y)$. In the case discussed here, the two-dimensional variable space $(\cos \theta_i, \phi_i)$ can be mapped to a one-dimensional space due to azimuthal symmetry considerations in the elastic energy. We computed the CDF(X) numerically, using the adaptive Gaussian integration method [25], and subsequently inverted it.

In order to show how the mapping from the two-dimensional variable space $(\cos \theta_i, \phi_i)$ to a one-dimensional space is carried out, we first consider the case of the chain without the excluded volume constraint. Namely, we compute the inverse of the following function:

$$I(\cos \theta_i) = \frac{\int_{-1}^{\cos \theta_i} d \cos \theta \int_0^{2\pi} d\phi e^{-a(1-\cos \theta)}}{\int_{-1}^1 d \cos \theta \int_0^{2\pi} d\phi e^{-a(1-\cos \theta)}}, \quad (27)$$

which allows us to extract $\cos \theta_i$, provided that some random number y corresponding to $I(\cos \theta_i)$ is generated in the range $(0, 1)$. Finally, we generate ϕ_i from a uniform distribution over the entire range $\phi_i \in [0, 2\pi)$ to complete the two-angle set.

If we reintroduce the hard-wall potential, the above integral now changes to

$$I(\cos \theta_i) = \frac{\int_{-1}^{\cos \theta_i} d \cos \theta \int_{\text{allowed } \phi \text{ for } \theta} d\phi e^{-a(1-\cos \theta)}}{\int_{-1}^1 d \cos \theta \int_{\text{allowed } \phi \text{ for } \theta} d\phi e^{-a(1-\cos \theta)}}. \quad (28)$$

The integral over ϕ is no longer over the entire $[0, 2\pi)$ range, as some directions in space cannot be assumed by the new link due to the excluded volume constraint. Thus, for each possible θ angle of the new link, there is a different range of ϕ values

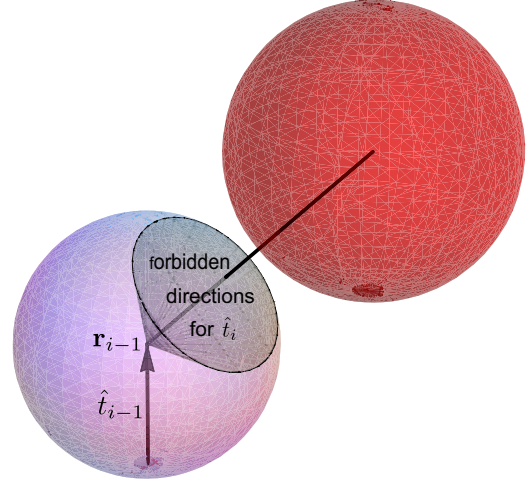


FIG. 3. (Color online) Forbidden directions for \hat{t}_i . Purple (light gray) sphere is the excluded volume around \mathbf{r}_{i-1} [the endpoint of the $(i-1)$ th link]. The red (dark gray) sphere is a potentially colliding object. If \hat{t}_i assumes any direction inside the black (dark gray) cone this would result in a collision of the hard-wall sphere at \mathbf{r}_i with the red sphere.

that are allowed. The situation is illustrated in Fig. 3. The red sphere is an object in close proximity to the endpoint of the $(i-1)$ th link. The black cone maps the directions forbidden for \hat{t}_i , as that would cause an overlap between the i th link and the red sphere. These directions must be excluded from the integration in (28). As a result, we may proceed similarly to the unconstrained case, except that in this case the angle ϕ_i is no longer sampled from the full $[0, 2\pi)$ range, but from a smaller range of possible values.

In order to compute the integral in (28) and to generate ϕ_i , we first compute the set of values that are allowed for ϕ_i for each value of θ_i . Since there can be more than one colliding object, the set of allowed values for ϕ_i need not be one continuous range. Rather, these values can be organized into $m \geq 1$ consecutive ranges $[r_{i,0}^a, r_{i,0}^b], \dots, [r_{i,m-1}^a, r_{i,m-1}^b]$, such that

$$\phi_i^{\text{tot}} = \sum_{j=0}^{m-1} (r_{i,j}^b - r_{i,j}^a) \leq 2\pi. \quad (29)$$

Thus, for each θ_i the ϕ_i angle can be sampled from a stepwise uniform distribution:

$$P_i^{\phi}(\theta_i, \phi_i \in (0, 2\pi)) = \frac{\sum_{j=0}^{m-1} \Theta(\phi_i - r_{i,j}^a) \Theta(r_{i,j}^b - \phi_i)}{\phi_i^{\text{tot}}}. \quad (30)$$

1. Mapping disallowed directions for the i th link

After generating $i-1$ links of the chain, the endpoint is located at \mathbf{r}_{i-1} and the link direction is \hat{t}_{i-1} . We now have to map all the directions that \hat{t}_i cannot assume (as shown in Fig. 3). In the following discussion we focus on one of the chain joints in the vicinity of \mathbf{r}_{i-1} which limits the range of \hat{t}_i . We denote the location of this potentially colliding joint as \mathbf{r}_p and the vector from the endpoint \mathbf{r}_{i-1} to the center of the

potentially colliding sphere as \mathbf{r} :

$$\mathbf{r} = \mathbf{r}_p - \mathbf{r}_{i-1}. \quad (31)$$

We denote by γ the angle between the direction of the last link \hat{t}_{i-1} and \mathbf{r} :

$$\cos \gamma = \frac{\mathbf{r}}{|\mathbf{r}|} \cdot \hat{t}_{i-1}. \quad (32)$$

We note that the forbidden directions for \hat{t}_i form a cone around \mathbf{r} (Fig. 3). We denote half of the opening angle of the cone by α which can then be calculated from simple geometrical considerations as

$$\cos \alpha = \frac{|\mathbf{r}|^2 - w^2 + l^2}{2|\mathbf{r}|l}. \quad (33)$$

A certain (θ_i, ϕ_i) direction is disallowed for \hat{t}_i if the vector $\mathbf{r}_i = \mathbf{r}_{i-1} + l\hat{t}_i$ is at a distance $|\mathbf{r} - \mathbf{r}_i|^2 \leq w^2$ from \mathbf{r}_p . We proceed to work in the local coordinates of the $(i-1)$ th link, assuming for now that \mathbf{r} is in the XZ plane in these coordinates:

$$\mathbf{r} = |\mathbf{r}|(\sin \gamma, 0, \cos \gamma), \quad (34)$$

$$\mathbf{r}_i = l\hat{t}_i = l(\sin \theta_i \cos \phi_i, \sin \theta_i \sin \phi_i, \cos \theta_i). \quad (35)$$

Disallowed values of (θ_i, ϕ_i) satisfy the following inequality:

$$|\mathbf{r} - \mathbf{r}_i|^2 = (|\mathbf{r}| \sin \gamma - l \sin \theta_i \cos \phi_i)^2 + (|\mathbf{r}| \cos \gamma - l \cos \theta_i)^2 + l^2 \sin^2 \theta_i \sin^2 \phi_i \leq w^2. \quad (36)$$

We can thus map the disallowed range for \hat{t}_i in two steps. We first determine the range of $\cos \theta_i$ for which a collision can occur at all. Then, for each θ_i in this range we determine the range of ϕ_i for which a collision does occur.

2. The θ_i range for which a collision occurs

The critical values $\cos \theta_{\pm}$ for which $|\mathbf{r} - \mathbf{r}_i|^2 = w^2$ can be derived from Eq. (36) as follows:

$$\cos \theta_{\pm} = \cos(\gamma \mp \alpha) = \cos \gamma \cos \alpha \pm \sin \gamma \sin \alpha. \quad (37)$$

Caution is needed when interpreting the results in (37). It is true that choosing $\cos \theta_i \in [\cos \theta_-, \cos \theta_+]$ will result in collision for some values of ϕ_i (Fig. 4). However, this does not take into account additional θ_i ranges for which there is a collision for all values of ϕ_i . Such a situation is shown in Fig. 4(b): For $\cos \theta_i \in [\cos \theta_-, \cos \theta_+]$ obviously a collision occurs for some values of ϕ_i , but for $\cos \theta_i \in [\cos \theta_+, 1]$ a collision occurs for all values of ϕ_i .

If the colliding object is located at $\cos \gamma > \cos \alpha$ or at $\cos \gamma < -\cos \alpha$ there is a range of θ_i values for which a collision occurs for all values of ϕ_i . To summarize,

(i) There are disallowed ϕ_i angles for $\cos \theta_i \in [\cos \theta_-, \cos \theta_+]$.

(ii) If $\cos \gamma > \cos \alpha$, for $\cos \theta_i \in [\cos \theta_+, 1]$ there is always a collision.

(iii) If $\cos \gamma < -\cos \alpha$, for $\cos \theta_i \in [-1, \cos \theta_-]$ there is always a collision.

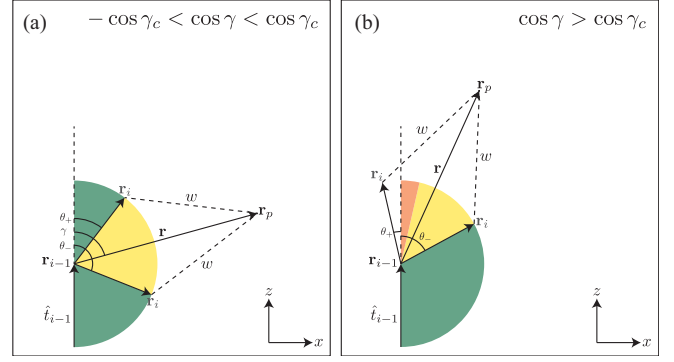


FIG. 4. (Color online) θ_i ranges mapping. θ_i ranges that do not cause a collision for all values of ϕ_i are marked in green (dark gray). θ_i ranges that cause a collision for some values of ϕ_i are marked in yellow (light gray). θ_i ranges that cause a collision for all values of ϕ_i are marked in red (medium gray). (a) If $\theta_i \in [\theta_+, \theta_-]$, a collision occurs for some values of ϕ_i . (b) If $\theta_i \in [0, \theta_+]$, a collision occurs for all values of ϕ_i .

3. The ϕ_i range in which a collision occurs, for specific θ_i

We begin with Eq. (36) and find the critical value of ϕ_i (ϕ_{ic}) for a specific value of θ_i :

$$\cos \phi_{ic}(\theta_i) = \frac{\cos \alpha - \cos \gamma \cos \theta_i}{\sin \gamma \sin \theta_i}. \quad (38)$$

A full collision (no possible values for ϕ_{ic}) occurs when

$$\frac{\cos \alpha - \cos \gamma \cos \theta_i}{\sin \gamma \sin \theta_i} < -1. \quad (39)$$

No collision occurs when

$$\frac{\cos \alpha - \cos \gamma \cos \theta_i}{\sin \gamma \sin \theta_i} > 1. \quad (40)$$

The disallowed angles for a specific θ_i are

$$-\phi_{ic}(\theta_i) \leq \phi_{\text{disallowed}}(\theta_i) \leq \phi_{ic}(\theta_i). \quad (41)$$

We can now generalize to the case when \mathbf{r} is not in the XZ plane. If $\mathbf{r} = |\mathbf{r}|(\sin \gamma \cos \beta, \sin \gamma \sin \beta, \cos \gamma)$ then

$$\beta - \phi_{ic}(\theta_i) \leq \phi_{\text{disallowed}}(\theta_i) \leq \beta + \phi_{ic}(\theta_i). \quad (42)$$

C. Faithful ensemble sampling

For the simulation of the WLC model we used importance sampling (IS; see Appendix A 1). For the simulation of the SAWLC model (and the thick freely jointed self-avoiding chain) we used weighted-biased sampling (WBS; see Appendix A 2) [26], which is based on a method developed by Rosenbluth and Rosenbluth [27]. A variation of this method was recently used to sample dsDNA configurational space [12] for much longer chains than the ones used in this work to examine the cyclization process.

The partition function for an ensemble of N_c sampled chains of N links in WBS is given by

$$Z_N = \sum_{j=1}^{N_c} W(\{\theta_N, \phi_N\}_j), \quad (43)$$

where $W(\{\theta_N, \phi_N\}_j)$ is the Rosenbluth factor of the j th chain in the ensemble:

$$\begin{aligned} W(\{\theta_N, \phi_N\}) &= \prod_{i=1}^N \int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d \phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}) \\ &= \prod_{i=1}^N \int_{\text{non-colliding directions}} d \theta_i d \phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)], \end{aligned} \quad (44)$$

with Θ_i^{hw} defined in Eq. (7).

The average value $\langle f \rangle$ of a physical property is calculated as

$$\langle f \rangle = \frac{\sum_{j=1}^{N_c} f(\{\theta_N, \phi_N\}_j) W(\{\theta_N, \phi_N\}_j)}{\sum_{j=1}^{N_c} W(\{\theta_N, \phi_N\}_j)}. \quad (45)$$

This allows the computation of $\langle R^2 \rangle$. The power-law exponent ν was extracted from the computed $\langle R^2 \rangle$ (see Appendix B 1).

To approximate $C(\mathbf{r}, \omega)$, the configurational space for $\mathbf{r} = (r, \omega')$ and ω is partitioned into bins, with bin volume $d\mathbf{r} \times d\omega$. In each bin we compute the probability of the end-to-end vector \mathbf{r} and relative bond orientation to be contained in the bin:

$$\begin{aligned} P_{\text{bin}}(r, \omega', \omega) &= \lim_{N_c \rightarrow \infty} \frac{\sum_{j=1}^{N_c} f_{\text{bin}}(\mathbf{r}, \omega, \{\theta_N, \phi_N\}_j) W(\{\theta_N, \phi_N\}_j)}{\sum_{j=1}^{N_c} W(\{\theta_N, \phi_N\}_j)}, \end{aligned} \quad (46)$$

where

$$f_{\text{bin}}(\mathbf{r}, \omega, \{\theta_N, \phi_N\}) = \begin{cases} 1 & \mathbf{r}_N \text{ and } \hat{t}_N \text{ relative to } \hat{t}_1 \text{ are contained} \\ & \text{in the bin defined by } \mathbf{r}, \omega \text{ and } d\mathbf{r}, d\omega. \\ 0 & \text{otherwise} \end{cases} \quad (47)$$

$C(\mathbf{r}, \omega)$ is then computed by

$$C(\mathbf{r}, \omega) = \frac{P_{\text{bin}}(r, \omega', \omega)}{d\mathbf{r} \times d\omega}. \quad (48)$$

IV. RESULTS

A. DNA end-to-end distance analysis and comparison to RG results

In order to verify that our algorithm generates a faithful swollen chain ensemble, we simulated chains of up to 5000 links and compared our simulation results to previously published renormalization group (RG) calculations [9] and Monte Carlo results [12]. We simulated chains with the touching-beads model [$l = w$; see Fig. 1(b)] and dsDNA-like parameters of $b \approx 106$ nm and $w \approx 4.6$ nm.

In Fig. 5 we plot the RG predictions and our simulation results for the polymer end-to-end separation power-law exponent ν as a function of the normalized chain length L/b . We display three cases: a “thick” polymer with dsDNA-like parameters $w/b \approx 4.6/106 \ll 1$ (SAWLC), a zero-thickness dsDNA (WLC), and a fully flexible “thick” polymer with $w/b = 1$ (self-avoiding freely jointed chain, SAFJC). The simulation shows that the SAWLC (dashed green curve and green squares)

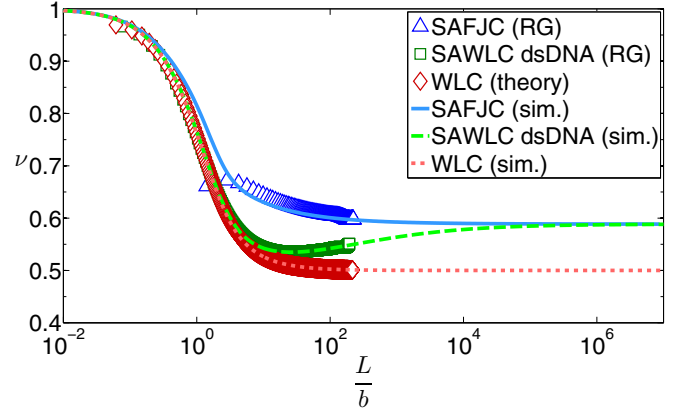


FIG. 5. (Color online) Comparison of the simulation results to the RG calculations and WLC theory. Power-law exponent ν of the end-to-end distance $\sqrt{\langle R^2 \rangle}$ is plotted as a function of chain length L/b . We compare three cases (top to bottom): a freely jointed “thick” polymer (SAFJC) with $w/b = 1$ and $l = w$ (solid blue line, RG; blue triangles, simulation), a “thick” polymer (SAWLC) with dsDNA-like parameters $w/b = 4.6/106 \ll 1$ and $l = w$ (dashed green line, RG; green squares, simulation), and a zero-thickness dsDNA (WLC) with $l/b = 4.6/106$ (dotted red line, WLC; red diamonds, simulation). The excluded volume parameter in the RG calculations was rescaled to match the excluded volume in the simulations with $\bar{u}_{\text{RG}} = 1.5u$ following Tree *et al.* [12].

maintains a WLC-like behavior (dotted red curve and red diamonds) for lengths that are up to several times larger than the persistence length $l_p \equiv b/2$. At $L/b \approx 30$ the curve and simulation begin to bend up and approach the SAFJC prediction until convergence of the SAWLC and the SAFJC is predicted to occur by RG at $L/b \approx 10^4$. It is important to note that while our simulation does not capture the exact convergence to the predicted SAWLC behavior due to chain-length limitations of our algorithm, the trends are precisely tracked over the range of the RG predictions. Moreover, the length at which the divergence away from the WLC behavior occurs (the length where the dashed green and dotted red lines separate) is highly dependent on the width parameter (w/b). This result is consistent with experimental observations [10] and simulation [12].

For the case of the SAFJC ($w/b = 1$), our simulation produces a decaying power law that converges to the Flory prediction of $\sqrt{\langle R^2 \rangle} \propto N^{-0.5876}$ for long chain lengths. Note that our simulation for the semiflexible chain deviates from the RG predictions only for very short chains ($N < 5$). However, for the fully flexible chain our simulation converges on the RG predictions for values $N > 20$. These deviations are a natural consequence of the discrete simulation process, as our simulation employs finite size links while the RG and WLC predictions are for continuous chain models, which correspond to the case $l \rightarrow 0$ and $N \rightarrow \infty$. The larger short-scale deviation from the RG predictions generated by our simulation for the SAFJC (deviation of the blue triangles from the solid blue line for $L/b < 10^2$) can be explained by short-range excluded volume interactions, which are properly simulated by our algorithm and are for the most part neglected in the RG model [9].

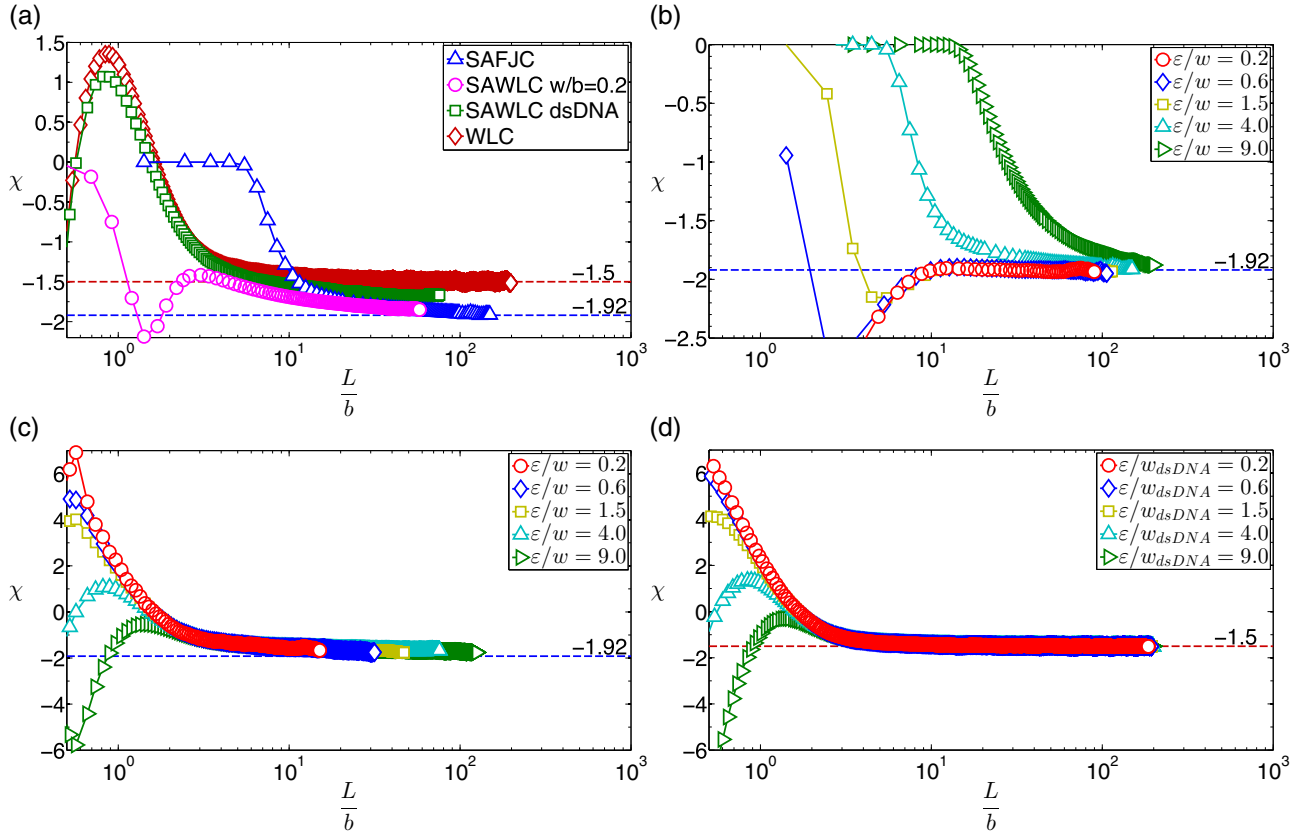


FIG. 6. (Color online) The J -factor power law. (a) Comparison of J -factor power-law exponents χ for the WLC ($\frac{w}{b} = 0$) (red diamonds), SAWLCs with $\frac{w}{b} = 0.046$ (green squares) and $\frac{w}{b} = 0.2$ (magenta circles), and the SAFJC ($\frac{w}{b} = 1$) (blue triangles), using $d_{\min} = w$ and $\frac{\varepsilon}{w} = 4$. The blue (bottom) and red (top) dashed lines correspond to the SAFJC and WLC power laws of -1.92 and -1.5 , respectively. (b)–(d) Effect of ε on the J -factor power-law exponent. (b) J -factor power-law exponent for the SAFJC with various $\frac{\varepsilon}{w}$. (c) J -factor power-law exponent for the SAWLC with $\frac{w}{b} = \frac{4.6}{106}$ and with various $\frac{\varepsilon}{w}$. (d) J -factor power-law exponent for the WLC ($\frac{w}{b} = 0$) with $b = 106$ nm and with various $\frac{\varepsilon}{w_{\text{dsDNA}}}$, where w_{dsDNA} is taken to be equal to that of dsDNA.

B. The SAWLC J factor in the entropic regime

We next proceeded to model the looping or cyclization J factor as defined in Eq. (23). We first examined the entropic regime, namely $\frac{L}{b} \gg 1$. For the sake of simplicity, for this regime we set $d_{\min} = w$ and $\delta\omega' = \delta\omega = 4\pi$ in (14) and (15). In Fig. 6 we plot the J -factor power-law exponent χ as a function of the normalized chain length $\frac{L}{b}$. Previous theoretical studies have shown that in the entropic regime, the J -factor scales like $J \propto N^\chi = N^{-\frac{3}{2}}$ for the WLC model and $J \propto N^{-(3\nu+\gamma-1)} = N^{-1.9196}$ for a flexible swollen chain (see Appendix B 2), where ν is the end-to-end power-law exponent and γ is the “universal” exponent [8]. In Fig. 6(a), we plot the J -factor power-law exponents for the WLC ($\frac{w}{b} = 0$, red diamonds), SAWLCs with $\frac{w}{b} = \frac{4.6}{106}$ (green squares) and $\frac{w}{b} = 0.2$ (magenta circles), and the SAFJC ($\frac{w}{b} = 1$, blue triangles). The panel shows that the behavior of the power-law exponent dramatically changes as the value of $\frac{w}{b}$ increases from zero for the ideal WLC behavior to one for the fully swollen chain. Both the semiflexible WLC and the fully flexible SAWLC J factors converge on their respective theoretical predictions. Moreover, our calculations indicate that the SAWLC J factor converges to the asymptotic value of -1.92 in a manner which is highly dependent on the thickness of the respective chains.

To gain a better insight into the power-law behavior of the J factor, we plot in Figs. 6(b)–6(d) the J -factor power-law exponents for the SAFJC, the SAWLC with dsDNA aspect ratio $\frac{w}{b} = \frac{4.6}{106}$, and the WLC, for various values of $\frac{\varepsilon}{w}$, corresponding to the size of the sphere where bond formation between the two chain termini occurs [see Eq. (14)] divided by the chain diameter. We notice that in all three figures the power-law behavior in the far entropic regime converges to the scaling theory’s predicted limit. Instead, the effects of $\frac{\varepsilon}{w}$ seem to be localized to the shorter range elastic regime and the intermediate (elastic-entropic) transition region. However, the effects of varying $\frac{\varepsilon}{w}$ on the J -factor power-law exponent for the SAFJC are striking as compared with the other cases. Here [Fig. 6(b)] we observe a sharp dependence in the elastic regime. Moreover, the value of $\frac{L}{b}$ at which the transition to the entropic regime and the convergence to the scaling-theory power-law exponent occur, is a strong function of $\frac{\varepsilon}{w}$. Consequently, we conclude that this parameter plays an important role in the physical properties of polymers the closer the value of $\frac{w}{b}$ is to one.

C. The SAWLC J factor in the elastic regime

In order to complete our description of the effects of the excluded volume on the J factor, we generated ensembles of short chains up to $\frac{L}{b} \approx 3$. It was previously predicted [8]

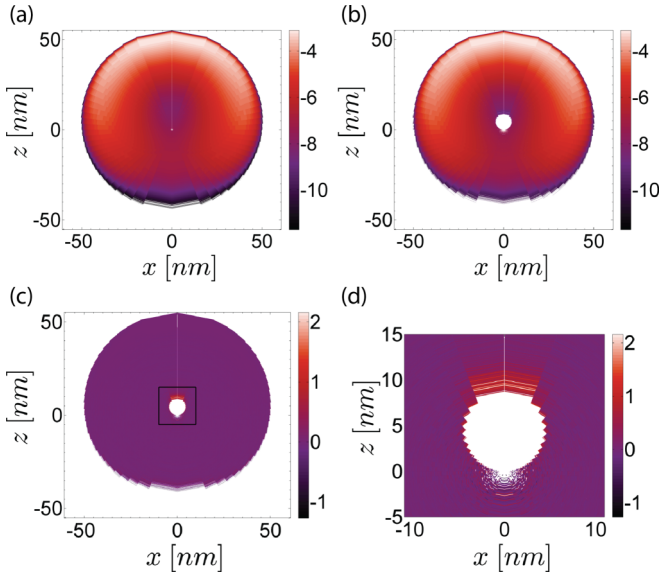


FIG. 7. (Color online) WLC and SAWLC ($d_{\min} = 0$) comparison in the elastic regime. The insets show a cross section of the data in the xz plane. \mathbf{r}_0 is in the origin and $\hat{t}_1 = \hat{z}$. (a) $\log_{10}(\mathbf{C}_{\text{WLC}}(\mathbf{r})/N_A)$ in the xz plane for WLC with $L = 162$ bp. (b) $\log_{10}(\mathbf{C}_{\text{SAWLC}-0}(\mathbf{r})/N_A)$ in the xz plane for SAWLC with $L = 162$ bp, $\frac{w}{b} = \frac{4.6}{106}$ and $d_{\min} = 0$. (c) $\log_{10}(\frac{\mathbf{C}_{\text{WLC}}(\mathbf{r})}{\mathbf{C}_{\text{SAWLC}-0}(\mathbf{r})})$. (d) A blowup of the rectangular section marked in (c). Note, shorter chains were also examined, however, the accumulated statistical data around the origin in those chains was insufficient, and precluded a comprehensive analysis. Qualitatively, however, the shorter chains exhibited the same behavior as shown below in the data ranges that contained significant statistical data.

that for short chains the excluded volume effect should be negligible, as the elastic energy is expected to be the dominant contribution to the probability of looping. Figure 7 shows cross sections of $\mathbf{C}(\mathbf{r}) \equiv \int \mathbf{C}(\mathbf{r}, \omega) d\omega$ in the xz plane for both the WLC and the SAWLC with $\frac{L}{b} \approx \frac{55.2\text{nm}}{106\text{nm}} \approx 0.52$ and $d_{\min} = 0$ and $\delta\omega' = \delta\omega = 4\pi$ in the SAWLC case, with \mathbf{r}_0 at the origin and $\hat{t}_1 = \hat{z}$. Figures 7(a) and 7(b) show that the spatial distributions for the end-to-end separations are highly anisotropic functions in the z direction for both the SAWLC and the WLC. This implies that the probability of looping itself must be highly anisotropic. In addition, both distributions look similar, where the only significant deviation observed for the self-avoiding from the non-self-avoiding distribution emerges from the space occupied by the chain around the origin [white circle in Fig. 7(b)]. In order to quantify this observation, we divided the two distributions, as shown in Fig. 7(c). It is evident from this panel that while both the self-avoiding and non-self-avoiding $\mathbf{C}(\mathbf{r})$ functions are similar, they differ around the origin in the positive z direction, where the SAWLC has reduced probability [Fig. 7(d)]. This is expected, since $\mathbf{r}_1 = l\hat{z}$, rendering the volume around the origin inaccessible for \mathbf{r}_N .

We next examined the case in which $d_{\min} \neq 0$ (and $\delta\omega' = \delta\omega = 4\pi$), implying that looping is now defined for some finite volume around the origin for the WLC. In the SAWLC model, this condition translates to a larger volume around the origin for which the distribution vanishes [see Fig. 8(a)]. However, the actual deviation from the non-self-avoiding WLC distribution function is for the most part negligible, as can be seen from

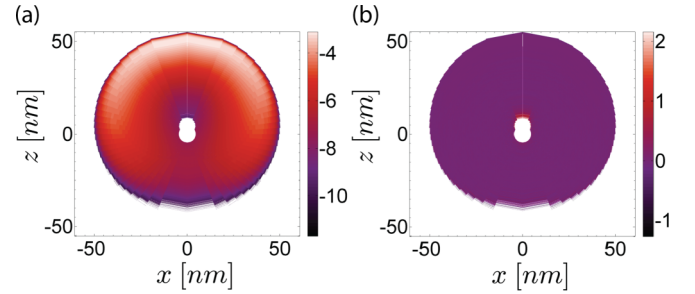


FIG. 8. (Color online) WLC and SAWLC ($d_{\min} = w$) comparison in the elastic regime. The insets show a cross section of the data in the xz plane. \mathbf{r}_0 is at the origin and $\hat{t}_1 = \hat{z}$. (a) $\log_{10}(\mathbf{C}_{\text{SAWLC}-w}(\mathbf{r})/N_A)$ in the xz plane for SAWLC with $L = 162$ bp, $\frac{w}{b} = \frac{4.6}{106}$, and $d_{\min} = w$. (b) $\log_{10}(\frac{\mathbf{C}_{\text{WLC}}(\mathbf{r})}{\mathbf{C}_{\text{SAWLC}-w}(\mathbf{r})})$.

Fig. 8(b), and is similar in its scope to the minimal effect described for $d_{\min} = 0$ above. Consequently, any deviation of the SAWLC J factor from the WLC prediction in the elastic regime is not due to some global modification of the underlying distribution, but rather arises from a local redefinition of $\delta\mathbf{r}$ in Eq. (23), which is imperative due to the presence of the excluded volume around the origin.

Computation of $\mathbf{C}_{\text{SAWLC}}(\mathbf{r})$ is extremely time-consuming for chain lengths shorter than ≈ 160 bp. We therefore utilized the apparent identity of $\mathbf{C}_{\text{SAWLC}}(\mathbf{r})$ and $\mathbf{C}_{\text{WLC}}(\mathbf{r})$ far enough away from the origin to approximate $\mathbf{C}_{\text{SAWLC}}(\mathbf{r})$ for shorter chain lengths. We modeled the effect of the excluded volume on the SAWLC J factor by computing $\mathbf{C}_{\text{WLC}}(\mathbf{r})$ and excluded volumes inaccessible to the SAWLC due to both $w > 0$ and $d_{\min} > 0$ in the integration in Eq. (23). The removed volumes were approximated by a cylinder with diameter $2w$ oriented along \hat{z} with the center of the lower base at the origin, and a sphere with radius d_{\min} centered at the origin. Figure 9(a) shows a schematic for the volume inaccessible to the SAWLC superimposed on $\mathbf{C}_{\text{WLC}}(\mathbf{r})$.

Using this approximation for $\mathbf{C}_{\text{SAWLC}}(\mathbf{r})$, we were able to study the SAWLC J factor for short chain lengths. In Fig. 9(b), we study the case in which looping is defined for some constant bond-center to bond-center separation ($d_{\min} + \varepsilon = 4.6$ nm) for varying values of shell thickness ε . The figure shows that at some shell thickness values (dotted blue line), the SAWLC J factor is larger by about half an order of magnitude as compared to the WLC J factor (solid aqua line) for loop lengths $L \sim 100$ bp. In Fig. 9(c), we compare the J factor for several SAWLCs to their WLC counterparts, keeping ε constant. The figure shows that as the looping criterion for bond-center to bond-center separation ($d_{\min} + \varepsilon$) increases, the deviation from the WLC J -factor prediction increases as well (compare red and red-dashed to blue and blue-dashed lines). This result indicates that the deviation of the SAWLC J factor from the WLC J factor can be made to approach an order of magnitude or more if the bond-center to bond-center looping criterion is increased sufficiently.

V. DISCUSSION

We carried out a simulation study of the looping or cyclization J factor for a self-avoiding wormlike chain (SAWLC)

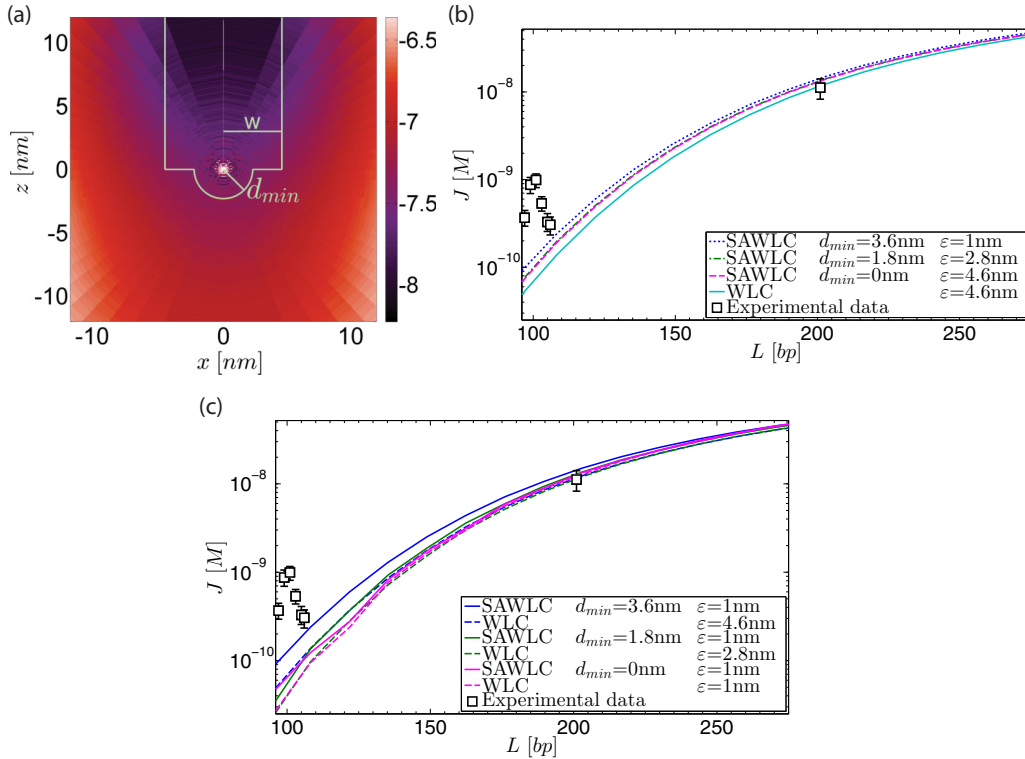


FIG. 9. (Color online) Effect of d_{\min} on the J factor in the elastic regime. (a) The volume enclosed in the light-green outline is inaccessible to a SAWLC chain. (b) J factor for the WLC and SAWLC, varying d_{\min} , constant $d_{\min} + \epsilon = 4.6$ nm. Recent experimental data [7] are shown with black squares. (c) J factor for the WLC and SAWLC, varying d_{\min} , constant $\epsilon = 1$ nm. Recent experimental data [7] are shown with black squares. Note the particular choice of $d_{\min} + \epsilon$ is based on accepted physical values for the effective chain thickness in standard saline conditions. Even though the effects are slightly different in both cases, the magnitude of the effects is less than half an order of magnitude for these values.

model, taking excluded volume effects into consideration for the first time. The distribution of DNA configurations generated by our numerical algorithm was carefully tested by comparing to scaling theory [8] and renormalization group [9] results for the entropic regime, and indeed faithfully represented the swollen chain partition function. In particular, our algorithm was able to extract the universal exponent γ [8], and to reproduce the predicted $N^{-1.92}$ [8,19] J -factor power law for the entropic regime (i.e., where $L \gg l_p$). This stronger power-law drop-off is a direct consequence of the self-avoiding assumption, under which the end-to-end separation grows like $\approx N^{0.5876}$ as compared with the ideal chain scaling of $\approx N^{0.5}$. We further showed that the rate of convergence of the SAWLC J -factor power-law exponent to the asymptotic value of -1.92 is highly dependent on the semiflexible chain aspect ratio. This result implies that a highly precise experimental determination of the dsDNA aspect ratio can be made by simply measuring this power-law exponent for various looping lengths. Our model and its numerical implementation thus enabled the study of the swollen-chain J factor in the elastic regime, where swollen-chain analysis to our knowledge has not been carried out previously.

While the statistics in the elastic regime are dominated by elastic energy considerations, our results indicate that the SAWLC model generates a small but detectible effect on the propensity of the chain to loop. Our computations show that in this regime, the angle-independent probability

density function of the end-to-end vector $\mathbf{C}(\mathbf{r})$ is almost identical for the WLC and SAWLC at points which are accessible to both types of chains. However, there is a range of end-to-end vector \mathbf{r} which is inaccessible to the SAWLC [Fig. 9(a)]. The excluded volume of the SAWLC model leads to a redistribution of $\mathbf{C}_{\text{SAWLC}}(\mathbf{r})$, as shown in Fig. 9, with the combined weight of the $\mathbf{C}_{\text{WLC}}(\mathbf{r})$ within the excluded volume being shifted outside of the SAWLC excluded volume. This could result in an increase in $\int_{\delta\mathbf{r}^{\text{accessible}}} \mathbf{C}_{\text{SAWLC}}(\mathbf{r})d\mathbf{r}$. Comparison of our numerical calculations of $\mathbf{C}(\mathbf{r})$ for both models (Fig. 9) shows that this redistribution of weight is roughly uniform over the entire range of \mathbf{r} . This uniformity is not surprising, since we could have generated the ensemble of short chains using importance sampling (IS; see Appendix A 1) and then discarded all self-crossing chains, thereby increasing the weight of all allowed chains uniformly. Moreover, if we assume that $\int_{\delta\mathbf{r}^{\text{inaccessible}}} \mathbf{C}_{\text{WLC}}(\mathbf{r})d\mathbf{r}$ is redistributed uniformly over the \mathbf{r} range, we obtain a relative increase in the J factor on the order of $\Delta J/J \approx 10^{-5}$. This is clearly not the main increase in the SAWLC J factor observed in Fig. 9(b). Instead, the increase that we observe is likely due to the anisotropic nature of $\mathbf{C}(\mathbf{r})$. Namely, the excluded volume for the SAWLC model overlaps the region for which the probability density $\mathbf{C}_{\text{WLC}}(\mathbf{r})$ is particularly low [the region enclosed in the light-green outline in Fig. 9(a), in the positive \hat{z} direction]. Thus $\mathbf{C}_{\text{WLC}}(\mathbf{r})$ averaged over the part of $\delta\mathbf{r}$ accessible to the SAWLC model increases, which in turn leads

to an increase in the J factor [see Eq. (23)]. Interestingly, the particular choice of the boundary conditions for looping (e.g., minimal bond-to-bond separation) determines the extent of the increase. Consequently, the semiflexible chain SAWLC J factor generates the following picture as compared with the WLC J factor: In the elastic regime, there is a slightly increased probability for looping, while in the entropic regime this trend switches and the SAWLC J factor falls off with a power law greater than the -1.5 value predicted for the WLC.

Previous analyses have shown that the value of the J factor at the elastic regime can be made very sensitive to boundary conditions and local deformations on the DNA [7,16,17]. Our result adds another possible contribution that, together with the previous explanations, could generate a cumulative effect that might explain the seemingly anomalous bendability effect observed in cyclization experiments [4,6,7].

Using this insight, what are the experimental implications for looping or cyclization experiments? First, experiments that aim to test looping in the elastic regime will be strongly dependent on boundary conditions. This implies that any observed deviation from a consensus model should first be considered in the context of better models that simulate the effects of the particular boundary conditions. Second, cyclization experiments that are carried out at larger polymer lengths should be considered as the effects of the boundary conditions become negligible for longer polymer contour lengths. In this regime, cyclization measurements can yield accurate estimates for the persistence length of the polymer and effective width or thickness based on measuring the looping probability and comparing to the theoretical predictions shown in Fig. 6. Finally, our model represents a realistic depiction of DNA molecules at length scales that are relevant to biological regulation. We believe that it can form the basis for a more elaborate model that incorporates protein-DNA interactions.

ACKNOWLEDGMENTS

This research was supported by the Israel Science Foundation through Grant No. 1677/12, by the I-CORE Program of the Planning and Budgeting Committee and The Israel Science Foundation (Grant No. 152/11), and the Marie Curie Reintegration Grant No. PCIG11-GA-2012-321675. Y.P. would also like to acknowledge support provided by the Russell Berrie Nanotechnology Institute, Technion. S.G. is supported in part at the Technion by an Aly Kaufman Fellowship.

APPENDIX A: FAITHFUL SAMPLING

1. Importance sampling

For the simple (non-self-avoiding) 3D freely jointed chain (FJC) model [28], faithful ensembles of polymer chains can be

generated computationally by random step-by-step sampling of the angles of each successive link. In this method, a complete N -mer chain is constructed in sequence, where the i th step samples the θ_i, ϕ_i angles to construct an i -mer chain. After sampling a sufficient number of chains, the partition function can be evaluated from the N_c sampled chains as follows:

$$Z_N = \sum_{j=1}^{N_c} e^{-\beta E_j} = N_c, \quad (\text{A1})$$

where $E_j = 0$ in the FJC case. Any physical property $\langle f \rangle$ can be likewise computed:

$$\langle f \rangle = \frac{\sum_{j=1}^{N_c} f(\{\theta_N, \phi_N\}_j) e^{-\beta E_j}}{\sum_{j=1}^{N_c} e^{-\beta E_j}} = \frac{\sum_{j=1}^{N_c} f(\{\theta_N, \phi_N\}_j)}{N_c}. \quad (\text{A2})$$

When trying to collect an ensemble of plausible polymer configurations for the WLC model, the simple sampling algorithm described above is insufficient. Unlike the FJC model, in the WLC model link orientations depend strongly on the elastic energy of bending [see Eq. (2)]. Namely, sharp bends are highly improbable as compared to small perturbations away from the nonbending minima. As a result, a uniform sampling procedure of the relative link orientations will yield a disproportionate number of improbable chains that are characterized by very large bending energies, which will increase by orders of magnitude the computation time necessary to obtain a statistically significant and representative ensemble.

In order to alleviate this problem, and generate a faithful sampling algorithm for the WLC model (i.e., $E^{hw} = 0$), we employ a sampling method termed importance sampling (IS) [29]. In IS we assign the following Boltzmann weight to the i th link:

$$\exp[-\beta E^{el}(\theta_i, \phi_i)]. \quad (\text{A3})$$

We then sample the chain links according to the following probability distribution:

$$p^{el}(\theta_i, \phi_i) = \frac{\exp[-\beta E^{el}(\theta_i, \phi_i)]}{\int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d \phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)]}. \quad (\text{A4})$$

As a result, the orientation angles that are more likely to occur due to low bending energies will be chosen more frequently, thus reflecting faithfully the underlying distribution. Furthermore, since each sampled link appears in a chain of N sampled links, the cumulative probability to obtain each chain is a product of the individual link probabilities sampled at every stage of the chain construction, as follows:

$$\begin{aligned} P_{\text{tot}}^{el}(\{\theta_N, \phi_N\}) &\equiv \prod_{i=2}^N p^{el}(\theta_i, \phi_i) = \prod_{i=2}^N \frac{\exp[-\beta E^{el}(\theta_i, \phi_i)]}{\int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d \phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)]} \\ &= \frac{\exp[-\beta \sum_{i=2}^N E^{el}(\theta_i, \phi_i)]}{\int_{-1}^1 d \cos \theta_1 \int_0^{2\pi} d \phi_1 \cdots \int_{-1}^1 d \cos \theta_N \int_0^{2\pi} d \phi_N \exp[-\sum_i \beta E^{el}(\theta_i, \phi_i)]} = \frac{\exp[-\beta E_{\text{chain}}]}{\int_{\text{all configurations}} \exp[-\beta E_{\text{chain}}]}. \end{aligned} \quad (\text{A5})$$

Thus, this method generates complete WLC chains according to their statistical probabilities in the ensemble.

According to statistics theory [30], if $X : \Omega \rightarrow \mathbb{R}$ is a random variable in some probability space (Ω, \mathcal{F}, P) and one wishes to estimate the expected value of X under P , provided that one has random samples x_1, \dots, x_n generated according to P , then an empirical estimate of $E[X; P]$ is

$$\hat{E}_n[X; P] = \frac{1}{n} \sum_{i=1}^n x_i. \quad (\text{A6})$$

This simplifies the computation of any physical observable $\langle f \rangle$ for the IS to

$$\langle f \rangle = \frac{1}{N_c} \sum_{j=1}^{N_c} f(\{\theta_N, \phi_N\}_j). \quad (\text{A7})$$

2. Weighted-biased sampling

The IS method can be used to generate chains for the SAWLC model by following (A4) and then discarding chains with overlapping links. However, this method is extremely inefficient and leads to extreme sampling attrition [31].

Naively one could replace Eq. (A4) with

$$p_i(\{\theta_i, \phi_i\}) = \frac{\exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\})}{\int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\})}, \quad (\text{A8})$$

where $\Theta_i^{hw}(\{\theta_i, \phi_i\})$ was defined previously in Eq. (7), and follow the same (IS) procedure. However, since the hard-wall potential eliminates a certain percentage of the possible configurations, a misrepresentation of the sampled chains in the ensemble results, as discussed in [27]. Therefore, in order to produce a faithful representation of the partition function, each generated chain has to be assigned with a weight for the excluded volume case.

In order to gain insight into the calculation of these weights and their necessity, we consider the general case of self-avoiding chains. We denote the probability of a chain j with N links to appear in a statistical ensemble as P_j^S :

$$P_j^S = \frac{e^{-\beta E_j}}{Z_N^S}, \quad (\text{A9})$$

where Z_N^S is the partition function of the statistical ensemble. We denote the probability of a chain j with N links to be generated with IS as P_j^{IS} . Due to the way the chain links are sampled,

$$\begin{aligned} P_j^{IS} &= \prod_{i=2}^N p_i(\{\theta_i, \phi_i\}_j) = \prod_{i=2}^N \left(\frac{\exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}_j)}{\int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}_j)} \right) \\ &= \frac{e^{-\beta E_j}}{\prod_{i=2}^N \left(\int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}_j) \right)}. \end{aligned} \quad (\text{A10})$$

While the numerators in (A10) and (A9) agree, the denominators differ. Furthermore, the denominator in (A10) is different for each chain configuration, unlike the denominator in (A9). Thus, to remove the bias for chain j , it is sufficient to take the denominator of (A10) as the counterweight $W(\{\theta_N, \phi_N\})$:

$$W(\{\theta_N, \phi_N\}) = \prod_{i=2}^N \int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}), \quad (\text{A11})$$

which allows one to define the link-by-link counter-weight as

$$\omega_i(\{\theta_i, \phi_i\}) = \int_{-1}^1 d \cos \theta_i \int_0^{2\pi} d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)] \Theta_i^{hw}(\{\theta_i, \phi_i\}) = \int_{\text{non-colliding directions}} d\theta_i d\phi_i \exp[-\beta E^{el}(\theta_i, \phi_i)], \quad (\text{A12})$$

and the Rosenbluth factor as

$$W(\{\theta_N, \phi_N\}) = \prod_{i=1}^N \omega_i(\{\theta_i, \phi_i\}). \quad (\text{A13})$$

The partition function takes on the following form in the WBS case:

$$Z_N = \sum_{j=1}^{N_c} W(\{\theta_N, \phi_N\}_j), \quad (\text{A14})$$

and the average value $\langle f \rangle$ of a physical property is calculated as

$$\langle f \rangle = \frac{\sum_{j=1}^{N_c} f(\{\theta_N, \phi_N\}_j) W(\{\theta_N, \phi_N\}_j)}{\sum_{j=1}^{N_c} W(\{\theta_N, \phi_N\}_j)}. \quad (\text{A15})$$

APPENDIX B: ADDITIONAL COMPUTATIONS

1. Computation of power-law exponents

To compute the power-law exponents ν of $\sqrt{\langle R^2 \rangle}$ and χ of the J factor in Secs. IV A and IV B, respectively, we started off by taking a logarithm of both the data and L/b . The power-law exponent could then theoretically be read from the slope of the logarithm of the data as a function of $\log(L/b)$. The data, however, are very sensitive to noise in the sampled data. Thus, we employed a smoothing algorithm to deduce the slope of the data at each point: We used a sliding window along the values of L/b . The range of the plot covered by the sliding window was fitted to a linear function, and the slope of the linear function was taken as the slope of the plot at the center of the sliding window. The size of the sliding window was increased as L/b increased in order to account for increased noise at greater values of L/b .

2. A scaling theory derivation for the asymptotic power law for looping in the swollen coil regime

The WLC model predicts that the J factor will scale like $N^{-\frac{3}{2}}$ in the entropic regime, when the length of the polymer is much longer than the persistence length ($N \gg l_p$) [28]. Likewise, the swollen chain also exhibits some power-law dependency in the entropic regime. The asymptotic behavior of the SAWLC model in the entropic regime can be inferred from the existing theory for a freely jointed chain on a 3D lattice. In the following derivation we rely on the arguments given in [8,32]. We first assume that in the entropic regime, the probability of looping scales as

$$p_{\text{loop}}(N) \propto N^{-\alpha}. \quad (\text{B1})$$

Since the probability for looping is defined as a ratio of the number of looped polymer configurations to the total number

of polymer configurations, we begin with the asymptotic expression for the total number of available configurations for a self-avoiding chain of N links [33]:

$$Z = \tilde{z}^N N^{\nu-1}, \quad (\text{B2})$$

where \tilde{z} is a number corresponding to the effective lattice space that the chain is defined on (i.e., $\tilde{z} = 6$ for an ideal chain defined on a cubic lattice). Next, the mean square end-to-end distance for a swollen chain follows a scaling law, originally derived by Flory [2]:

$$R_f \propto N^\nu. \quad (\text{B3})$$

Combining these two results, it is possible to derive an expression for the total number of looped states for a self-avoiding random-walk chain. This result is derived for a grid with pixels of size a , such that a loop is formed when the two ends are separated by a distance a . Thus, on a grid, a self-avoiding random-walk chain will form a closed polygon of $N + 1$ edges, which implies [8]

$$Z_{\text{loop}} \propto \tilde{z}^N \left(\frac{a}{R_f} \right)^3, \quad (\text{B4})$$

with a normalizing term which takes into account that the loop can be closed anywhere within a sphere of radius which is roughly equal to the Flory radius. Given this expression, we can now write the expression for the probability for looping in the asymptotic regime [8]:

$$p_{\text{loop}}(N) \equiv \frac{1}{a^3} \frac{Z_{\text{loop}}}{Z} \propto \frac{1}{R_f^3 N^{\nu-1}} \propto N^{-(3\nu+\gamma-1)}, \quad (\text{B5})$$

which is the desired power law. Finally, using the RG derived values for $\nu = 0.5876$ and $\gamma = 1.1568 \simeq 7/6$, [23,34,35] we obtain

$$\chi = 3\nu + \gamma - 1 = 1.9196. \quad (\text{B6})$$

The universal nature of exponent γ implies that the J -factor power law for semiflexible swollen chains ($\frac{w}{b} < 1$) can be predicted simply from the end-to-end exponent ν .

-
- [1] Homer Jacobson and Walter H. Stockmayer, *J. Chem. Phys.* **18**, 1600 (1950).
- [2] Paul J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience Publishers, Olney, 1969).
- [3] Donald M. Crothers, Jacqueline Drak, Jason D. Kahn, and Stephen D. Levene, in *Methods in Enzymology*, Vol. 212, *DNA Structures Part B: Chemical and Electrophoretic Analysis of DNA*, edited by James E. Dahlberg and David M. J. Lilley (Academic Press, Waltham, 1992), pp. 3–29.
- [4] T. E. Cloutier and J. Widom, *Proc. Natl. Acad. Sci. USA* **102**, 3645 (2005).
- [5] John F. Marko and Eric D. Siggia, *Macromolecules* **28**, 8759 (1995).
- [6] Paul A. Wiggins, Thijn van der Heijden, Fernando Moreno-Herrero, Andrew Spakowitz, Rob Phillips, Jonathan Widom, Cees Dekker, and Philip C. Nelson, *Nature Nanotechnology* **1**, 137 (2006).
- [7] Reza Vafabakhsh and Taekjip Ha, *Science* **337**, 1097 (2012).
- [8] Pierre-Gilles Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [9] Zheng Yu Chen and Jaan Noolandi, *J. Chem. Phys.* **96**, 1540 (1992).
- [10] Manish Nepal, Alon Yaniv, Eyal Shafran, and Oleg Krichevsky, *Phys. Rev. Lett.* **110**, 058102 (2013).
- [11] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, *Proc. Natl. Acad. Sci. USA* **90**, 5307 (1993).
- [12] Douglas R. Tree, Abhiram Muralidhar, Patrick S. Doyle, and Kevin D. Dorfman, *Macromolecules* **46**, 8369 (2013).
- [13] P. J. Flory and J. A. Semlyen, *J. Am. Chem. Soc.* **88**, 3209 (1966).

- [14] Hiromi Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971).
- [15] Roee Amit, Hernan G. Garcia, Rob Phillips, and Scott E. Fraser, *Cell* **146**, 105 (2011).
- [16] Jie Yan, Ryo Kawamura, and John F. Marko, *Phys. Rev. E* **71**, 061905 (2005).
- [17] Paul A. Wiggins, Rob Phillips, and Philip C. Nelson, *Phys. Rev. E* **71**, 021909 (2005).
- [18] Paul A. Wiggins and Philip C. Nelson, *Phys. Rev. E* **73**, 031906 (2006).
- [19] Andrew M. Sinclair, Mitchell A. Winnik, and Gerard Beinert, *J. Am. Chem. Soc.* **107**, 5798 (1985).
- [20] O. Kratky and G. Porod, *Rec. Trav. Chim. Pays-Bas.* **68**, 1106 (2010).
- [21] P. J. Flory, U. W. Suter, and M. Mutter, *J. Am. Chem. Soc.* **98**, 5733 (1976).
- [22] E. R. Semlyen, *Cyclic Polymers*, 2nd ed. (Springer, Dordrecht/Boston, 2000).
- [23] Nathan Clisby, *Phys. Rev. Lett.* **104**, 055702 (2010).
- [24] Luc Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).
- [25] Wilhelm M. Pieper, *Communications in Numerical Methods in Engineering* **15**, 77 (1999).
- [26] Juan J. De Pablo and Fernando A. Escobedo in *Advances in Chemical Physics*, Vol. 105, edited by I. Prigogine and Stuart A. Rice (John Wiley & Sons, Hoboken, 1998), pp. 337–367.
- [27] Marshall N. Rosenbluth and Arianna W. Rosenbluth, *J. Chem. Phys.* **23**, 356 (1955).
- [28] Rob Phillips, Jané Kondev, and Julie Theriot, *Physical Biology of the Cell* (Garland Science, New York, 2009).
- [29] Peter Grassberger, *Phys. Rev. E* **56**, 3682 (1997).
- [30] Bert E. Fristedt and Lawrence F. Gray, *A Modern Approach to Probability Theory* (Birkhauser, Boston, 1996).
- [31] Jiro Sadanobu and William A. Goddard, III, *J. Chem. Phys.* **106**, 6722 (1997).
- [32] J. des Cloizeaux, *Phys. Rev. A* **10**, 1665 (1974).
- [33] D. S. McKenzie, *Physics Reports* **27**, 35 (1976).
- [34] Sergio Caracciolo, Maria Serena Causo, and Andrea Pelissetto, *Phys. Rev. E* **57**, R1215(R) (1998).
- [35] Nathan Clisby, Richard Liang, and Gordon Slade, *J. Phys. A: Math. Theor.* **40**, 10973 (2007).