

Extreme bendability of DNA double helix due to bending asymmetry

H. Salari,¹ B. Eslami-Mossallam,² S. Naderi,¹ and M. R. Ejtehadi^{1,3,a)}

¹Department of Physics, Sharif University of Technology, P.O. Box 11155-9161, Tehran, Iran

²Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands

³Center of Excellence in Complex Systems and Condensed Matter (CSCM), Sharif University of Technology, Tehran 145889694, Iran

(Received 27 May 2015; accepted 21 August 2015; published online 10 September 2015)

Experimental data of the DNA cyclization (J-factor) at short length scales exceed the theoretical expectation based on the wormlike chain (WLC) model by several orders of magnitude. Here, we propose that asymmetric bending rigidity of the double helix in the groove direction can be responsible for extreme bendability of DNA at short length scales and it also facilitates DNA loop formation at these lengths. To account for the bending asymmetry, we consider the asymmetric elastic rod (AER) model which has been introduced and parametrized in an earlier study [B. Eslami-Mossallam and M. R. Ejtehadi, *Phys. Rev. E* **80**, 011919 (2009)]. Exploiting a coarse grained representation of the DNA molecule at base pair (bp) level and using the Monte Carlo simulation method in combination with the umbrella sampling technique, we calculate the loop formation probability of DNA in the AER model. We show that the DNA molecule has a larger J-factor compared to the WLC model which is in excellent agreement with recent experimental data. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4929994>]

I. INTRODUCTION

Studying the elastic behavior of DNA molecule is important for understanding its biological functions. One of the most popular theoretical models to explain the elastic behavior of DNA is the harmonic elastic rod model, also called the wormlike chain (WLC) model.^{1,2} In this model, it is assumed that the elastic energy is a harmonic function of local deformations. The WLC model can accurately predict the elastic properties of long DNA molecules and yields a persistence length of about 50 nm for DNA.^{2,3} However, recent experimental^{4–8} and simulational⁹ studies suggest that short DNA molecules are much more flexible than what is predicted by the WLC model. For example, loop formation probability, i.e., the J-factor,¹⁰ for DNA molecules shorter than 100 bp (~34 nm) is several orders of magnitude higher than the prediction of the WLC model.^{4,5,8,11} Although there is some doubt about the results of some of these experiments,^{12–14} they motivated us to generalize the harmonic elastic rod model in order to obtain more realistic models for DNA elasticity.

Different experiments have been carried out to measure the cyclization probability for short DNA molecules. For example, in the work of Cloutier and Widom,⁵ the DNA molecules have short sticky ends. Therefore, when the two DNA ends are close to each other, torsional and axial alignments are required to form a DNA loop, which is then stabilized by the ligase. Thus, the J-factor depends on the concentration of the ligase in the experiment.¹⁵ On the other hand, Vafabakhsh and Ha have used DNA molecules with long sticky ends.⁸ In this case, it is expected that the rate of loop formation depends only on the probability that the two DNA ends reach to each

other and thus is directly related to DNA elasticity. In both experiments, the persistence length of short DNA molecules appears to be much shorter than 50 nm.

The anomalous elastic behavior of short DNA molecules has been the subject of many studies. Mostly, it is suggested that the structural properties^{16,17} or formation of low-energy kinks in highly bent DNA molecules^{15,18,19} may explain this anomaly. It has been shown that local DNA melting of the cyclized DNA increases the J-factor at short length scales.^{20,21} Also, there have been efforts to explain this anomaly by introducing more structural details to the elastic model (e.g., considering anisotropy^{22,23} in bending rigidity of DNA). But these efforts were not successful, as it has been shown that anisotropy has no significant effect in these dimensions²⁴ and even leads to the stiffening of DNA if the molecule is confined in a two dimensional surface.²⁵ The effect of electrostatic interaction²⁶ and shear flexibility of DNA²⁷ have been studied which may increase DNA flexibility, but in this study, we assume that the DNA molecule is neutral and unsharable.

The DNA molecule in its B form suggests that it has an asymmetric structure, in the sense that the opposite grooves of the DNA are not equal in size and the structure.²⁸ Thus, one expects that the energy required to bend the DNA is not only anisotropic but also asymmetric, i.e., the energy required to bend the DNA over its major groove is not equal to the energy needed to bend it over its minor groove by the same angle. There are theoretical analysis,²⁹ experimental evidences,³⁰ and simulation studies³¹ which show that the bending asymmetry may affect the elastic behavior of DNA. In a previous work, we have introduced the asymmetric elastic rod (AER) model to account for the asymmetric bending of DNA.³² In this paper, we evaluate the elastic properties of the AER model, namely, the DNA looping probability, to reveal the relevance of the

^{a)}Electronic address: ejtehadi@sharif.edu

asymmetric bending for short extremely bent DNA molecules. To this end, we exploit the Monte Carlo (MC) simulation in combination with the umbrella sampling (US) technique, which enables us to efficiently sample the rare cyclization events. We show that the AER model is in excellent agreement with the experimental data of the J-factor at short length scales.

II. MODEL AND METHOD

A. Asymmetric elastic rod model

In the elastic rod model, a DNA is represented by a flexible inextensible rod,^{1,2} which can be deformed in response to the external forces or torques. Here, we use the discrete elastic rod model,^{2,33} where the rod is discretized into segments each representing a DNA base pair. In this model, the internal degrees of freedom of the base pairs are neglected, and each base pair is considered as a rigid body. A local coordinate system (material frame) with an orthonormal basis $\{\hat{d}_1, \hat{d}_2, \hat{d}_3\}$ is attached to each base pair. As depicted in Figure 1, \hat{d}_3 is perpendicular to the base pair surface, \hat{d}_1 lies in the base pair plane and points toward the major groove, and \hat{d}_2 is defined as $\hat{d}_2 = \hat{d}_3 \times \hat{d}_1$. Since it is assumed that the DNA is inextensible, each base pair only has three rotational degrees of freedom, and the position of the $(k+1)$ th base pair with respect to the k th base pair is given by the vector $\vec{r}(k)$,²

$$\vec{r}(k) = l_0 \hat{d}_3(k), \quad (1)$$

where $l_0 = 0.34$ nm is the base pair separation. The orientation of the $(k+1)$ th base pair with respect to the k th base pair is determined by a rotation transformation $\mathbf{R}(k)$, which can be parametrized by a vector $\vec{\Theta}(k)$. The direction of $\vec{\Theta}(k)$ is normal to the rotation plane of the k th base pair, and its magnitude determines the rotation angle. The components of $\vec{\Theta}(k)$ in the local coordinate system attached to the k th base pair are denoted by $\Theta_1(k)$, $\Theta_2(k)$, and $\Theta_3(k)$. These components can be regarded as three rotational degrees of freedom of the base pairs around \hat{d}_1, \hat{d}_2 , and \hat{d}_3 , and are called tilt, roll, and twist, respectively.³³ If the values of these three angles are known for all base pairs, the conformation of the DNA can be uniquely determined.

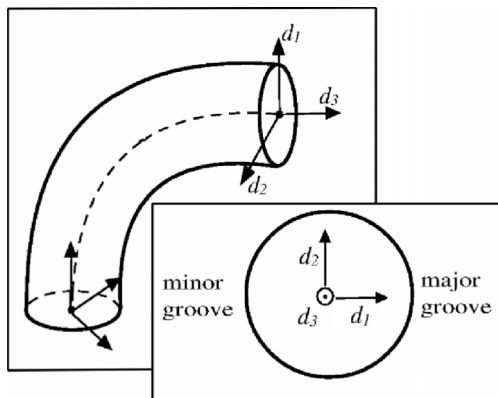


FIG. 1. Parametrization of the elastic rod. The local frame $\{\hat{d}_1, \hat{d}_2, \hat{d}_3\}$ is attached to the rod.

For an inextensible DNA with N base pair steps, the elastic energy depends on the spatial angular velocity $\vec{\Omega}(k) = \frac{\vec{\Theta}(k)}{l_0}$. Therefore, the elastic energy of the AER model³² can be written as

$$E = \sum_{k=1}^N l_0 \mathcal{E}_k^{\text{AS}}[\vec{\Omega}(k)], \quad (2)$$

with

$$\begin{aligned} \mathcal{E}_k^{\text{AS}}[\vec{\Omega}(k)] = k_B T & \left[\frac{1}{2} A_1 \Omega_1(k)^2 + \frac{1}{2} A_2 \Omega_2(k)^2 \right. \\ & + \frac{1}{2} C (\Omega_3(k) - \omega_0)^2 + \frac{1}{3!} F^2 \Omega_2(k)^3 \\ & \left. + \frac{1}{4!} G^3 (\Omega_1(k)^4 + \Omega_2(k)^4) \right], \quad (3) \end{aligned}$$

where k_B is the Boltzmann constant and T is the absolute temperature. The first three terms in Equation (3) correspond to the harmonic parts of the elastic energy, which also appear in the WLC model. A_1 , A_2 , and C are the harmonic elastic constants of DNA, and ω_0 is its intrinsic twist. The remaining terms constitute the anharmonic parts of the elastic energy. The term $+1/3! F^2 \Omega_2^3$ accounts for the asymmetric structure of DNA. The cubic term in Equation (3) will eventually dominate the quadratic term for $\Omega_2 < 0$; therefore, we need to keep the fourth-order correction, $\frac{1}{4!} G^3 (\Omega_1(k)^4 + \Omega_2(k)^4)$, with $G > 0$, for stability of the model.³²

Since there is no coupling term in the model, roll, tilt, and twist can be regarded as independent deformations, and the energy density can be decomposed into three separate terms,

$$\mathcal{E}^{\text{AS}}[\Omega_1, \Omega_2, \Omega_3] = \mathcal{E}_1^{\text{AS}}[\Omega_1] + \mathcal{E}_2^{\text{AS}}[\Omega_2] + \mathcal{E}_3^{\text{AS}}[\Omega_3], \quad (4)$$

where

$$\mathcal{E}_1^{\text{AS}}[\Omega_1] = k_B T \left[\frac{1}{2} A_1 \Omega_1^2 + \frac{1}{4!} G^3 \Omega_1^4 \right], \quad (5)$$

$$\mathcal{E}_2^{\text{AS}}[\Omega_2] = k_B T \left[\frac{1}{2} A_2 \Omega_2^2 + \frac{1}{3!} F^2 \Omega_2^3 + \frac{1}{4!} G^3 \Omega_2^4 \right], \quad (6)$$

and

$$\mathcal{E}_3^{\text{AS}}[\Omega_3] = \frac{1}{2} k_B T C (\Omega_3 - \omega_0)^2. \quad (7)$$

Here, we use the parameters of Ref. 32 obtained by fitting the AER model to the experimental data of Wiggins *et al.*⁶ (see the first row of Table I as model ‘‘A’’). In this parametrization, the bending anisotropy is also considered, where $A_1 \neq A_2$. With this parametrization, the roll energy, \mathcal{E}_2 , has two minima (solid, red curve in Figure 2); one at $\Omega_2 = 0$ and another at $\Omega_2 = -3.3 \text{ nm}^{-1}$, which corresponds to a negative roll of about

TABLE I. Two different parameter sets for the AER and WLC models, indicated as models ‘‘A’’ and ‘‘W,’’ respectively, in this study. The parameter set ‘‘A’’ is obtained from the fitting of the model to the experimental data of Wiggins *et al.*³² and its effective persistence length is about 51 nm, same as model ‘‘W.’’

Model	A_1 (nm)	A_2 (nm)	F (nm)	G (nm)	C (nm)	ω_0 (1/nm)
A	87.00	43.50	7.90	3.20	100	1.8
W	51.00	51.00	0.0	0.0	100	1.8

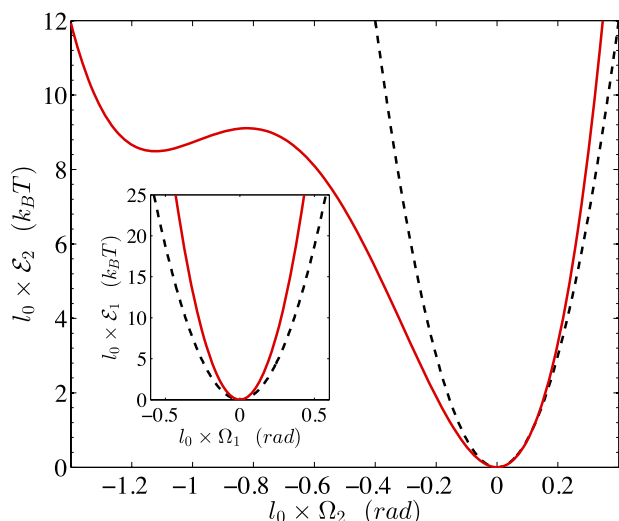


FIG. 2. \mathcal{E}_2 as a function of Ω_2 for two different models, “W” and “A” (Table I). The symmetric model “W,” dashed (black) curve, has one minimum and its curvature is positive everywhere. The asymmetric model “A,” solid (red) curve, has two minima. \mathcal{E}_1 as a function of Ω_1 for both models remain always convex (inset).

60° , with a roll energy about $8 k_B T$. The energy barrier between the two minima is about $9 k_B T$. The existence of a second minimum in \mathcal{E}_2 can lead to the formation of kinks in the minor groove direction. With a large energy barrier between the two minima, one expects that the kinks rarely form in a free DNA at room temperature. However, if the DNA is forced to adopt a tightly bent conformation, the probability of kink formation increases significantly.

The possibility of kink formation in the DNA structure has been originally proposed by Crick and Klug,²⁹ who suggest that DNA can form a kink in the minor groove direction without breaking the base pairs. Also, molecular dynamics simulations on a 94 bp minicircle³¹ show that kinks are formed, with the same structure predicted by Crick and Klug. It is known in the structure of nucleosomal DNA that kinks occur toward the minor groove;³⁰ however, other studies of protein-DNA complexes have shown a tendency to kink in the major groove direction.^{34,35} Many theoretical efforts have tried to explain DNA softening at short lengths by including kinks in the elastic rod model. A simple model has been presented by Nelson, Wiggins, and Phillips to describe the elasticity of kinkable elastic rods.¹⁸ This model is mathematically equivalent to the models of local DNA melting.^{20,21,36} Recently, Vologodskii and Frank-Kamenetskii have proposed another model for the kink formation in DNA.¹⁵ These modified elastic models were rather successful to predict the extreme bendability of short DNA in the way that the kinks absorb much of the curvature required to form a sharp bend. But, in all of these models, the kinks are isotropic, i.e., they can be formed in any direction with equal probability. On the contrary, all atomistic simulations³¹ and experimental data^{29,30,34,35} show that there is a privileged direction for the kink formation. This issue introduces an implicit coupling between the bend and twist and consequently the kink and twist. This coupling has been considered in the AER model in the sense that the DNA kinks form toward the minor groove.

In order to compare the AER and WLC models, we also use another set of parameters, model “W” in Table I. As we will show in Section III, at long length scales, these two models are equivalent and yield the same persistence length, $l_p = 51$ nm. The tilt and roll energies, \mathcal{E}_1 and \mathcal{E}_2 , of these two models are compared in Figure 2.

B. Calculation of J-factor

The loop formation probability of a polymer, which is known as J-factor,³⁷ is defined as the probability that the two ends of the polymer meet each other with axial and torsional alignments. For simplicity, we neglect the torsional alignment and only require that the two ends are close to each other, while the two terminal tangent vectors are parallel. Denoting the separation between the two ends by r and the angle between the two terminal tangent vectors by $\theta = \cos^{-1}(\hat{d}_3(0) \cdot \hat{d}_3(L))$ (see Figure 3), the J-factor for a DNA of length L in molar unit is given by^{37,38}

$$J(L) = \lim_{\substack{r_0 \rightarrow 0 \\ \gamma_0 \rightarrow 1}} \frac{1}{N_a} \frac{3}{4\pi r_0^3} \int_0^{r_0} K(r, L) dr \\ \times \frac{2}{1 - \gamma_0} \int_{\gamma_0}^1 P_\gamma(\gamma, L) d\gamma, \quad (8)$$

where $\gamma = \cos \theta$, N_a is Avogadro’s number, $K(r, L)$ is the normalized distribution function of r , and $P_\gamma(\gamma, L)$ is the normalized distribution function of γ under condition of $r = 0$.

The bending energy of DNA depends on the bending direction. Thus, there is an implicit bend-twist coupling in this model. However, this coupling can barely affect the J-factor if the DNA length is much larger than the helical pitch. Therefore, we expect that for both models, the average trend of the J-factor is given by Equation (8). For end distances near zero ($r/L \approx 0$), the radial distribution function, $K(r, L)$, is proportional to r^2 ,

$$K(r, L) \rightarrow 4\pi r^2 K_0(L) \quad \text{for } r/L \rightarrow 0, \quad (9)$$

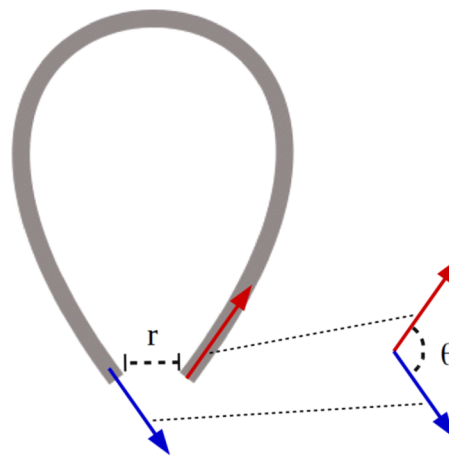


FIG. 3. Schematic figure of a chain with terminal tangent vectors, $\hat{d}_3(0)$ and $\hat{d}_3(L)$, which are indicated by red and blue arrows, respectively. The end-to-end distance is shown by r , and θ is the angle between the tangent vectors.

where $K_0(L)$ is an r -independent function.³⁹ In addition, we have

$$\lim_{\gamma_0 \rightarrow 1} \frac{1}{1 - \gamma_0} \int_{\gamma_0}^1 P_\gamma(\gamma, L) d\gamma = P_\gamma(\gamma = 1, L), \quad (10)$$

then, Equation (8) can be written as

$$J(L) = J_0(L) \times 2P_\gamma(1, L), \quad (11)$$

where

$$J_0(L) = \frac{1}{N_a} K_0(L) \quad (12)$$

is the unconstrained J-factor which does not involve axial and torsional alignments.

C. Simulations

We exploited a Metropolis MC simulation to evaluate the statistical properties of DNA. We do not include the self avoiding in the simulations, since the probability of self crossing is small for the short simulated DNA molecules. For short DNA molecules, the loop formation probability is very low, and thus, the DNA cyclization events are too rare to be observed in the simulations. To overcome this problem, we used the method of US⁴⁰ to evaluate the distribution functions $K(r, L)$ and $P_\gamma(\gamma, L)$. To calculate $K(r, L)$, the reaction coordinate is the end-to-end distance, r . We divided r into 100 successive windows, and for each window performed, a separate MC simulation in which a harmonic bias potential is applied to the end-to-end distance of the DNA. All the harmonic potentials have a common spring constant $\mathcal{A}_r^b = 0.3 k_B T / \text{nm}^2$, and the minimum of each potential lies on the center of the corresponding window. We then found the biased distribution for each individual simulation and used the Weighted Histogram Analysis Method (WHAM)⁴¹ to reconstruct the unbiased distribution function. To calculate $P_\gamma(\gamma, L)$, we set the end-to-end distance to zero, then perform another US by dividing the range of variation of γ into 100 windows, and apply a harmonic bias potential with spring constant $\mathcal{A}_\gamma^b = 40 k_B T$ in each window. The unbiased distribution function $P_\gamma(\gamma, L)$ is then found by WHAM.

In each individual simulation during the umbrella sampling procedure, the first 10^5 MC steps were disregarded to ensure the equilibration of the system and the next 2×10^6 MC steps were considered for sampling. We performed 5 independent simulations in each window to estimate error bars.

III. RESULTS AND DISCUSSION

A. Persistence length and effective bending energy

For a long DNA of length L , the persistence length, l_p , is defined as¹⁹

$$\langle \cos(\theta) \rangle = \exp(-L/l_p). \quad (13)$$

With the parameters given in Table I, the asymmetric model (model “A”) and the wormlike chain model (model “W”) have a common persistence length of about 51 nm. This means that for long DNA molecules with small deformations, the two models are equivalent, and thus, the asymmetric model is effectively

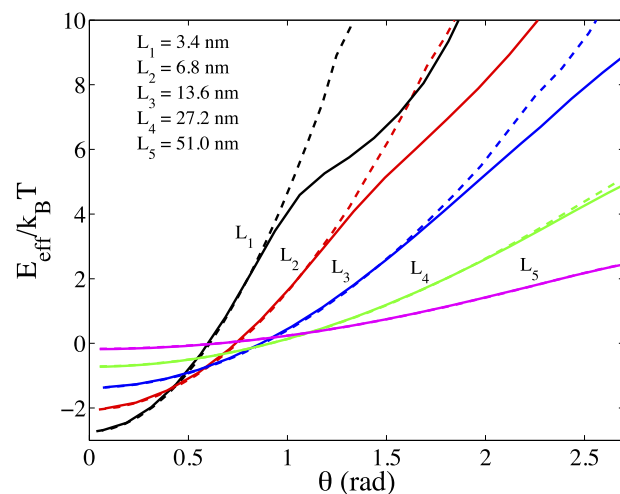


FIG. 4. Monte Carlo simulation results of the effective bending energy, E_{eff} , as a function of the bending angle, θ , for various chain lengths $L = 3.4$ nm (black), $L = 6.8$ nm (red), $L = 13.6$ nm (blue), $L = 27.2$ nm (green), and $L = 51.0$ nm (magenta). Solid and dashed curves correspond to “A” and “W” models, respectively (see Table I).

reduced to an isotropic wormlike chain model. On the other hand, at large bending angles, one expects that the asymmetric structure of DNA affects its elastic properties. To show this, we evaluated the effective bending energy as a function of the bending angle θ , defined as⁴²

$$E_{\text{eff}}(\theta, L) = -k_B T \ln \frac{P(\theta, L)}{\sin \theta}, \quad (14)$$

where $P(\theta, L)$ is the distribution function of the bending angle θ of a DNA of length L . Figure 4 compares the effective energies of the AER model (solid curves) with the WLC model (dashed curves) for different DNA lengths. One can see that at small bending angles, both models follow a common parabola. However, at large bending angles, the effective bending energy of the asymmetric model falls beneath the parabola, which leads to extreme bendability of DNA or formation of the kinks.³² The effect is suppressed as the DNA length increases. It is well expected that for long enough DNA, the effective energy is independent of the structural details and it converges to a parabola.⁶

As Figure 4 shows, the transition between the harmonic and non-harmonic region is smooth. This is because in the AER model, DNA preserves its resistance against bending even in kink region. In the versions of kinkable elastic rod models,^{18,19} where the kinks are assumed to be completely flexible, there is a sharp transition in the curve of the effective bending energy between a parabola and a straight line with zero slope.¹⁸

B. The end-to-end distribution functions

Figures 5(a) and 5(b) compare the radial distribution function, $K(r, L)$, for a DNA of length $L = 33.66$ nm (=99 bp). The triangles (blue) and squares (red) show MC simulation results for models “A” and “W,” respectively (Table I). The solid (black) curve corresponds to the theoretical treatment of Samuel and Sinha⁴³ for the WLC model, which perfectly

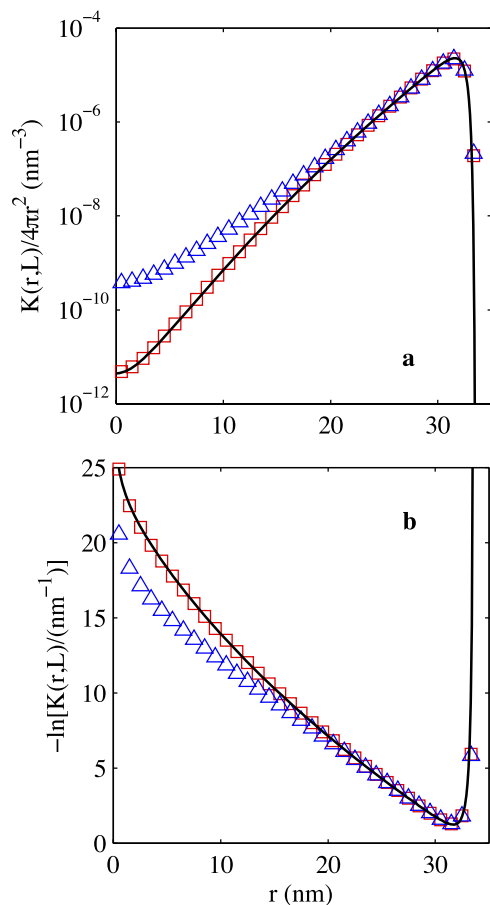


FIG. 5. The radial distribution (a) and free energy (b) as a function of end-to-end distance, r , for a DNA of length $L = 33.66$ nm ($=99$ bp). Data points represent MC simulation results, where triangles (blue) and squares (red) correspond to the models “A” and “W,” respectively. Solid (black) curve corresponds to the theoretical prediction of the WLC model with $l_p = 51$ nm.⁴³ Error bars (not shown) are about the size of the markers.

matches the simulation data. As can be seen in Figure 5(a), there is no significant difference between the two models at large end-to-end distances, while at short end-to-end distances, the radial distribution function in the AER model significantly deviates from that of the WLC model.

Figure 5(b) shows $-\ln(K(r,L))$, the free energy of the DNA as a function of its end-to-end distance. The position of the minimum in the free energy curve corresponds to the most probable end-to-end separation. A relaxed DNA molecule which is shorter than the persistence length tends to be almost straight. As can be seen in Figure 5(b), for $L = 99$ bp, the free energy minimum is very close to the total length of DNA. In this case, we found that the average and the variance of the end-to-end distances in the WLC and AER models differ by less than 0.1%. Therefore, in the experiments which involve long free DNA molecules, such as the DNA stretching experiment, the two models are indistinguishable. On the other hand, the asymmetric bending can significantly affect the outcome of the experiments performed on short and tightly bent DNA molecules, such as the DNA cyclization.

Figure 6(a) compares the distribution functions, $K(r,L)/4\pi r^2$, of models “A” (solid curves) and “W” (dashed curve) for different lengths $L = 16.66, 33.66, 50.66, 67.66, 84.66,$ and

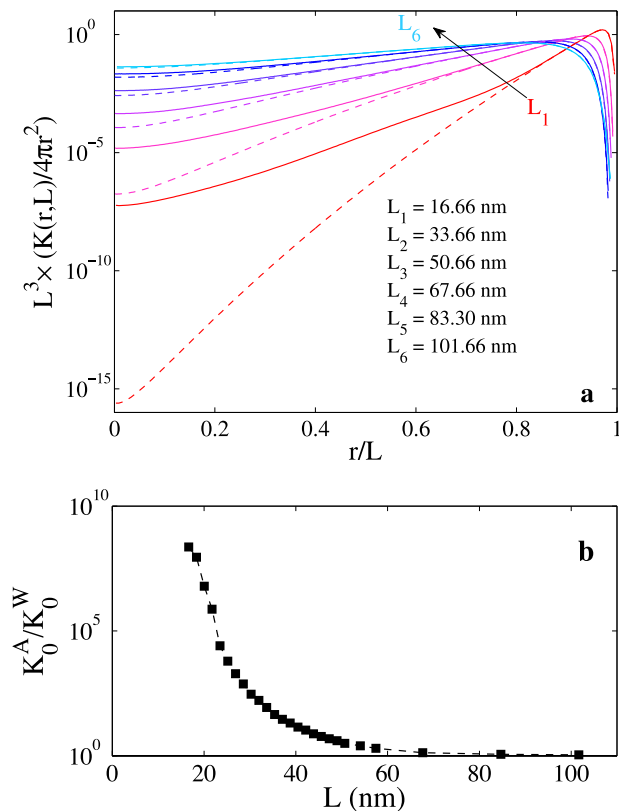


FIG. 6. (a) $K(r,L)/4\pi r^2$ as a function of r for different DNA lengths. Solid and dashed curves, respectively, correspond to the models “A” and “W” in Table I. (b) K_0^A/K_0^W as a function of DNA length.

101.66 nm (or 49, 99, 149, 199, 249, and 299 bp, respectively). One can see that the difference between the two models is disappeared as the DNA length increases. As expected for small end-to-end distances, $K(r,L)/4\pi r^2$ converges to a constant $K_0(L)$ (see Equation (9)) which is proportional to the unconstrained J-factor $J_0(L)$ (Equation (12)). To calculate $K_0(L)$, we average $K(r,L)/4\pi r^2$ in the vicinity of $r = 0$ over the range of $0 \leq r \leq r_0$, where r_0 is chosen such as $\frac{K(r_0,L) - 4\pi r_0^2 K_0(L)}{K(r_0,L)} \approx 0.01$. Figure 6(b) shows the ratio K_0^A/K_0^W as a function of DNA length, where the superscripts “A” and “W” refer to the AER and WLC models as parametrized in Table I. As can be seen, while K_0^A is several orders of magnitude larger than K_0^W for short DNA molecules, the ratio K_0^A/K_0^W approaches unity as the DNA length increases.

C. The distribution function of the angle between the terminal tangent vectors

In Figure 7, $P_\gamma(\gamma, L)$ is plotted against γ for a DNA with $L = 33.66$ nm ($=99$ bp), while its end-to-end distance is kept at $r = 0$. In this figure, triangles (blue) and squares (red), respectively, correspond to the models “A” and “W.” As can be seen, the two distributions are significantly different, but in the vicinity of $\gamma = 1$, they asymptotically approach each other (inset of Figure 7). We found that the peak of the distribution for the AER model generally occurs at a smaller γ compared to the WLC model at short length scales (below the persistence length). For example, for $L = 33.6$ nm, the most probable

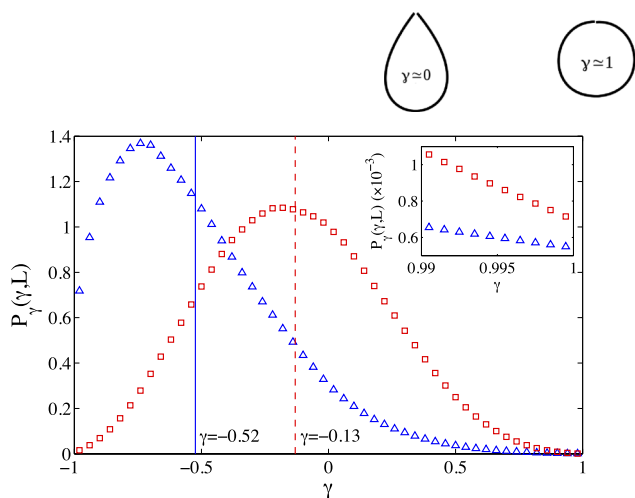


FIG. 7. Top: Different schematic configurations of the chain for different angles between the two DNA tangents. Bottom: The MC simulation results for $P_\gamma(\gamma, L)$ as a function of γ for a DNA loop of length 33.66 nm ($=99$ bp). Triangles (blue) and squares (red) correspond to the models “A” and “W” in Table 1, respectively. Solid (blue) and dashed (red) vertical lines (at -0.52 and -0.13 , respectively) are the mean values of the distribution functions. Inset shows the tails of the distributions near $\gamma = 1$. Error bars (not shown) are about the size of the markers.

values of γ for the models “A” and “W” are -0.74 and -0.18 , respectively, and with $\langle \gamma \rangle^A = -0.52$ and $\langle \gamma \rangle^W = -0.13$ (as indicated in Figure 7 by solid (blue) and dashed (red) vertical lines, respectively). This indicates that when the two ends of a stiff chain meet each other, the angle between the terminal tangent vectors tends to be smaller in the AER model than in the WLC model. This difference reflects the effect of the kink formation on the equilibrium structure of the DNA loop. The same structure also has been reported by other studies considering the kink in the model.^{44,45}

D. Loop formation probability

Figure 8 compares the J-factor of the DNA in the AER (triangles, blue) and the WLC (squares, red) models, as obtained in the MC simulations, where filled and open markers correspond to J and J_0 , respectively (see Sections II B and II C and Equations (11) and (12)). The solid black curves are the theoretical predictions for J and J_0 in the WLC model¹⁰ which perfectly match the simulation data. In the case of the AER model, the dashed curves are shown as eye-guides. As can be seen at short lengths (below 100 bp), the J-factor in the AER model (with or without axial alignment) is several orders of magnitude larger than in the WLC model. As expected, the difference between the two models decreases as the DNA length increases, and for lengths larger than the DNA persistence length (~ 150 bp), the models are essentially indistinguishable. The same result can be obtained by other kinkable models.^{15,19} In this study, we showed that the asymmetry in DNA structure may promote the kink formation, in particular, largely increases the J-factor at short length scales. As was discussed, the torsional constraint is not considered for both looping probabilities J and J_0 , which leads to oscillations of the J-factor as a function of DNA length with a period equal to the DNA helical pitch.¹⁰

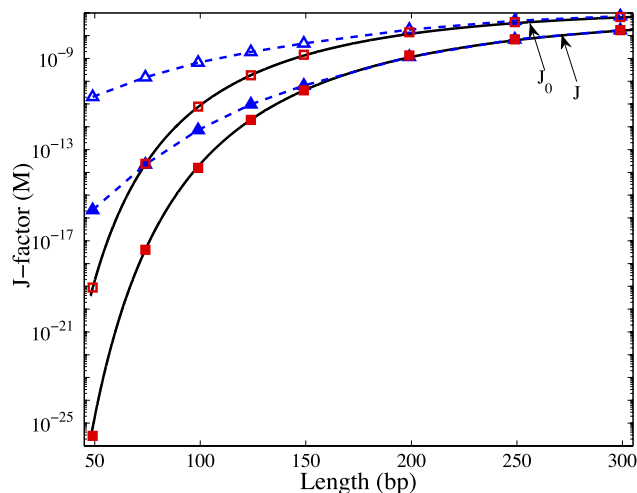


FIG. 8. The J-factor as a function of DNA length. MC result for models “A” and “W” are represented by triangles (blue) and squares (red), respectively, where filled and open markers correspond to the cyclization with (J) and without (J_0) axial alignment. Shimada-Yamakawa’s theoretical predictions for the J-factor in the model “W” have been also shown by solid (black) curves.¹⁰ Dashed (blue) curves show the trend of the simulation data and do not correspond to a theoretical model. Error bars (not shown) are about the size of the markers.

Recent experimental data of the J-factor of DNA molecules have shown that short DNA molecules are much more cyclizable than the prediction of the WLC model (Vafabakhsh and Ha⁸). In this experiment, the DNA probe was a duplex with two complementary single-stranded overhangs on both ends (two sticky ends). Because the single-stranded overhangs are 10 nucleotides, they are considered as long sticky ends. It is expected that joining of such long sticky ends does not

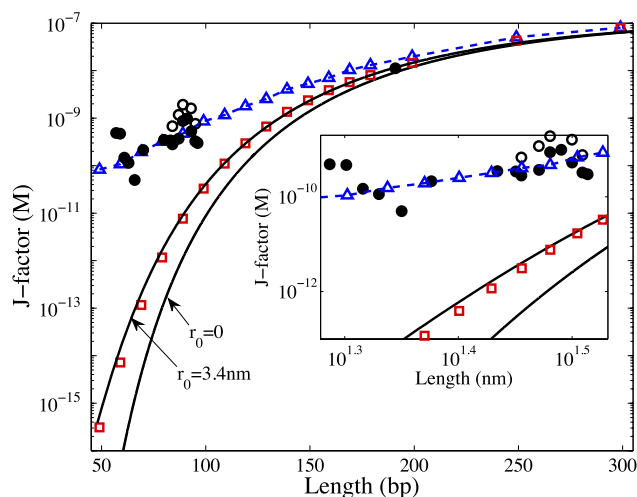


FIG. 9. J-factor comparison. The triangles (blue) and squares (red) correspond to the simulation data of the models “A” and “W,” respectively, for arbitrary angle between the duplex ends and the capture radius $r_0 = 3.4$ nm ($=10$ bp). Circles show the experimental data of Vafabakhsh and Ha for surface tethered (filled) and vesicle encapsitated (open) experiments.⁸ For consistency with our simulation data, we shifted the original data of the sticky ends by 10 bp to the left. The solid (black) curves are the theoretical predictions of the WLC model for $r_0 = 0$ and 3.4 nm.⁴⁴ Inset represents the length range of 50–100 bp in full logarithmic plot. Dashed (blue) curves show the trend of the simulation data and do not correspond to a theoretical model. Error bars (not shown) are about the size of the markers.

require the axial alignment of the duplex ends,¹³ and the effect of the torsional alignment could be considered as an oscillation factor in the J-factor. Also they can join each other when the end-to-end distance of the duplex is less than the capture radius, r_0 , which is 10 bp in this experiment. We thus evaluate the J-factor with free boundary conditions and 3.4 nm capture radius for both parameter models, i.e., models “A” and “W.” As Figure 9 shows, while the experimental data significantly deviate from the prediction of the WLC model for short DNA molecules, they show a considerable agreement with the AER model at all length scales. The oscillations in experimental data are believed to result from the torsional alignment between the DNA ends. The length dependence of the J-factor in the length range of 57–96 bp is much weaker in the AER model compared to the WLC model. The inset of Figure 9 shows a zoomed view in the length range of 50–100 bp in full logarithmic plot. To quantify the length dependence, we fit power law functions to the simulations data points in this range and found $J_0^A \sim L^{3.9}$ and $J_0^W \sim L^{16}$, where the superscripts indicate the model parameters “A” and “W.” In this issue, it has been suggested that underlying mechanism in the case of surface tethered may increase the rate of cyclization,^{46,47} but this effect cannot explain the anomalous behavior of DNA. Other kinkable model shows a sharp deviation from the WLC model at a critical length,⁴⁸ but the AER model shows a smooth deviation (see Figures 8 and 9).

IV. CONCLUSION

In this paper, we proposed that the asymmetric structure of DNA can significantly affect the elasticity of DNA at short length scales. To account for the bending asymmetry, we exploited the AER model, which was introduced and parametrized in a previous study.³² The kinks in the AER model are allowed to form in the minor groove direction, unlike the other kinkable models.^{15,18} By evaluating the effective bending energy and the distribution function of the end-to-end distance, we show that although the AER model is equivalent to a WLC model at large length scales, for tightly bent short DNA molecules, the DNA is much more flexible in the AER model than in the WLC model. Using the umbrella sampling method, we evaluated the loop formation probability, i.e., the J-factor, as a function of the DNA length. We found that the unconstrained J-factor in the AER model with capture radius about 3.4 nm is in excellent agreement with the measured experimental data presented in Ref. 8 at all length scales. This implies that the axial alignment of the two ends is not required to join the two juxtaposed DNA ends in this experiment. Enforcing an axial alignment can induce a 1000-fold change in the J-factor (see Figure 7). This may explain the large dispersity in the experimental data where DNA molecules with short sticky ends are cyclized.^{4,5,11} The results presented in this paper suggest that the asymmetric elastic rod model, as parametrized in Ref. 32, is a realistic model to explain the elastic behavior of the DNA double helix at short length scales.

ACKNOWLEDGMENTS

We thank Sayeh Rajabi for intently reading the manuscript and helpful grammar comments.

- ¹J. F. Marko and E. D. Siggia, *Macromolecules* **27**, 981 (1994).
- ²K. B. Towles, J. F. Beausang, H. G. Garcia, R. Phillips, and P. C. Nelson, *Phys. Biol.* **6**, 025001 (2009).
- ³J. F. Marko and E. D. Siggia, *Macromolecules* **28**, 8759 (1995).
- ⁴T. E. Cloutier and J. Widom, *Mol. Cell* **14**, 355 (2004).
- ⁵T. E. Cloutier and J. Widom, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3645 (2005).
- ⁶P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, *Nat. Nanotechnol.* **1**, 137 (2006).
- ⁷C. Yuan, H. Chen, X. W. Lou, and L. A. Archer, *Phys. Rev. Lett.* **100**, 018102 (2008).
- ⁸R. Vafabakhsh and T. Ha, *Science* **337**, 1097 (2012).
- ⁹A. Noy and R. Golestanian, *Phys. Rev. Lett.* **109**, 228101 (2012).
- ¹⁰J. Shimada and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- ¹¹Q. Du, A. Kotlyar, and A. Vologodskii, *Nucleic Acids Res.* **36**, 1120 (2008).
- ¹²Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskiaia, and A. Vologodskii, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5397 (2005).
- ¹³A. Vologodskii, Q. Du, and M. D. Frank-Kamenetskii, *Artif. DNA, PNA, XNA* **4**, 1 (2013).
- ¹⁴R. M. Harrison, F. Romano, T. E. Ouldridge, A. A. Louis, and J. P. Doye, preprint [arXiv:1506.09008](https://arxiv.org/abs/1506.09008) (2015).
- ¹⁵A. Vologodskii and M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **41**, 6785 (2013).
- ¹⁶P. Ranjith and G. I. Menon, *Biophys. J.* **104**, 463 (2013).
- ¹⁷X. Xu, B. J. R. Thio, and J. Cao, *J. Phys. Chem. Lett.* **5**, 2868 (2014).
- ¹⁸P. A. Wiggins, R. Phillips, and P. C. Nelson, *Phys. Rev. E* **71**, 021909 (2005).
- ¹⁹P. A. Wiggins and P. C. Nelson, *Phys. Rev. E* **73**, 031906 (2006).
- ²⁰J. Yan and J. F. Marko, *Phys. Rev. Lett.* **93**, 108108 (2004).
- ²¹P. Ranjith, P. B. S. Kumar, and G. I. Menon, *Phys. Rev. Lett.* **94**, 138102 (2005).
- ²²B. Eslami-Mossallam and M. R. Ejtehadi, *J. Chem. Phys.* **128**, 125106 (2008).
- ²³D. Norouzi, F. Mohammad-Rafiee, and R. Golestanian, *Phys. Rev. Lett.* **101**, 168103 (2008).
- ²⁴J. D. Moroz and P. C. Nelson, *Macromolecules* **31**, 6333 (1998).
- ²⁵G. Bijani, N. H. Radja, F. Mohammad-Rafiee, and M. Ejtehadi, preprint [arXiv:cond-mat/0605086](https://arxiv.org/abs/cond-mat/0605086) (2006).
- ²⁶A. G. Cherstvy, *J. Phys. Chem. B* **115**, 4286 (2011).
- ²⁷A. Fathizadeh, B. Eslami-Mossallam, and M. Ejtehadi, *Phys. Rev. E* **86**, 051907 (2012).
- ²⁸C. R. Calladine and H. R. Drew, *Understanding DNA* (Academic Press, Cambridge, 1999).
- ²⁹F. H. C. Crick and A. Klug, *Nature* **255**, 530 (1975).
- ³⁰T. J. Richmond and C. A. Davey, *Nature* **423**, 145 (2003).
- ³¹F. Lankas, R. Lavery, and J. H. Maddocks, *Structure* **14**, 1527 (2006).
- ³²B. Eslami-Mossallam and M. R. Ejtehadi, *Phys. Rev. E* **80**, 011919 (2009).
- ³³B. Mergell, M. R. Ejtehadi, and R. Everaers, *Phys. Rev. E* **68**, 021911 (2003).
- ³⁴R. E. Dickerson, *Nucleic Acids Res.* **26**, 1906 (1998).
- ³⁵W. K. Olson, A. A. Gorin, X. Lu, L. M. Hock, and V. B. Zhurkin, *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11163 (1998).
- ³⁶N. Destainville, M. Manghi, and J. Palmeri, *Biophys. J.* **96**, 4464 (2009).
- ³⁷P. J. Flory, U. W. Suter, and M. Mutter, *J. Am. Chem. Soc.* **98**, 5733–5739 (1976).
- ³⁸A. A. Podtelezhnikov and A. Vologodskii, *Macromolecules* **33**, 2767 (2000).
- ³⁹S. Mehraeen, B. Sudhanshu, E. F. Koslover, and A. J. Spakowitz, *Phys. Rev. E* **77**, 24 (2008).
- ⁴⁰G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- ⁴¹S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comput. Chem.* **13**, 1011 (1992).
- ⁴²J. Curuksu, K. Zakrzewska, and M. Zacharias, *Nucleic Acids Res.* **36**, 2268 (2008).
- ⁴³J. Samuel and S. Sinha, *Phys. Rev. E* **66**, 050801 (2002).
- ⁴⁴N. Douarache and S. Cocco, *Phys. Rev. E* **72**, 061902 (2005).
- ⁴⁵J. D. Kahn and D. M. Crothers, *J. Mol. Biol.* **276**, 287 (1998).
- ⁴⁶J. T. Waters and H. D. Kim, *Macromolecules* **46**, 6659 (2013).
- ⁴⁷T. T. Le and H. D. Kim, *Biophys. J.* **104**, 2068 (2013).
- ⁴⁸T. T. Le and H. D. Kim, *Nucleic Acids Res.* **42**, 10786 (2014).