View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UT Digital Repository

PHYSICAL REVIEW E **93**, 022411 (2016)

Competition between B-Z and B-L transitions in a single DNA molecule: Computational studies

Ah-Young Kwon,[1] Gi-Moon Nam,[2] Albert Johner,[1,3] Seyong Kim,[1] Seok-Cheol Hong,[4,5] and Nam-Kyung Lee[1,3]
[1]*Institute of Fundamental Physics, Department of Physics, Sejong University, Seoul 143-743, Korea*
[2]*Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA*
[3]*Institute Charles Sadron, Université de Strasbourg, CNRS UPR22, 23 rue du Loess 67034, Strasbourg cedex 2, France*
[4]*Center for Molecular Spectroscopy and Dynamics, Institute for Basic Science (IBS), Seoul 02841, Korea*
[5]*Department of Physics, Korea University, Seoul 02841, Korea*
(Received 7 June 2015; revised manuscript received 1 December 2015; published 19 February 2016)

Under negative torsion, DNA adopts left-handed helical forms, such as Z-DNA and L-DNA. Using the random copolymer model developed for a wormlike chain, we represent a single DNA molecule with structural heterogeneity as a helical chain consisting of monomers which can be characterized by different helical senses and pitches. By Monte Carlo simulation, where we take into account bending and twist fluctuations explicitly, we study sequence dependence of B-Z transitions under torsional stress and tension focusing on the interaction with B-L transitions. We consider core sequences, $(GC)_n$ repeats or $(TG)_n$ repeats, which can interconvert between the right-handed B form and the left-handed Z form, imbedded in a random sequence, which can convert to left-handed L form with different (tension dependent) helical pitch. We show that Z-DNA formation from the $(GC)_n$ sequence is always supported by unwinding torsional stress but Z-DNA formation from the $(TG)_n$ sequence, which are more costly to convert but numerous, can be strongly influenced by the quenched disorder in the surrounding random sequence.

DOI: 10.1103/PhysRevE.93.022411

I. INTRODUCTION

Since the double helical structure of DNA has been discovered by Watson and Crick, the importance of DNA structures has also been recognized in relation to its cellular functions. DNA is a highly malleable molecule [1] and exploits its various forms during the replication and transcription processes. In contrast to standard B-DNA which has the right-handed helical form, Z-DNA with a zigzag backbone has the left-handed helical form. During the transcription process, torsional stress of negative superhelicity formed behind the R loop can be relaxed by forming such left-handed helical structures in DNA. Z-DNA can be formed at DNA sections where specific sequences such as repeats of (GC) or (TG) dinucleotides are located [2].

In vitro, Z-DNA can be stabilized by torsional stress that unwinds DNA (like in magnetic tweezers experiments). Z-DNA can be also favored by high tension, by high salt (molar range) and also by adsorption of Z-DNA binding proteins (ZBPs), such as ADAR-1 [3]. At moderate salt condition ($\sim$100 mM), the free energy cost for switching one base pair (bp) to Z-DNA is known as $\sim 0.51 k_B T$ (0.3 kcal/mol) for $(GC)_n$ repeats and $\sim 1.2 k_B T$ (0.67 kcal/mol) for $(TG)_n$ repeats [4]. It is more costly to convert $(TG)_n$ repeats into Z form than $(GC)_n$ repeats. Short Z-DNA forming sequences of $(GC)_n$ or $(TG)_n$ adopt Z form in the presence of ZBPs such as ADAR-1 and typically the concentration of ZBPs required for the transition in $(TG)_n$ repeats is more than that of $(GC)_n$ repeats [3].

While Z-DNA can be formed at specific sequences, more commonly a left-handed helical form exists called L-DNA [5]. In principle, L-DNA can be formed from any sequence, but practically it is formed at AT rich sequences. Negative torsional stress can be released by opening the right-handed helical structure of standard B-DNA. The separated strands resulting from denaturation further twist to form L-DNA. In contrast to the Z form which has a well defined structure (Watson Crick pairs), helical pitch and length, the structure of L-DNA is ill-defined due to the lack of hydrogen bondings. Whether it is in natural DNA or in a DNA prepared for *in vitro* assay, a Z-DNA forming sequence is surrounded by random sequences which may switch to L-DNA under negative torsional stress.

Due to the development of single molecule experiments, it is possible to apply torsional stresses directly to a DNA molecule and to measure various characteristics of the B-Z transitions. Lee *et al.* investigated the kinetics and energetics of Z-DNA [6] using the hybrid technique of magnetic tweezers and single molecule FRET (smFRET). More recently, in a series of papers, Bryant's group reported torque measurements on a single DNA molecule upon applying negative superhelical stress [7–10]. Both B-L transition and B-Z transition are cooperative due to the similarly large domain wall energy E_J which is the penalty for breaking DNA stacking in B-DNA. However, the critical torques measured for the structural transitions are quite different. For a molecule bearing an insert of Z-DNA forming sequences $(GC)_n$, the torque remains $\sim -(2–3)$ pN nm during the B-Z transition. For the random sequence of 50% (A · T) contents, the critical torque for the B-L transition is measured to be $-(10–12)$ pN nm (150 mM salt). These values of critical torque reflect the free energy cost per unit twist angle accompanied by structural changes. The torque experiments imply that these two transitions are clearly separated for DNA of tens of kilo base pairs.

Recent experiments [11] report that the elastic response of the B-Z transition in a $(TG)_n$ sequence is qualitatively different from that in a $(GC)_n$ sequence: at moderate tension, Z-DNA from $(GC)_n$ shows a steplike transition with increasing torsion but $(TG)_n$ repeats undergo the B-Z transition over a broader range of torsions. This suggests that there is competition between Z-DNA and L-DNA in a certain range of unwinding torsion and such competition is more relevant for a $(TG)_n$

 ©2016 American Physical Society

sequence. Furthermore, while the transition torque of $(GC)_n$ repeats is well established, the critical torque measurements for $(TG)_n$ repeats have not been successful up to date. For $(TG)_n$ repeats, it seems to be difficult to isolate the B-Z transition from the B-L transition and thus to determine the critical torque for the B-Z transition. Both Z-DNA and L-DNA are favored over B-DNA under large tension. The rise of a base pair in Z form, a_Z, is slightly longer than that in B form, a_B. In general, L-DNA can be more elongated than B-DNA by tension, which makes it more favorable at large tension as seen from the twist-extension relation [12–14]. Since the total linking number should be conserved, there is obvious coupling between the B-Z and B-L transitions. The sequence dependent B-L transition, hence, will influence the Z-DNA formation under given unwinding conditions in a nontrivial way. We thus perform computational studies to elucidate the stability of Z-DNA in relation to B-L transitions.

Recently we have introduced a copolymer model to describe the B-Z transitions [15] for a chain consisting of switchable monomers between two states. The essential quantities for characterizing Z-DNA and B-DNA are their intrinsic twist rates, t_Z and t_B, together with twist fluctuations around them. The model thus accounts for elastic twist responses during the B-Z transition related to the annealed structural heterogeneity involving interconversion between B-DNA and Z-DNA states.

In this paper, we apply this model [15] for a numerical study of the system comprising Z-DNA and L-DNA forming sequences. Although the helical pitch is not strictly defined for L-DNA, the effective pitch at a given tension can be described using the two strand model [16]. One of the main parameters is the free energy cost required to convert one monomer from B- to L-DNA. The unpairing energy of the base pair strongly depends on the salt condition and the type of nearest neighbor bases [17]. The unpairing energy is 0.43 kcal/mol/bp for d(AA) [(AA) with complementary sequence of (TT)]. In contrast, the binding energy is 0.27 kcal/mol/bp for d(AT) at low salt conditions (∼20 mM). At intermediate salt conditions (∼75 mM), the unpairing energy is in the range 0.6–0.7 kcal/mol. At higher salt conditions (∼1 M), the binding energy is as large as ∼1 kcal/mol. The G · C pair binding energy is found in the range 1.6–2.0 kcal/mol. We consider heterogeneous values of binding energy ϵ_L along the sequence, corresponding to a type of base pair. We also consider different values of $u_{\rm AT}$ to take into account the shift of the unpairing energy depending on salt conditions as described above.

Using the Wang and Landau Monte Carlo method [18], the equilibrium conformational statistics of DNA is obtained via the calculation of the density of states in two dimensional phase space of energy E and extra linking number δLk. Torsional stress is imposed by sampling the conformations falling in the range of the desired linking number δLk. We first check that our model properly describes the reported torsional responses of DNA undergoing the B-Z transitions [6,8,12,13,19]. Our main efforts are dedicated to the investigation of B-Z transitions for $(TG)_n$ and $(GC)_n$ repeats interacting with L-DNA forming sequences. We show that the stability of the Z-DNA from $(TG)_n$ repeat can be strongly influenced by the sequence of the surrounding random DNA with quenched disorder.

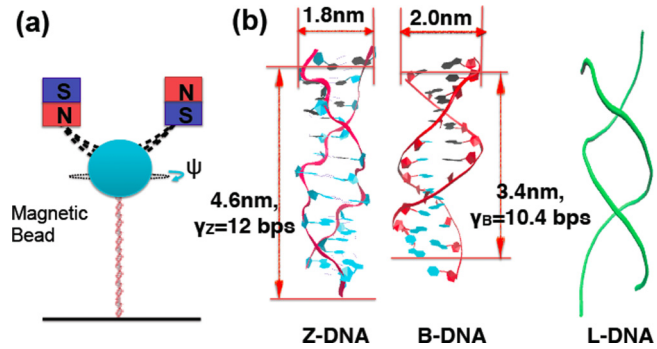

FIG. 1. (a) Schematic representation of magnetic tweezers experiments applying the torsion angle ψ. (b) Z-DNA, B-DNA, and L-DNA. The number of monomers in a helical pitch γ_B, γ_Z, and γ_L are different.

II. MODEL

We consider a single DNA molecule of length S stretched under tension f above the supercoiling threshold [Fig. 1(a)]. The molecule is also subject to prescribed excess turns $\delta\mathrm{Lk} = \psi/(2\pi)$. Under the torsional stress, the torsional response of the DNA chain can be well described using the wormlike-chain (WLC) model [20] where the DNA's elastic properties are defined through the bending and twist persistence lengths, $l_p = B/k_{\rm B}T$ and $l_t = C/k_{\rm B}T$ with the local curvature κ and the twist rate $\tau = d\psi/ds$. In order to take into account the structural transition in DNA, we extend the WLC model for a copolymer consisting of several types of monomers with the following Hamiltonian up to the junction energy [see Eq. (6)]:

$$\frac{E}{k_{\rm B}T} = \int_0^S ds \left\{ \frac{1}{2} l_{p,\nu}\kappa^2 + \frac{1}{2} l_{t,\nu} \left(\frac{d\psi}{ds} - \frac{t_i(s)}{a_\nu} \right)^2 + \frac{\epsilon_\nu}{a_\nu} \right\}, \quad (1)$$

where the index "ν" indicates the state of monomers (e.g., B, Z, and L) which varies along the chain. The monomers in the core sequence (Z core) can be in two different states: B-DNA and Z-DNA. These monomer states have different intrinsic twist rate t_ν, free energy ϵ_ν, and monomer extension a_ν, depending on the index of state $\nu \in \{B,Z\}$. Similarly a random sequence can be in two different states: B form or L form, $\nu \in \{B,L\}$. If the core sequence is a Z-DNA forming sequence, $(GC)_n$ or $(TG)_n$, we used $\epsilon_Z = 0.51k_{\rm B}T$ for $(GC)_n$ and $\epsilon_Z = 1.2k_{\rm B}T$ for $(TG)_n$. For L-DNA, we choose ϵ_L as the unpairing energy of a bp. For an (A · T) bp, we use $u_{\rm AT} = 0.6k_{\rm B}T$ and $1.0k_{\rm B}T$, which are typical values of the unpairing energy at the monovalent ion concentrations of ∼20 and ∼100 mM, respectively [17]. The former is comparable to the experimental condition in [21]. For a G · C bp, we use $u_{\rm GC} = 2.4k_{\rm B}T$ (1.4 kcal/mol).

For B-DNA, the number of base pairs per helical turn is $\gamma_B = 10.5$ and the intrinsic twist rate is $t_B = 2\pi/\gamma_B$. For Z-DNA, $\gamma_Z = 12$ and $t_Z = -2\pi/\gamma_Z$ [Fig. 1(b)]. The positive (negative) sign of the twist rate implies a right-handed (left-handed) chirality. The stacking distance between neighboring base pairs defines the monomer sizes a_Z,a_L. Note that the tension affects the free energy cost of monomer switching because each state has different elongation. The Gibbs free energy of the corresponding monomer state is measured with

respect to the B-DNA state and rescales by tension as

$$\epsilon'_{Z(L)} = \epsilon_{Z(L)} - \mathbf{f} \cdot [\mathbf{a}_{Z(L)}(f) - \mathbf{a}_B]. \tag{2}$$

Considering that $a_Z \approx 0.38$ nm and $a_B \approx 0.33$ nm, the cost ϵ'_Z for switching a bp from B-DNA to Z-DNA is expected to become negative in the tension regime larger than ~40 pN, which is yet far below the tension required for the B-S transition (~60 pN). Here, we discuss the B-Z transition in the tension regime below any other transitions.

Although the helical pitch and monomer size for L-DNA are not as definite as those of Z-DNA, it is yet possible to describe L-DNA as a helical ply if a large enough tension is applied [5,16,22]. The twist rate $t_L = -2\pi/\gamma_L$ and the monomer length a_L of L-DNA is expected to vary with tension and a similar energy rescaling can be expected. At typical tensions considered, 8 pN, the shift in ϵ'_L turns out to be negligible and the number of monomers in a helical pitch in L-DNA is similar to γ_B. In the large tension limit, the helical pitch of L form formally diverges.

In this paper, we assume that a sufficiently large tension is applied to the chain so that it is almost stretched along the direction of tension. The upper limit of the critical torque for supercoiling under given tension f is $\mathcal{T}_{PL} = \frac{4}{\pi}\sqrt{l_L k_B T f}$ [16,23–27] and is dictated by the small persistence length of L-DNA, $l_L \approx 3$ nm. The estimates of the force preventing supercoiling at the critical torque of the B-L transition (−12 pN nm) is at least $f_L^{sc} \approx 8$ pN. For force below f_L^{sc}, the L-DNA is expected to be supercoiled upon the B-L transition.

Proceeding to MC simulation of B-L and B-Z transitions, we estimate the tension dependent twist rate of L-DNA using the L-DNA model developed from our previous study [16]. We use these estimates as the references to choose the valid input parameters (γ_L and a_L) for simulation of the L-DNA state. Assuming the geometry of L-DNA to be a pair of interwinding single strands [Fig. 1(b)], we compute the number of monomers in a helical pitch γ_L and the spacing between bps, a_L, along the direction of the force. The central parameter is the helical angle α of this vertical ply. The optimal size of L-DNA and helical angle α are obtained self-consistently by minimizing the free energy of the DNA where we prescribe the mean value of unpairing energy per bp to be $\tilde{\epsilon}_L$. Note that $\tilde{\epsilon}_L$ is an average free energy cost per bp for "local" section which converts to L-DNA. The details of the model are described in Appendix A and B and Ref. [16]. From the simple model, a rough estimate for the critical torque for the B-Z and B-L transitions can be made as

$$\tau^*_{Z(L)} = -\frac{\tilde{\epsilon}'_{Z(L)}}{\Delta\psi_{Z(L)}} \tag{3}$$

where we define $\Delta\psi_{Z(L)} \equiv 2\pi(\gamma_B^{-1} + \gamma_{Z(L)}^{-1})$ which corresponds to the change in the twist angle due to a monomer switching from B-DNA to Z-DNA (L-DNA). We also define $\gamma'^{-1}_{Z(L)} \equiv \Delta\psi_{Z(L)}/(2\pi)$ as the change in linking number due to a monomer switching.

Table I shows some representative values of γ_L, γ'_L, and a_L for various $\tilde{\epsilon}_L$ under applied tensions of $f = 13$ and 30 pN. Each strand of L-DNA consists of n_d nucleotides of size $b_0 = 0.5$ nm [28]. Fluctuations present in two interwinding single strands can be taken into account by rescaling the nucleotide size b_0 by the ratio between single strand extension (z_L) and length (S_L) under the given tension f, such that $b = b_0 z_L/S_L$. The ratio, z_L/S_L, is obtained using the force-extension relation of WLC. Here, we set the parameter v accounting for the excluded volume repulsion between the two single strands of L-DNA to be $v = 0.9k_BT$, which reproduces the realistic values of helical pitches of L-DNA as previously discussed in [16] (see Appendix B). At moderate tension, 10 pN $< f <$ 20 pN, the helical pitch of L-DNA is in the range similar to B-DNA, $\gamma_L \approx 9$–11, and at large tension, two L-DNA strands are almost straight. For an alternating (A · T)/(G · C) sequence ($\tilde{\epsilon}_L \approx 1.5)k_BT$, γ_L is 9–11 in the tension below 20 pN. Under the condition where single strand fluctuations are negligible ($b \approx b_0$), the interwinding helix is more straightened by tension. At 30 pN, $\gamma_L \approx 20$ with $\tilde{\epsilon}_L = 0.6k_BT$. In the regime of straightened L-DNA, the change of twist due to monomer state switching is mainly from unwinding of B-DNA. This leads to $\Delta\psi_L \approx 2\pi/\gamma_B$ for large tension. Hence $\Delta\psi_L$ is only half of the largest twist rate as $\gamma_L \to \infty$ [Eq. (3)].

TABLE I. The rise of a monomer a_L, helical pitch γ_L, and γ'_L for L-DNA sequences with various mean values of free energies $\tilde{\epsilon}_L$ with $b = b_0 z_L/S_L$. These values are computed from Eq. (B2) with $v = 0.9k_BT$ for two different tensions, $f = 13$ pN and $f = 30$ pN (see Appendix B). The lower lines are for the condition that single strands in L-DNA are fully stretched ($b = b_0$).

		a_L (nm)		γ_L		$\gamma'_L = \frac{2\pi}{\Delta\psi_L}$	
$\tilde{\epsilon}_L(k_BT) \setminus f$(pN)		13	30	13	30	13	30
	0.6	0.36	0.41	10.9	15.4	5.4	6.2
	0.9	0.35	0.41	10.4	13.3	5.2	5.9
$b < b_0$	1.2	0.34	0.40	10.1	12.3	5.2	5.7
	1.5	0.34	0.39	9.8	11.4	5.1	5.5
	2.4	0.33	0.38	9.3	10.2	4.9	5.2
	0.6	0.46	0.49	11.7	21	5.5	7.0
$b = b_0$	1.5	0.45	0.48	10.7	15.3	5.3	6.2
	2.4	0.44	0.47	10.1	13.0	5.2	5.8

While the helical pitch increases with tension, the effective free energy cost ϵ' for monomer switching decreases according to Eq. (2). Due to these competing influences, the critical torque is a slow varying function for some range of tension ($f < 25$ pN) which is consistent with experimental studies [5,7]. A virtually flat insensitive dependence of the critical torque on the applied tension can be obtained with an *ad hoc*, slightly small, parameter (i.e., $b_0 = 0.45$ nm), which compensates some weakness of the WLC model for single strands. Such a change, however, has little influence on the central discussion of the competition between B-Z and B-L transitions.

In Fig. 2, we plot critical torques for B-L and B-Z transitions, τ^*_L and τ^*_Z, as a function of tension f. We note that τ^*_L exceeds τ^*_Z of $(TG)_n$ sequence ($\epsilon_Z \approx 1.2k_BT$) if $\tilde{\epsilon}_L$ is large enough. This implies that Z-DNA and L-DNA population can be influenced by tension and more significantly by sequence compositions for $(TG)_n$ sequences. Due to the quenched disorder in the sequence, the free energy cost per monomer switching ϵ_L generally deviates from its average.

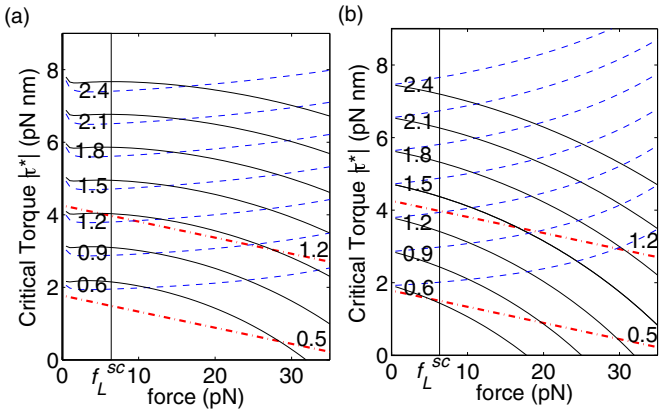

FIG. 2. Critical torques of B-Z and B-L transitions computed for various tensions f are shown for (a) $b < b_0 = 0.5$ nm and (b) for $b = b_0$, the completely straightened chain. The dot-dashed (red) lines are for the B-Z transitions with $\epsilon_Z = 0.5$ and $1.2k_BT$ where the critical torques are linearly decreasing. The black solid lines are the critical torques for B-L transitions with various $\tilde{\epsilon}_L$ computed by Eq. (3) and the blue dotted lines are the estimates without tension correction in $\tilde{\epsilon}_L$. The corresponding values of $\tilde{\epsilon}_L$ are indicated along with the solid lines. The curves on the left of $f_L^{sc} \approx 8$ pN are not accurate due to supercoiling.

The nucleation of L-DNA occurs at the longest (A · T) region with smallest $\tilde{\epsilon}_L$, for which the critical torque is less than τ_Z^*. The sequence dependence will be further exploited in a later part of this paper.

In the following simulations, we first model different states of DNA with various values of helical twist angles and switching energies without explicitly referring to changes in tension. Since ϵ_L depends on various parameters such as salt concentrations, pH of solution, and sequence composition, ϵ_L and γ_L are treated independently. In the second part of the simulation (Fig. 10), we rescale the tension dependent parameters ϵ_L and γ_L simultaneously under given tension to investigate the influence of change of monomer size. Our model assumes that all bond vectors are unit vectors and thus we do not distinguish the variations of monomer size by state switching. However, we incorporate the influence of tension and the variation of monomer size by considering the aforementioned free energy rescaling [Eq. (2)].

III. SIMULATION MODEL

For numerical studies, we model DNA as a wormlike chain (WLC) consisting of N heterogeneous monomers. The position vectors of monomers are represented by $\mathbf{r}_i$ ($i = 0,1,2,\ldots N$) and the bond vectors $\mathbf{u}_i$ are defined as $\mathbf{u}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$. In order to take into account the twist of the centerline, we use three orthonormal vectors for each bond i: $\{\mathbf{f}_i,\mathbf{v}_i,\mathbf{u}_i\}$ [29]. The other unit vectors $\mathbf{f}$ and $\mathbf{v}$ point two perpendicular directions to the tangent and $\mathbf{u}_i = \mathbf{f}_i \times \mathbf{v}_i$. The strain tensor $\Omega_\mathbf{i} = (\Omega_{1,i},\Omega_{2,i},\Omega_{3,i})$ is defined for each body frame as $\Omega_{1,i} = (\mathbf{u}_{i+1} - \mathbf{u}_i)\cdot\mathbf{f}_i$, $\Omega_{2,i} = (\mathbf{u}_{i+1} - \mathbf{u}_i)\cdot\mathbf{v}_i$, $\Omega_{3,i} = (\mathbf{v}_{i+1} - \mathbf{v}_i)\cdot\mathbf{f}_i$. The elastic energy of the WLC [20] is written as a function of the local curvatures $\kappa = \sqrt{\Omega_1^2 + \Omega_2^2}$ and twist $\tau = \Omega_3$ as

$$E_{el} = \frac{1}{2}\sum_{r=1}^{N-2}\left\{l_{p,r}\left(\Omega_{1,r}^2 + \Omega_{2,r}^2\right) + l_{t,r}(\Omega_{3,r} - \omega_{3,r})^2\right\}. \quad (4)$$

For different monomer states, the preferred twists $\omega_{3,r}$ take the values of $t_B = 2\pi/\gamma_B$, $t_Z = -2\pi/\gamma_Z$, and $t_L = -2\pi/\gamma_L$ for B-DNA, Z-DNA, and L-DNA, respectively. During MC simulations, the monomer switching from type ν to μ involves changes in δLk and E such that

$$E \to E + (\epsilon_\mu - \epsilon_\nu), \quad \delta\mathrm{Lk} \to \delta\mathrm{Lk} + \frac{(\omega_{3,\mu} - \omega_{3,\nu})}{2\pi}.$$

During simulation, we only trace the relative twist with respect to the reference twist angle t_ν, and update the topological constraints according to the monomer switching. With $n_{Z(L)}$ monomers,

$$\delta\mathrm{Lk} = \mathrm{Wr} + \delta\mathrm{Tw} + n_{Z(L)}\frac{\Delta\psi_{Z(L)}}{2\pi} - n_J\frac{\phi_J}{2\pi}. \quad (5)$$

The last term applies only for B-Z junctions where an extra twist angle ϕ_J is located [30] and n_J stands for the number of B-Z junctions. We measure $\langle\delta\mathrm{Tw}\rangle$ as $\delta\mathrm{Tw} = \Sigma_r(\Omega_{3,r} - \omega_{3,r})$. The mean values of writhe number, $\langle\mathrm{Wr}\rangle$, are obtained by evaluating the Gauss linking integral along the contour of the molecule.

Although there is an indication that the modulus of Z-DNA is somewhat smaller than that of B-DNA, the modulus of Z-DNA is not yet established and even less is for L-DNA. In the study, we do not introduce heterogeneous twist modulus in local scales. For given tension, L-DNA twist rate γ_L is obtained under the equilibrium condition between L-DNA and B-DNA with a larger twist modulus. In this sense the L-DNA twist modulus is absorbed in γ_L. Within the current copolymer model, we do not access local fluctuations of stiffness but can address global twist fluctuations of the whole chain. Most of the single-molecule experiments use DNA molecules of at least several kilo base pairs long or core sequences flanked by such long DNA handles. The local change in stiffness (l_s) over a small segment of length s influences global fluctuations of the total chain of length S via harmonic average of the stiffness l such that $S/l' = (S-s)/l_t + s/l_s$[15]. Therefore the torsional stiffness of the L form does not significantly alter the global fluctuation of twist angle unless s is considerably large and the free energy is even less affected by the twist rigidity changes at equilibrium. We thus apply the same mechanical parameters l_p,l_t before and after the B-Z and B-L transition. For the calculation of elastic energy, we use the persistence lengths $l_p = 50$ nm and $l_t = 80$ nm.

We use the spin type variable s_i at each monomer state: $s_i = 1$ or -1, for B form or Z form (L form) state. The energy associated with the monomer state is written as

$$E_b = \sum_i \epsilon_i\frac{1}{2}(1 - s_i) + E_J\frac{1}{2}(1 - s_i s_{i+1}). \quad (6)$$

A junction energy of $E_J = 9k_BT$ [21] is assigned for each B-Z junction and B-L junction.

We obtain the equilibrium conformational statistics of DNA via calculation of the density of states, $g(E,\delta\mathrm{Lk})$ [18,20] in the two-dimensional phase space of E and δLk. At given δLk,

each conformation with energy E_r contributes with statistical weight of $\exp(-\beta E_r)$ and the average values of various physical quantities $x = \delta\mathrm{Tw}, \mathrm{Wr}, \ldots$ can be computed by

$$\langle x(\delta\mathrm{Lk})\rangle = \frac{\sum_E x(E,\delta\mathrm{Lk})g(E,\delta\mathrm{Lk})e^{-\beta E}}{\sum_E g(E,\delta\mathrm{Lk})e^{-\beta E}}. \quad (7)$$

Each Monte Carlo (MC) step includes attempts of (1) switching the type of monomer and (2) changing the conformation of the monomer. There are two types of MC moves for conformational changes: pivot and pure twist [20]. In each pivot move, a chosen monomer rotates by arbitrary angle with respect to the axis defined by two neighboring monomers. In a pure twist move, **f** and **v** are updated through rotation by random twist angle with respect to the local tangent **u**. Independent of conformational moves, the spin type variables of each monomer state are also updated for every MC step.

Each computation of $g(E,\delta\mathrm{Lk})$ of a system of typical size (100-mer) is performed with accuracy of the modification factor $f_{\mathrm{final}} = 1.000\,01$. Computations are performed on Intel Xenon CPU at 2.40 GHz, 8192 MB RAM running Linux, with parallel algorithm (∼100 nodes).

IV. RESULTS

A. Validation of the model: B-Z transition

We first apply our model to study the B-Z transition. The sequence of interest consists of N_c monomers of size a and each monomer corresponds to a base pair of DNA. Similar to the actual molecule of single molecule experiments, this specific sequence (core sequence) is inserted in natural random sequences, which serves as handles for magnetic tweezers experiments [see Fig. 3(a)]. We assume that the switching of monomer type occurs only in the core sequence and monomers in handles remain in B form. We first consider a core consisting of only Z-DNA forming sequences (Z core). In order to take into account the fluctuations of the whole chain including handles, for some simulations, we introduced a pair of handles of total size N_h. The total chain length is thus $S = (N_h + N_c)a$.

Throughout simulations, we apply a large enough tension ($> f_L^{\mathrm{sc}}$) to prevent any supercoiling. To assure that L-DNA is stretched, we not only subject the chain to the tension along the z direction but also confine the chain end in a harmonic potential in the radial direction so that the chain end stays along the axis of tension. Practically $f = 8$ pN ($f_{\mathrm{L}}^{\mathrm{sc}}$ with $\tau \sim -12$ pN nm) is sufficient to suppress supercoiling of L-DNA. We also measure the chain extension $\langle z\rangle \approx S$ and the average value of $\langle \mathrm{Wr}\rangle \approx 0$ to ensure the supercoiling is suppressed. The total excess linking number $\delta\mathrm{Lk}$ is thus distributed to the linking number change associated with the structural transition of the DNA and the twist of the center line, $\delta\mathrm{Tw}$. The average torque $\langle\tau\rangle$ under a given $\delta\mathrm{Lk}$ can be measured by computing $\delta\mathrm{Tw}$ using the relation $\tau \approx 2\pi(l_t/S)\delta\mathrm{Tw}$.

In Fig. 3(b), we plot the measured torque $\langle\tau\rangle$ and average value of $\langle n_Z\rangle$ as a function of excess linking number $\delta\mathrm{Lk}$. For small twist deformations, the torsional response reveals the linear response of the standard B-DNA. With higher levels of unwinding, the chain undergoes structural transitions and the torque $\langle\tau\rangle$ is approximately constant in the regime of the B-Z transition. This plateau value of the torque is referred to as the critical torque. We compare torque-twist relations for two different Z cores. The critical torque is measured to be -2 and -4 pN nm for $(\mathrm{GC})_n$ core and $(\mathrm{TG})_n$ core, respectively. The $\tau_Z^* \approx -2$ pN nm of $(\mathrm{GC})_n$ matches the reported value from experimental results [8,10]. The larger free energy ϵ_Z of $(\mathrm{TG})_n$ accounts for the large critical torque as predicted from Eqs. (3) and (A3).

If the status of end monomers are fixed as B-DNA (fixed boundary conditions), the formation of Z-DNA in the core accompanies at least two B-Z junctions. The magnitude of the transition torque (at the onset of the B-Z transition) ~ -10 pN nm is much larger than the critical torque. Such a large transition torque results from the large domain wall energy $E_J \sim 5.3$ kcal/mol (8–9$k_{\mathrm{B}}T$), which also accounts for

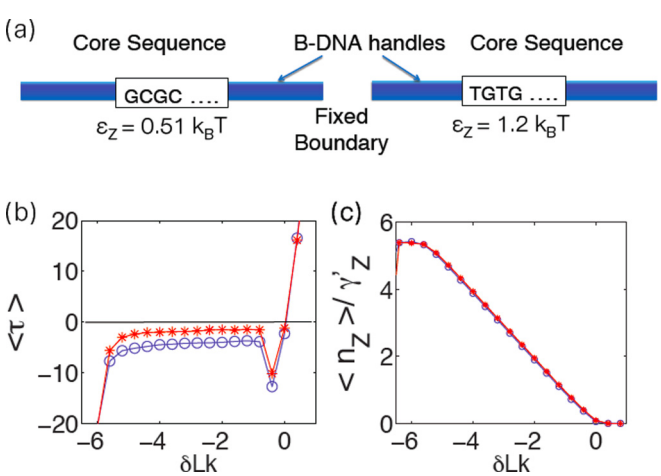

FIG. 3. (a) Z cores [GC type (⋆) and TG type (○)] flanked by B-DNA handles. (b) Torque-twist relation for short cores of length $N_c = 30$, $N_h = 70$ [$\epsilon_Z = 0.51k_{\mathrm{B}}T$ for $(\mathrm{GC})_{25}$ and $1.2k_{\mathrm{B}}T$ for $(\mathrm{TG})_{25}$] with the tension of $f = 8$ pN. (c) Linking numbers distributed to Z-DNA $\langle n_Z\rangle/\gamma'_Z$ vs $\delta\mathrm{Lk}$.

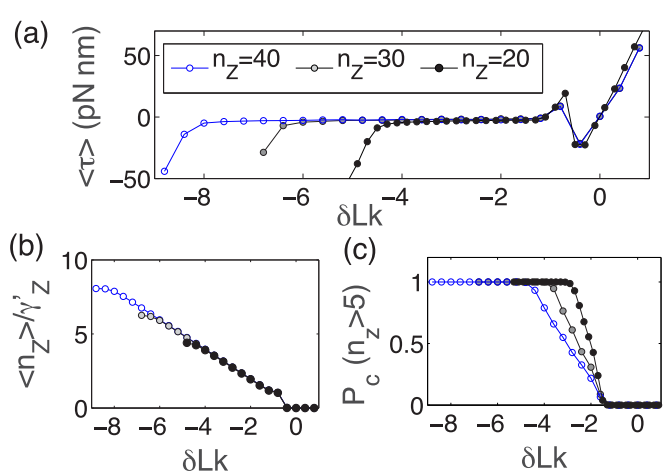

FIG. 4. B-Z transitions of $(\mathrm{GC})_n$ cores with various core sizes: $n = 10$, 15, and 20 ($N_c = 20$, 30, and 40) under the tension of $f = 8$ pN. The total chain length N_t is set to be 100 for all cases. (a) Torque-twist relation, (b) average values of n_Z, and (c) probability $P_c|_{n_Z>5}$ to form Z-DNA in the central core section.

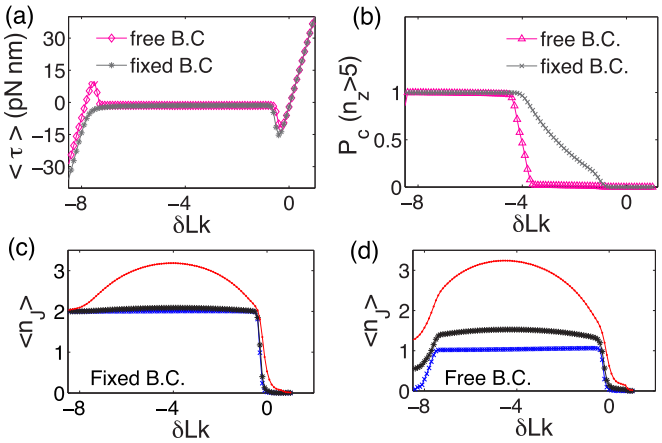

FIG. 5. Comparison between fixed and free boundary conditions for the chain with $(GC)_{22}$ core ($N_c = 44$). The shift in the junction angle is chosen to be $\phi_J = -0.1$ turn and the tension $f = 8$ pN. (a) $\langle\tau\rangle$ vs δLk and (b) the probability P_c of triggering Z-DNA ($n_Z > 5$) in the central part of the core. Panels (c) and (d) show the number of B-Z junctions $\langle n_J\rangle$ under fixed and free boundary conditions. Red, black, and blue lines correspond to $E/k_BT = 3$, 4.5, and 9 (from top to bottom).

high cooperativity of the B-Z transition. The mean number of monomers in the Z state $\langle n_Z\rangle$ [Fig. 3(c)] linearly increases with increasing unwinding torsional stress after nucleation. Figure 4 shows torque-twist relations for chains with various lengths of (GC) core, $N_c = 20$, 30, and 40, in a similar geometry ($N_t = 100$) as shown in Fig. 3(a).

For small $|\delta\text{Lk}|$, the torsional responses are linear and their slopes scale with l_t/S. The torque plateau indicating the B-Z transition extends in proportion to the lengths of core sequences. Similar to $N_c = 30$ [Fig. 3(c)], the average values of $\langle n_Z\rangle$ for all core lengths increase linearly with unwinding torsional stress after the initial nucleation and saturate at $\sim N_c/\gamma_Z'$. [The levels of saturation are indicated by dotted lines in Fig. 4(b).] At the transition, we observe a torque overshoot ($\sim$5–10 pN nm) towards positive torque. This results from the necessary countertwist to compensate for the abrupt twist angle shift at B-Z junctions $\phi_J = -2\pi\times 0.4$ over a single bond length. However, it turned out to be an artifact of choosing the junction size as a unit bond length and this overshoot disappears if the twist at a single bonded junction is chosen to be less than 0.2 turns [see, for example, Fig. 3(b) with $\phi_J = 0$, Fig. 5(a) with $\phi_J = -2\pi\times 0.1$, and Fig. 7(a) with $\phi_J = -2\pi\times 0.2$]. The experiments [8,9] do not observe such an overshoot at the start of the transition. The B-Z junction size is still under debate [31] but our numerical study and earlier theoretical study [15] suggest that the junction size should span at least over two bps to accommodate the junction twist of -0.4 turns, reported experimentally [32].

In single molecule experiments using smFRET [6], transition signals are captured by a pair of dye molecules which are tagged on the molecule several bps ($\sim$14) apart. The signal of transition relies on the extension of this central part. To make a connection with these FRET measurements [6], we compute the probability $P_c|_{n_Z>5}$ to find more than five central monomers in Z form [Fig. 4(c)]. When all core monomers switch to Z-DNA, $\langle n_Z\rangle$ saturates and the corresponding δLk value coincides with the end of the torque plateau. The longer the core sequence is, the more negative turns are required to reach $P_c|_{n_Z>5} = 1$. As expected, $P_c|_{n_Z>5}$ appears to be more cooperative than average values $\langle n_Z\rangle$. The overshoot in $\langle\tau\rangle$ (at the onset of transition) is more pronounced for small N.

Keeping in mind some experiments [33] using short oligomers consisting of Z-DNA, we consider a short core sequence [$(GC)_{22}$, $N_c = 44$] separated from handles. As the boundaries of the core permit free interconversion, no junction energy penalty applies at the boundaries. The junction energy penalty E_J applies if a B-Z junction locates in the middle of the chain but not at either end. The fixed boundary condition refers to a core sequence bounded by B-DNA. There are at least two junctions where some part of the core sequence is switched to Z-DNA.

In Fig. 5, we compare the results obtained from the same sequence but under different boundary conditions. A noticeable difference in the torque-twist relation is that there is a torque overshoot towards the end of the B-Z transition ($\delta\text{Lk} \sim -7$) for a chain under the free boundary conditions. As the torsional stress accumulates, the whole chain converts to Z-DNA removing the B-Z junction at the expense of converting the B-DNA section in the rest of the core to Z-DNA [Fig. 5(a)]. The probability of Z-DNA formation at the central section P_c is small under free boundary conditions, because Z-DNA domain forms near the ends rather than in the central section [Fig. 5(b)]. Figures 5(c) and 5(d) show the mean number of junctions $\langle n_J\rangle$ for two boundary conditions. Prescribing $E_J/k_BT = 9$, we find that in the coexistence regime $\langle n_J\rangle = 2$ under the fixed boundary conditions and $\langle n_J\rangle = 1$ under the free boundary conditions, respectively. The reduction of junction energy results in more than one Z-DNA domains with a peak of $\langle n_J\rangle \approx 3$ in the middle of the transition regime. These results are in good agreement with previous studies on the B-Z transitions [8,10,15]. We further apply the model to investigate sequence dependence of B-L transitions and B-Z transitions coupled with B-L transitions under unwinding torsional stress.

B. Sequence dependence: B-L transition

Next, we consider core sequences which interconvert between B-DNA and L-DNA, and thus are referred as L core below. Each sequence consists of two types of monomers, (A · T) and (G · C), distinguished by different free energy costs $\epsilon_L = u_{AT}$ and u_{GC}, respectively. In Fig. 6(a), we consider three different L-DNA sequences: (1) (A · T) only, (2) random sequence with 50% of (A · T), and (3) sequence of alternating (A · T)/(G · C). For all cases, we assume $\gamma_L = 20$. Similar to the B-Z transitions, the torque (measured from the center line twist of the chain) remains almost constant during the B-L transition. With $u_{AT} = 0.6k_BT$ and $u_{GC} = 2.4k_BT$, the measured critical torques for two sequences are ~ -2 pN nm for the type-(1) L core and ~ -5 pN nm for both type (2) and (3) L cores, respectively. The transition torque of the random sequence is about 2.5 times larger than that of the homo-(A · T) sequence reflecting the fact that the average binding energy of the random sequence $\tilde{\epsilon}_L = 1.5k_BT$ is 2.5 times larger than the used value of homo-(A · T) bp unpairing energy $\tilde{\epsilon}_L = 0.6k_BT$. The critical torque of an unbiased random sequence only slightly deviates from the average torque obtained from the

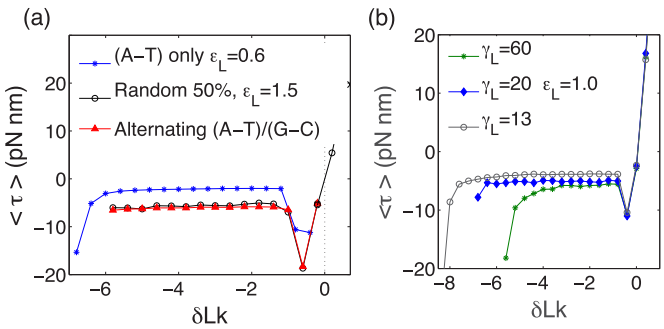

FIG. 6. (a) Torsional responses of a chain ($N = 135$) with an insert of different types of L-DNA forming sequences ($N_L = 45$): (A · T) only (blue star) and random sequence with 50% of (A · T) (black circle) and alternating (A · T)/(G · C) sequence (red triangle). $\gamma_L = 20$ for all three sequences. Here we assume $u_{\rm AT} = 0.6k_{\rm B}T$ and $u_{\rm GC} = 2.4k_{\rm B}T$. (b) Torsional responses of homo-(A · T) sequences of $u_{\rm AT} = 1.0k_{\rm B}T$ and $u_{\rm GC} = 2.4k_{\rm B}T$ with various values of $\gamma_L = 13$, 20, and 60. $N_L = 45$, $N_h = 55$. The tension is set to $f = 8$ pN.

alternating (A · T)/(G · C) sequence. The critical torque at the start of the transition is slightly smaller in magnitude for the random sequence. This reflects the fact that it is possible to find a facile sequence in a random L core to initiate the transition. For $u_{\rm AT} = 1.0k_{\rm B}T$, we observe the shift of the critical torque to more negative values. The critical torque is ~ -4 pN nm for sequence (1) [Fig. 6(b)] and $\sim$−(7–8) pN nm for sequences (2) and (3) (not shown).

The geometric parameters such as γ_L also influence the critical torque. To investigate the influence of γ_L on the critical torque [Eq. (3)], we measure the average torques of homo-(A · T) sequences [type (1)] with various helical pitches, $\gamma_L = 13$, 20, and 60 [$\Delta\psi_L/(2\pi) = 0.172$, 0.145, 0.112]. As mentioned above [Eq. (3)], small interstrand interactions, large tensions, and small values of ϵ_L result in large values of γ_L. Here we vary γ_L independently of ϵ_L and f. As shown in Fig. 6(b), the critical torque τ_L^* increases (in magnitude) with increasing γ_L and the width of the plateau regime extends further if γ_L is small.

C. Sequence dependence: Competitions between B-Z and B-L transitions

We now consider a Z core of size N_Z together with a Lcore of size N_L in the same chain. Similar to the scheme shown in Fig. 3(a), the molecule has two handles of total size N_h. In Fig. 7, we present the twist response of this combined sequence: Z core of $(\rm GC)_n$ combined with a L core where (A · T) and (G · C) monomers alternate along the sequence. As the free energy cost, we use $\epsilon_Z = 0.51k_{\rm B}T$ for Z core, and alternating values of $0.6k_{\rm B}T$ and $2.4k_{\rm B}T$ for monomers in the L core. Here we impose the fixed boundary conditions for the Z core, such that it is separated from the L core by several bps in the B-DNA state. As seen in the torque-twist relation, the B-Z transitions and B-L transitions are clearly separated (due to the high junction energy penalty). The average size of Z-DNA and L-DNA are shown in Fig. 7(b) as a function of δLk. As unwinding torsional stress applies, Z-DNA forms first and L-DNA forms after the size of Z-DNA is saturated

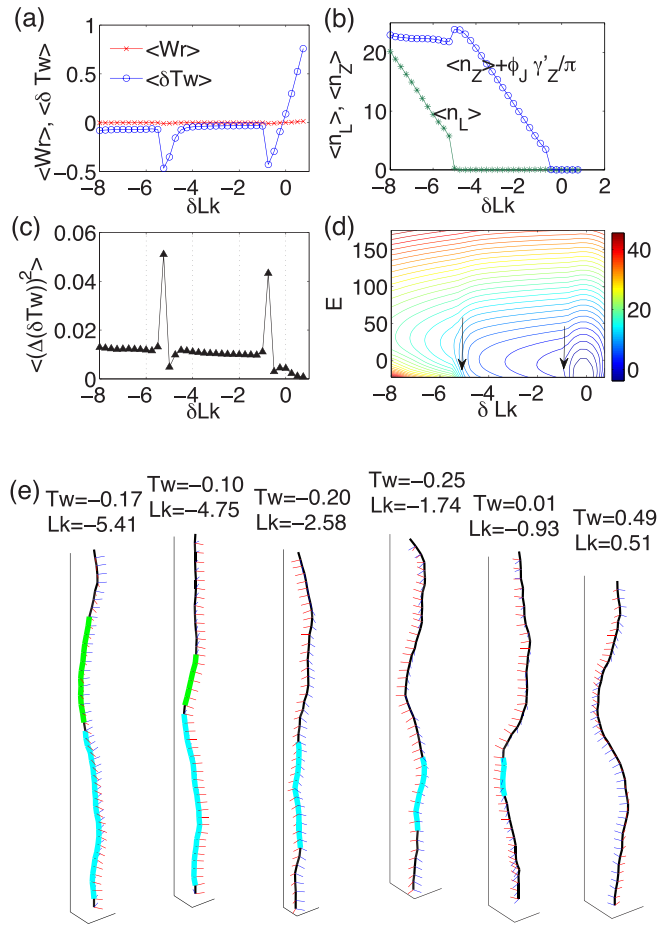

FIG. 7. (a) Twist δTw and writhe Wr as functions of δLk under the tension of $f = 8$ pN. The molecule consists of a Z core $(\rm GC)_{11}$ ($N_Z = 22$), a random sequence [50% of (A · T), $N_L = 23$], and flanking handles. For L-DNA $\gamma_L = 9$ and $\phi_J = -0.2\times 2\pi$. (b) Twist allocated in Z-DNA, $\langle n_Z\rangle/\gamma'_Z$, and L-DNA, $\langle n_L\rangle/\gamma'_L$, are shown for each γ_L. (c) Fluctuations in twist δTw ($\gamma_L = 9$ only). (d) Free energy landscape. Each colored line connects points of equal free energy. Two arrows indicate the onset of B-L and B-Z transitions. (e) Configurations of DNA cores at various values of δLk. The torsion to unwind DNA is increasing from right to left. Blue and green segments represent Z-DNA and L-DNA states, respectively. The centerline twist (besides the intrinsic twist) of each chain segment are depicted by a set of two unit vectors **f** and **v** orthogonal to the tangent **u**.

at $\delta{\rm Lk}^* = n_Z\Delta\psi_Z/(2\pi) + 2\phi_J/(2\pi)$. For $\delta{\rm Lk} < \delta{\rm Lk}^*$, both structures coexist as demonstrated in Fig. 7(e). At the onset of each transition, the fluctuations in δTw are sharply peaked [Fig. 7(c)]. The contour plot of the free energy curve $F(E,\delta{\rm Lk})$ [Fig. 7(d)] shows that the free energy minimum shifts to large values of E as more negative torsional stress is given. The changes in F are steep at the onset of B-Z, and B-L transitions [indicated by arrows in Fig. 7(d)]. Typical conformations of core parts are shown in Fig. 7(e). Along the chain, three orthonormal vectors for each bond display twisting and bending of the centerline of the chain.

The boundary condition of the Z core would have some influence on Z-DNA formation. To study this effect, we consider a Z core where all three states (B, L, Z) are allowed

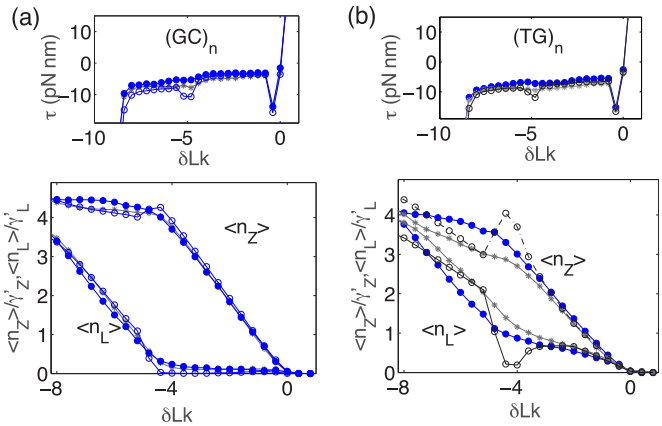

FIG. 8. Influence of boundary conditions on B-Z transition of Z core, $(GC)_n$ or $(TG)_n$, combined with L core with $\epsilon_L = 1.0k_BT$. $N_Z = N_L = 25$ and $N_h = 50$ and the L-DNA helical pitch is $\gamma_L = 20$. Torsional relations (top) and Z-DNA (L-DNA) population $\langle n_Z\rangle(\langle n_L\rangle)$ (bottom) are shown for Z cores of $(GC)_n$ (left) and $(TG)_n$ (right). The Z-L junction energy is set to be $E_{ZL} = 0.2E_J$ for the free boundary conditions (filled circle), and $E_{ZL} = 0.3E_J$ for the mixed boundary conditions (⋆). The Z-L junction is not allowed under fixed boundary conditions (open circle). The applied tension is $f = 8$ pN.

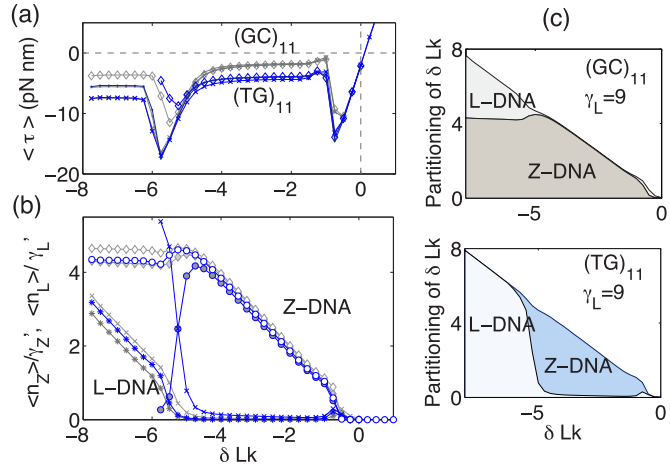

FIG. 9. Influence of the helical pitch of L-DNA on B, Z, and L phase. The chain contains a Z-DNA forming sequence of $N_Z = 22$ and a L-DNA forming sequence of $N_L = 78$. In the L-DNA sequences, (A · T) and (G · C) base pairs alternate ($\tilde{\epsilon}_L \approx 1.5k_BT$, $u_{AT} = 0.6k_BT$ and $u_{GC} = 2.4k_BT$). We assume that $\phi_J = -0.4\times 2\pi$ and $f = 8$ pN. The blue and grey curves represent data for $(TG)_n$ and $(GC)_n$. (a) Torque measurements of two types of Z cores embedded in L cores of various γ_L. Symbols ◇,·, and × correspond to data with $\gamma_L = 9$, 20, and 60. (b) $\langle n_Z\rangle$ and $\langle n_B\rangle$ in the B-Z-L transitions are shown for $(TG)_n$ type Z cores (blue) and for $(GC)_n$ type Z cores (gray). $\langle n_Z\rangle$ are represented as open symbols for $\gamma_L = 20$ and closed symbols for $\gamma_L = 9$. $\langle n_L\rangle$ are represented as ⋆ for $\gamma_L = 20$ and × for $\gamma_L = 9$. For $\gamma_L = 9$, the Z-DNA population of $(TG)_n$ drops at large torsional stress when L-DNA starts forming. (c) Area plots displaying the partitioning of δLk into Z-DNA and L-DNA for $(GC)_n$(top) and $(TG)_n$(bottom) cores. Here we use $\gamma_L = 9$.

and allocated at different locations with respect to the L core. When the Z core is in the middle of the L core, a Z-L junction can be formed. The energy cost of a Z-L junction E_{ZL} is expected to be considerably smaller than that of a B-L or B-Z junction, E_J. We thus refer to this condition as "free" boundary conditions here. If a L core is directly connected to only one end of the Z core, the "mixed" boundary condition applies. When a L core and a Z core are separated, the "fixed" boundary condition applies.

In Fig. 8, we show the torsional responses of these sequences with $\gamma_L = 20$ under various boundary conditions applied to Z cores with $N_Z = N_L = 25$ and $N_h = 50$. We choose the energetic cost of converting one bp to L-DNA inside the Z core according to the type of sequence: $1.7k_BT$ for $(TG)_n$, and $2.4k_BT$ for $(GC)_n$. In the L core, we assume the unpairing energy of a bp is $\tilde{\epsilon}_L = 1.0k_BT$. For the $(GC)_n$ core, Z-DNA and L-DNA populations are not much influenced by the junction condition, where the Z-DNA state is clearly dominant under the torsional stress. For the $(TG)_n$ core, either L-DNA or Z-DNA can be formed at small torsional stress with their respective statistical weights. If two cores are separated (fixed boundary), a left-handed structure (either L-DNA or Z-DNA block) would be formed first at a weak core and propagate to a nearby core. Formation of multiple blocks involves the large junction energy penalty and hence the corresponding transition torque is large. When the free boundary condition applies, a propagation of a left-handed structure to another core is smooth. With the prescribed energy parameters, L-DNA in a L core is less costly than Z-DNA and hence a left-handed structure forms at a L core first. At a larger torsional stress, Z-DNA with a smaller helical pitch is more efficient to relax the torsional stress. Due to the soft Z-L junction, Z-DNA can be easily accommodated in a Z core surrounded by AT rich patches. Besides the transition behavior, the overall B, Z, L populations are not very different from one condition to the other. As expected, the Z-DNA population increases if L-DNA is more stretched. The Z-DNA population changes from 70% to 80% when γ_L is changed from 13 to 20 (under the free boundary condition). Below, we stay with fixed boundary conditions for further investigation because the values of Z-L junction energy is yet poorly known.

In Fig. 9, we compare the B-L and B-Z transitions of two types of Z-DNA core sequences [$(GC)_n$ or $(TG)_n$] with longer (alternating) L-DNA forming sequences ($\tilde{\epsilon}_L \approx 1.5k_BT$). We now use several values ($\gamma_L = 9$, 20, and 60) for the helical pitch of L-DNA. In the regime of the B-Z transition, the critical torque τ_Z^* is ~ -2 pN nm for $(GC)_n$ cores, and ~ -4 pN nm for $(TG)_n$ cores. In the B-L transition regime (δLk < -6), the critical torque τ_L^* scales with $2\pi/\Delta\psi_L = 4.85$, 6.89, and 8.94, thus leading to $\tau_L^* \approx -4$, -6, and -8 pN nm for $\gamma_L = 9$, 20, and 60, respectively. These characteristic torque values τ_L^* for the B-L transitions are the same for both types of Z cores. The L-DNA population is large for small γ_L. For $\gamma_L = 9$, $\langle n_Z\rangle$ of the $(GC)_n$ core is slightly reduced after the L-DNA is formed. Similar to the $(GC)_n$ core, when combined with L-DNA with large $\gamma_L = 20,60$, Z-DNA from the $(TG)_n$ core is stable under all negative torsional stress and the coexistence regime is followed by the B-Z transition regime. With $\gamma_L = 9$, however, Z-DNA from the $(TG)_n$ core becomes unstable below the superhelical density for L-DNA nucleation. The linking numbers δLk partitioned to L-DNA and Z-DNA are shown as

an area plot in Fig. 9(c) for both types of Z cores with L-DNA of $\gamma_L = 9$.

There are local variations in the free energy ϵ_L along random L-DNA sequences. Sufficiently long L cores might include long AT rich regions of which the average values of $\tilde{\epsilon}_L$ are comparable to that of Z cores. To capture such an effect, we consider core sequences coupled to random L-DNA forming sequences with various (A · T) bp ratios. We take into account the tension dependent monomer size by rescaling the free energy cost under a given tension using Eq. (2). We now take the values of γ_L in relation with the applied tension and compute the corresponding helical pitch of L-DNA under a given tension. A specific case with moderate excluded volume parameter v is shown in Table I. According to the estimate given earlier, the critical torques for the B-Z transition in the $(TG)_n$ core crosses that for the B-L transition in the tension range 15–20 pN (Fig. 2). Below we present the results of B-Z and B-L transitions computed at two tension values $f = 13$ and 30 pN.

We rescale u_{AT}, u_{GC}, and ϵ_Z according to Eq. (2) for the given tension. The change in ϵ_Z is $-f\delta a$ and $\delta a = a_Z - a_B$ is 0.05 nm, independent of the values of tension. Assuming that L-DNA is first induced from the AT rich duplex, we apply $\delta a = 0.02$ nm for $f = 13$ pN and $\delta a = 0.07$ nm for $f = 30$ pN. Accordingly, we use $\gamma_L = 10$ and 15 for the tension of 13 and 30 pN, respectively, which are typical for local binding energy of (A · T) bp (see Table I). For L-DNA formed at large torsional stress, a larger $\tilde{\epsilon}_L$ should be involved. For $f = 13$ pN, the influence of $\tilde{\epsilon}_L$ on γ_L is almost negligible. For $f = 30$ pN, this choice of parameters slightly overestimates γ_L at large torsional stress. Because of the relatively high free energy cost, the Z-DNA population from $(TG)_n$ sequence is expected to be influenced by monomer composition of the neighboring regions. In Fig. 10(a), we show torque-twist relations for the B-L and B-Z transitions of a core sequence $(TG)_{11}$ connected to L-DNA forming random sequences with 50% (A · T) base pairs. The area plot shown in the right-hand side demonstrates the partitioning of the linking numbers to B-, Z-, or L-DNA structures. Unlike $(GC)_n$ cores (shown for example in Fig. 9), both Z-DNA and L-DNA form at small unwinding torsion and the Z-DNA fraction in $(TG)_n$ cores gradually increases with increasing unwinding torsional stress, which is consistent with experimental observations [11]. With $f = 30$ pN ($\gamma_L = 15$), yet more turns are partitioned into Z-DNA than L-DNA. For $f = 13$ pN, the turns used for conversion into Z-DNA are about 50% of the total applied linking number δLk. As more negative torsional stress is applied, L-DNA becomes dominant while Z-DNA population is drastically reduced. With a larger tension $f = 30$ pN, the dominance of the Z-DNA phase is further extended to the more negative torsional regime. In Fig. 10(b), B-Z-L transitions for the combined sequence where L cores contain 65% (A · T) are shown. For $(GC)_n$ cores, Z-DNA dominates at small negative superhelical stress. Z-DNA remains stable at larger negative superhelical stress where L-DNA also forms and shares the torsional stress. On the contrary, for $(TG)_n$ cores, L-DNA is dominant even at small negative δLk. With 80% of (A · T) base pairs (not shown), the population of Z-DNA diminishes for all negative twist regimes. The change of tension does not bring qualitative differences besides some shift of the midpoint

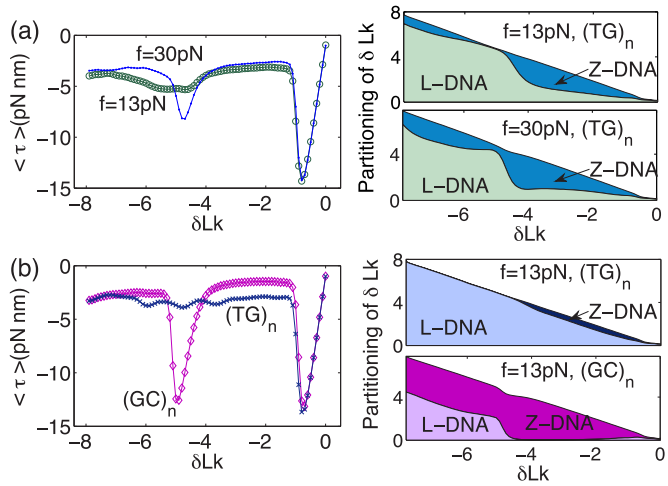

FIG. 10. Torsional responses (left panels) and partitioning of δLk (right panels) of Z cores combined with L cores ($u_{AT} = 0.6k_BT$, $u_{GC} = 2.4k_BT$) are shown for different tensions and sequences. The core lengths are $N_Z = 22$ and $N_L = 78$. In the right panels, the dark areas correspond to Z-DNA phase and the light areas to L-DNA phases. (a) Tension dependence of the stability of the Z-DNA formed by $(TG)_n$. The Z core is embedded into a random sequence of 50% of (A · T) content. (b) Either Z-core sequence $(GC)_n$ or $(TG)_n$ is embedded in a L core consisting of "random" sequences of 65% (A · T) content. The applied tension is set to be $f = 13$ pN.

torque because both free energy costs ϵ_L and ϵ_Z are rescaled by similar reduction due to tension. The competition between Z-DNA and L-DNA seems to be more strongly correlated with the *local* sequence composition than tension.

With 50% of (A · T) in sequence composition, the free energy cost is expected to be almost self-averaging along the sequence as long (A · T) blocks are unlikely to be present. For 65% of (A · T) in sequence composition, the relevant segment of L core is the longest block of (A · T) rich sections. For the specific realization shown in Fig. 10, the longest (A · T) block is 7 bps long. The size of AT rich region, separated by no more than a single (G · C) type bp, is ∼20. The average free energy cost $\tilde{\epsilon}_L$ of this section which easily turns to L-DNA is $\sim 0.8k_BT$ and it is smaller than the free energy cost of the Z-DNA core of $(TG)_n$, $\epsilon'_Z = 1.1k_BT$. The fluctuations in binding energy account for the dominance of L-DNA over Z-DNA from the $(TG)_n$ sequence, in a long DNA molecule.

V. CONCLUSIONS

We performed MC simulations for the B-Z and B-L transitions in DNA by adopting the recently developed WLC random copolymer model. Subject to the applied torsional stress, base pairs in the DNA chain interconvert between different helical states which are distinguishable by the preferred twist angles. The Wang-Landau type density of states method is applied in order to obtain the thermodynamic properties of molecules under imposed torsional stress. While imposing the topological constraint of external δLk, we take into account bending and twist fluctuations explicitly, in the tension regime above the supercoiling transition.

We first show that the model successfully captures the reported results of single molecule experiments. Torque-twist relations for the B-Z and B-L transitions, and Z-DNA populations we acquired in this study explain experimental results in Refs. [6,8,10] and are in good agreement with a theoretical study [15]. The critical torque for the B-Z transition in $(GC)_n$ obtained from our simulations results are ~ -2 pN.nm, which falls within the range of experimentally observed values. On the other hand, the estimated critical torques for the B-L transition are slightly smaller than experimental values available, which is $\sim -(7–12)$ pN nm. Among our simulation data, the small values of critical torques $-(2–4)$ pN nm correspond to the low salt concentration ($\sim$20–30 mM). With larger unpairing energies corresponding to high salt conditions, we found the critical torque to be in the range $-(6–10)$ pN nm. The denaturation energies of base pairs strongly depend on sequence composition and salt conditions [17]. The buffer conditions used in magnetic tweezers experiments [5,8,9] can be considered as high salt conditions. In Ref. [5], the critical torque for the B-L transition is measured to be $\sim -(10–12)$ pN nm or larger for a random sequence of 50% of $(A \cdot T)$ content (over the tension range from a few to 35 pN) with 150 mM salt. Variations in the critical torque is related to the disorder in the sequence compositions. A bubble starts in an AT rich region first. As L-DNA grows, the bubble extends to regions including stronger bonds. The critical torque will be shifted to more negative values. Nonuniform local unpairing energy also affects torque measurements away from equilibrium. Due to the finite speed of twisting, typically 10°/s, the system is not always at equilibrium. We expect the transition torque to be shifted to more negative value due to the dissipation. Furthermore base unpairing is dominated by breaking stronger bonds.

Then we further investigate the B-Z transitions of core sequence interacting with the B-L transitions occurring at neighboring DNA strands. The geometric and energetic properties of L-DNA may challenge the stability of Z-DNA, especially for those from the $(TG)_n$ sequence. For $(TG)_n$ type cores, we found that if DNA handles include long AT rich sequences, the L-DNA phase prevails over the Z-DNA phase and Z-DNA formation is suppressed. Naturally, a long random DNA sequence includes AT rich sections and the probability to find a AT rich region of a given length increases linearly with chain length before saturation. While Z-DNA from $(GC)_n$ cores is always supported by a negative superhelical twist, the presence of a long AT rich sequence would interfere with Z-DNA formation from $(TG)_n$ repeats. Ref. [11] pointed out that $(TG)_n$ is more commonly found in natural DNA [34,35] and the biological role of $(TG)_n$ is attributed to its volatile nature, that is, its rapid interconversion between B form and Z form under some torsional stress as demonstrated therein. It is of interest to quantify the level of torsion or tension under which Z-DNA states from various sequence types are stable. We show that Z-DNA and L-DNA populations are influenced by tension and more significantly by sequence compositions for $(TG)_n$ sequences. We also studied B-Z transitions of Z cores bounded with AT rich sequences, where the Z phase and the L phase can come directly into contact. The weak Z-L interfacial energy helps to form Z-DNA at smaller negative torsional stress.

In this study, we focus on large tension limit where the helical pitch of L-DNA is replaced by a effective twist rate γ_L. To understand Z-DNA formation at small tension, we would need to take into account the influence of supercoiling of B-DNA and L-DNA. Furthermore a proper modeling of theB-Z junction would be essential for a more precise description of the B-Z transition. The twist angle absorbed in the junction and junction modulus can be modeled by a L-DNA-like structure with soft stiffness.

ACKNOWLEDGMENTS

Our work was supported by National Research Foundation grants provided by the Korean government (Grants No. NRF-2012 R1A1A3013044 and No. NRF-2014R1A1A2055681). S.-C. H acknowledges Grant No. (NRF-2012R1A1A2021736) and also support by IBS-R023-D1. S.K. acknowledges Grant No. (NRF-2015R1A2A2A01005916) and KISTI for the access of KREONET/GLORIAD network. Authors acknowledge H.-C. Jeong (Sejong University) for the development of the parallel code.

APPENDIX A: ESTIMATES OF CRITICAL TORQUES

Below we provide simple estimates for some characteristic values representing B-Z and B-L transitions following the simple model described in [11,16,25]. At large tension regime ($f > f_L^{sc}$), the elastic energy of a stretched DNA molecule in pure B form can be written as $H_B = \frac{l_t k_B T}{2S}(2\pi)^2(\delta \mathrm{Lk})^2$. The reference state for potential energy is that the chain is fully stretched along the force. When n_Z bps turn to Z-DNA with n_J B-Z junctions, some of the torsional stress can be released where the topological constraint becomes $2\pi\delta \mathrm{Lk}' = 2\pi\delta \mathrm{Lk} + n_Z\Delta\psi_Z - n_J\phi_J$ where $\Delta\psi_Z = 2\pi(\gamma_B^{-1} + \gamma_Z^{-1})$, and ϕ_J stands for the extra twist located at each B-Z junction [30]. The elastic energy of DNA with n_Z bps of the Z state thus reads $H_Z = \frac{l_t k_B T}{2S}(2\pi\delta \mathrm{Lk} + n_Z\Delta\psi_Z - n_J\phi_J)^2 + n_Z\epsilon_Z + n_J E_J$. From the optimization of the energy with respect to n_Z, the torque exerted on the optimized shape can be written as

$$\tau_Z^* = 2\pi l_t k_B T \frac{n_0*}{S} = -\frac{\epsilon_Z'}{\Delta\psi_Z}. \tag{A1}$$

The twist of B-DNA strand $n_0^* = \delta \mathrm{Lk} + n_Z\Delta\psi_Z/(2\pi) - n_J\phi_J/(2\pi)$ is determined by the torsional equilibrium condition. At the onset of the B-Z transition, the free energy associated to the two structures is equal. By equating $H_B = H_Z$, the torque τ_Z^{tr} at the start of the B-Z transition can be obtained as

$$\tau_Z^{\mathrm{tr}} = 2\pi\frac{l_t}{S}(\delta \mathrm{Lk}) \approx -\frac{\epsilon_Z'}{\Delta\psi_Z} - \sqrt{\frac{l_t k_B T}{2S} n_J\left(E_J + \frac{\epsilon_Z'\phi_J}{\Delta\psi_Z}\right)}. \tag{A2}$$

The second term reflects the domain wall energy penalty E_J and extra turns ϕ_J at junctions.

Similar to B-Z transitions, we find the critical torque τ_L^* and the transition torque τ_L^{tr} for the B-L transition as follows:

$$\tau_L^* = -\frac{\tilde{\epsilon}_L'}{\Delta\psi_L}, \quad \tau_L^{\mathrm{tr}} = -\frac{\tilde{\epsilon}_L'}{\Delta\psi_L} - \sqrt{\frac{l_t k_B T}{2S} n_J E_J}, \tag{A3}$$

where $\Delta\psi_L = 2\pi(\gamma_B^{-1} + \gamma_L^{-1})$ and $\tilde{\epsilon}'_L$ corresponds to the mean value of ϵ'_L along the relevant L-DNA segment.

APPENDIX B: TENSION DEPENDENT HELICAL PITCH OF L-DNA

Here, we calculate the helical pitch γ_L of L-DNA of length S_L using the L-DNA model developed from our previous study [16]. Converting a monomer state from B-DNA to L-DNA accompanies the change of twist angle by $\Delta\psi_L = 2\pi(1/\gamma_B + 1/\gamma_L)$. If n_d base pairs are switched to L-DNA, the actual remaining torsion in the strand will be reduced to $\delta\mathrm{Lk} + n_d\Delta\psi_L/(2\pi)$. We consider the shape of L-DNA, under the tension f, as two interwinding single strands in a helical form with the helical angle α and the helical radius r aligned along the direction of force. Each single strand consists of n_d nucleotides of size $b_0 = 0.5$ nm and has the persistence length $l_s \approx 1$ nm. Fluctuations present in two interwinding single strands can be taken into account by rescaling of the nucleotide size to be $b = b_0 z_L/S_L$, where z_L is the single strand extension obtained by the force-extension relation of WLC, $\frac{\hat{f}2l_s}{k_\mathrm{B}T} = \frac{z_L}{S_L} + \frac{1}{4(1-z_L/S_L)^2} - \frac{1}{4}$ with the tension $\hat{f}$ projected along each single strand. The total energy H_L of the partly denatured DNA comprises the elastic energy of the denatured part H_b, the twist energy of the duplex H_B, interaction between bases in the single strands H_{ex}, and energy cost related to the bubble denaturation E_unp:

$$\begin{aligned} H_L(n_0, f, \alpha, n_d) &= H_b + H_B + H_{ex} + E_\mathrm{unp}, \\ H_b &= 2\frac{B_d n_d b}{2r^2}\sin^4\alpha + f n_d b(1-\cos\alpha) + f n_d(a-b), \\ H_B &= \frac{l_t k_\mathrm{B} T}{2S}(2\pi)^2(n_0 + n_d/\gamma'_L)^2, \\ H_\mathrm{ex} &= v n_d/\cos\alpha, \quad E_\mathrm{unp} = \tilde{\epsilon}_L n_d + 2E_J, \end{aligned} \tag{B1}$$

where $B_d = l_p k_\mathrm{B} T$. The first term in H_b is the bending energy of the two separated strands, $B_d\kappa^2 L$, with the curvature of the helix being $\kappa = \sin^2\alpha/r$. The next terms stand for the potential energy associated with the bubble conformation subject to the tension. Note that H_{ex} takes into account a moderate average interaction between bases (electrostatics, excluded volume) where v is of the order of thermal energy. From the minimization of the free energy with respect to α and n_d, we obtain an equation determining the helical angle α as a function of unpairing energy $\tilde{\epsilon}_L$ and tension f as below:

$$\frac{B_d b}{r^2}\sin\alpha^4 - \frac{4\pi}{6\gamma'_L(\alpha)}\frac{4B_d\frac{b}{r^2}\sin\alpha^3\cos\alpha + fb\sin\alpha + \frac{v\sin\alpha}{\cos^2\alpha}}{\cos(2\alpha)\frac{b}{r}} + \frac{v}{\cos\alpha} + f(a - b\cos\alpha) + \tilde{\epsilon}_L = 0. \tag{B2}$$

The helical pitch of L-DNA γ_L is then obtained in relation with the optimal helical angle α^* as $\gamma_L(f) = \frac{4\pi r}{3\sin[2\alpha^*(f)]b}$. The base-pair rise of the L form can be obtained as $a_L = b\cos\alpha^*$ for an L-DNA in a tight ply shape.

[1] R. R. Sinden, *DNA Structure and Function* (Elsevier, San Diego, 1994).
[2] M. Fuertes, V. Cepeda, C. Aloso, and J. Perez, Chem. Rev. **106**, 2045 (2006).
[3] A. Herbert, M. Schade, K. Lowenhaupt, J. Alfken, T. Schwartz, L. S. Shlyakhtenko, Y. L. Lyubchenko, and A. Rich, Nucl. Acids Res. **26**, 3486 (1998).
[4] P. Ho, Proc. Natl. Acad. Sci. USA **91**, 9549 (1994).
[5] M. Y. Sheinin, S. Forth, J. F. Marko, and M. D. Wang, Phys. Rev. Lett. **107**, 108102 (2011).
[6] M. Lee, S. H. Kim, and S.-C. Hong, Proc. Natl. Acad. Sci. USA **107**, 4985 (2010).
[7] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature (London) **424**, 338 (2003).
[8] F. C. Oberstrass, L. E. Fernandes, and Z. Bryant, Proc. Natl. Acad. Sci. USA **109**, 6106 (2012).
[9] F. C. Oberstrass, L. E. Fernandes, P. Lebel, and Z. Bryant, Phys. Rev. Lett. **110**, 178103 (2013).
[10] P. Lebel, A. Basu, F. Oberstrass, E. M. Tretter, and Z. Bryant, Nat. Meth. **11**, 456 (2014).
[11] S. H. Kim, A.-R. Lee, S. H. Lim, M. Cho, N.-K. Lee, J.-H. Lee, and S.-C. Hong (unpublished).
[12] T. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, Biophys. J. **74**, 2016 (1998).
[13] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[14] D. Salerno, A. Tempestini, I. Mai, D. Brogioli, R. Ziano, V. Cassina, and F. Mantegazza, Phys. Rev. Lett. **109**, 118303 (2012).
[15] A.-Y. Kwon, N.-K. Lee, S.-C. Hong, J. Fiering, and A. Johner, Biophys. J. **108**, 2562 (2015).
[16] A. Son, A.-Y. Kwon, A. Johner, S.-C. Hong, and N. K. Lee, Europhys. Lett. **105**, 48002 (2014).
[17] J. SantaLucia, Jr., Proc. Natl. Acad. Sci. USA **95**, 1460 (1998).
[18] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[19] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, Curr. Opin. Struct. Biol. **10**, 279 (2000).
[20] G.-M. Nam, N.-K. Lee, H. Mohrbach, A. Johner, and I. Kulić, Europhys. Lett. **100**, 28001 (2012).
[21] W. Bauer and C. J. Benham, J. Mol. Biol. **234**, 1184 (1993).
[22] J. F. Marko, Phys. Rev. E **76**, 021926 (2007).
[23] S. Neukirch, Phys. Rev. Lett. **93**, 198107 (2004).
[24] S. Neukirch and E. L. Starostin, Phys. Rev. E **78**, 041912 (2008).
[25] N. Clauvelin, B. Audoly, and S. Neukirch, Macromolecules **41**, 4479 (2008).
[26] G. van der Heiden and J. Thompson, Nonlin. Dyn. **21**, 71 (2000).
[27] J. F. Marko and S. Neukirch, Phys. Rev. E **88**, 062722 (2013).
[28] S. Cocco, J. Yan, J.-F. Leger, D. Chatenay, and J. F. Marko, Phys. Rev. E **70**, 011910 (2004).

[29] G. Chirico and J. Langowski, Biopolymers **34**, 415 (1994).
[30] G. Wang and K. Vasquez, Front. Biosci. **12**, 4424 (2007).
[31] J. Lee, Y.-G. Kim, K. K. Kim, and C. Seok, J. Phys. Chem. B **114**, 9872 (2010).
[32] L. Peck and J. Wang, Proc. Natl. Acad. Sci. USA **80**, 6206 (1983).
[33] A.-R. Lee, H.-E. Kim, Y.-M. Lee, M. Jeong, K.-H. Choi, J.-W. Park, Y.-G. Choi, H.-C. Ahn, B.-S. Choi, and J.-H. Lee, Biochem. Biophys. Res. Commun. **428**, 137 (2012).
[34] H. Hamada, M. G. Petrino, T. Kakunaga, M. Seidman, and B. Stollar, Mol. Cell Biol. **4**, 2610 (1984).
[35] J. Majewski and J. Ott, Genome Res. **10**, 1108 (2000).