

Supercoiling Induces Denaturation Bubbles in Circular DNA

Jae-Hyung Jeon,¹ Jozef Adamcik,² Giovanni Dietler,² and Ralf Metzler¹

¹*Department of Physics, Technical University of Munich, 85747 Garching, Germany*

²*Laboratoire de Physique de la Matière Vivante, Institute of Physics of Biological Systems (IPSB), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

(Received 8 June 2010; published 11 November 2010)

We present a theoretical framework for the thermodynamic properties of supercoiling-induced denaturation bubbles in circular double-stranded DNA molecules. We explore how DNA supercoiling, ambient salt concentration, and sequence heterogeneity impact on the bubble occurrence. An analytical derivation of the probability distribution to find multiple bubbles is derived and the relevance for supercoiled DNA discussed. We show that *in vivo* sustained DNA bubbles are likely to occur due to partial twist release in regions rich in weaker AT base pairs. Single DNA plasmid imaging experiments clearly demonstrate the existence of bubbles in free solution.

DOI: 10.1103/PhysRevLett.105.208101

PACS numbers: 87.15.A–, 36.20.Ey, 87.64.Dz

Because of the complementarity of bases allowing only the formation of AT and GC base pairs (bps) double-stranded (DS) DNA self-assembles into the duplex form from two single strands (SS) [1–3]. At physiological conditions, the thermodynamically stable duplex can be disrupted to allow local opening of bps, creating a SS denaturation bubble. Such bubbles are promoted solely by thermal fluctuation or by an applied mechanical torque. Their size fluctuations were measured by fluorescence correlation and described in terms of stochastic models [4,5]. Since DNA bubble formation is essential for biological processes such as transcription initiation, its properties are extensively studied [3]. While most studies on DNA bubbles concentrate on a free linear DNA with fully relaxed mechanical torsion, native DNA typically is subject to topological constraints. Eukaryotic linear DNA forms a chromosome structure tightly packaged around histones, while prokaryotic DNA typically occurs in circular form. Under such topological constraint twist and bending modes of DNA are intrinsically coupled, with significant changes in the thermodynamics of bubbles. As a prominent example single molecule experiments using magnetic tweezers showed that a linear supercoiled DNA forms a bubble to alleviate the twist strain imposed by an external unwinding torque when the torque exceeds a threshold and the induced superhelical coiling is mechanically removed by a stretching force [6,7].

DNA bubbles may also occur in circular supercoiled molecules simply due to the interplay of twist and writhe [8], as experimentally proven by different methods, e.g., electron microscopy [9], gel electrophoresis [10], and atomic force microscopy (AFM, Fig. 1). An important question is to reveal the location of bubbles for given sequences, expected to occur in weaker AT-rich regions, and to understand the biological role of bubbles for processes requiring opening up of the DS.

It is in fact nontrivial to develop a quantitative analytical model that fully takes into account the writhe, especially in the presence of spontaneous superhelical coiling. In principle, one should integrate out the bending fluctuation modes for a given topological state and obtain a renormalized twist energy. Except for the case when DNA undulations are negligible in presence of a stretching force [7], this integration remains formidable. Thus, theoretical studies usually rely on coarse-grained simulations [11] or effective models [12,13]. A very successful approach is the Benham model [12]: it predicts the location and sizes of bubbles for given supercoiled DNA, in good agreement with experimental data. Most studies of supercoiled circular DNA focus on critical exponents [13] or melting profiles [12].

Here we study the equilibrium properties of denaturation bubbles and show that their average number and size are determined by supercoiling and DNA length. We also show that the sequence heterogeneity and salt concentration crucially impact the bubble statistics. The claims from our model are corroborated by AFM measurements of

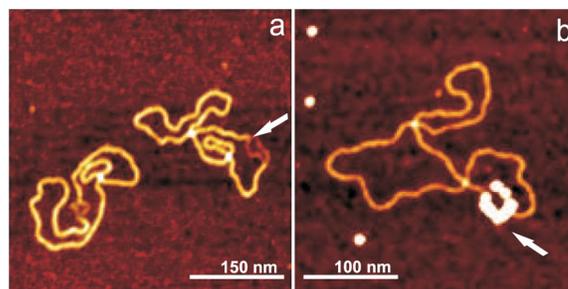


FIG. 1 (color online). AFM images of pUC19 plasmids in 1 mM Tris-HCl buffer at $pH = 7.8$ in (a) absence and (b) presence of *E. coli* SS DNA binding proteins (SSB), deposited on 3-aminopropyltriethoxysilane modified mica [8]. (a) Denaturation bubbles appear as fainter lines in the plasmids (arrow). (b) The bright ring is the SSB-covered bubble (arrow).

DNA plasmids. Our findings are important to understand better the formation of bubbles in supercoiled circular DNA as well as their biochemical relevance.

Consider a supercoiled circular DNA with N bps and superhelical density $\sigma = (Lk - Lk_0)/Lk_0$, where Lk_0 and Lk are the linking numbers of the relaxed and supercoiled DNA [8,14]. Natural DNA isolated from either prokaryotes or eukaryotes has superhelical density $\sigma \approx -0.06$. The DNA molecule's conformational state is described by the free energy $\mathcal{E}_0 = \mathcal{E}_{sc}^0 + \mathcal{E}_{ss} + \mathcal{E}_{bp}$, where \mathcal{E}_{sc}^0 is the supercoiling energy of the DS part of the DNA, \mathcal{E}_{ss} the inter-strand twist energy of SS bases in the bubble region, and \mathcal{E}_{bp} denotes the energy cost for bubble formation. Since the DNA is topologically constrained, the total linking number is conserved [8,14]. On breaking n bp the linking number satisfies $\Delta Lk_{ds} + \Delta Lk_{ss} - n/\theta_0 = \sigma Lk_0$, ΔLk_{ds} being the residual linking number in the DS, ΔLk_{ss} the linking number due to the spontaneous twist of the SS in bubbles, and n/θ_0 the linking number loss.

It was shown experimentally that \mathcal{E}_{sc}^0 is a quadratic function of ΔLk_{ds} , i.e., $\mathcal{E}_{sc}^0 = \frac{1}{2}K(\sigma Lk_0 + n/\theta_0 - \Delta Lk_{ds})^2$ with $K \approx 2200RT/N$ and $\theta_0 \approx 10.4$ bps/turn [12]; R is the gas constant and T the absolute temperature. With the (average) twist angle per broken bp τ the interstrand twisting energy becomes $\mathcal{E}_{ss} = \frac{1}{2}nC\tau^2$. The experimentally determined twist modulus is $C \approx 9.3 \times 10^{-21}$ erg · cm [12]. Defining a two-state variable n_i at location i , where $n_i = 0/1$ for a closed or open bp, the unbinding energy becomes $\mathcal{E}_{bp} = \sum_{i=1}^N [(\varepsilon_I + \varepsilon_{i,i+1})n_i - \varepsilon_I n_i n_{i+1}]$. Here, ε_I is the bubble initiation energy and $\varepsilon_{i,i+1}$ is one of ten nearest-neighbor energies between the i th and $(i+1)$ th bp.

Tracing out the twist energy of bubbles \mathcal{E}_{ss} via $e^{-\beta\mathcal{E}_{sc}} = \int e^{-\beta(\mathcal{E}_{sc}^0 + \mathcal{E}_{ss})} d\tau$ we obtain a renormalized supercoiling energy \mathcal{E}_{sc} . The effective energy $\mathcal{E}_{\text{eff}} = \mathcal{E}_{sc} + \mathcal{E}_{bp}$ becomes

$$\mathcal{E}_{\text{eff}}(\{n_i\}) = \frac{2\pi^2 CK}{4\pi^2 C + Kn} \left(\sigma Lk_0 + \frac{n}{\theta_0} \right)^2 + \mathcal{E}_{bp}, \quad (1)$$

where $n = \sum_i n_i$ is the total number of broken bps. From Eq. (1), the partition Z is obtained from all 2^N states as $Z = \sum_{\{n_i=0,1\}} \exp(-\beta\mathcal{E}_{\text{eff}}(\{n_i\}))$. We rewrite Z in the form $Z \equiv \sum_{m=0} Z^{(m)}$ where $Z^{(m)}$ corresponds to the partition for states with exactly m bubbles. At physiological conditions far below the melting temperature, multiple-bubble states are generally negligible due to the large bubble initiation [$\varepsilon_I \approx 9-11k_B T$ [15]]; $Z^{(m)}$ decreases rapidly with m , and Z is approximated by the first terms $Z^{(m)}$.

Homopolymer.—In homogeneous DNA $Z^{(m)}$ becomes

$$Z^{(m)} = e^{-m\beta\varepsilon_I} \sum_{n=m} \mathcal{D}^{(m)}(n) e^{-\beta\mathcal{E}(n)}, \quad (2)$$

where $\mathcal{E}(n) = \mathcal{E}_{\text{eff}} - m\varepsilon_I$ is the DNA energy with n broken bps in m bubbles, and $\mathcal{D}^{(m)}(n)$ is the degeneracy of the corresponding state. Since $\mathcal{E}(n) = \mathcal{E}_{sc}(n) + n\bar{\varepsilon}$ with average unbinding energy $\bar{\varepsilon}$ per bp, it suffices to count $\mathcal{D}^{(m)}(n)$

to evaluate $Z^{(m)}$. In circular DNA, $\mathcal{D}^{(1)}(n) = N$ and $\mathcal{D}^{(2)}(n) = N(N-n-1)(n-1)/2! \sim O(N^3)$. As $\mathcal{E}(n)$ is m -independent, $Z^{(m)}$ [$m > 1$] is comparable with $Z^{(1)}$ only when $\mathcal{D}^{(m)}(n)$ is of order $e^{(m-1)\beta\varepsilon_I} \approx 10^{4(m-1)} - 10^{5(m-1)}$. This gives a reasonable truncation value of m in calculating Z for given DNA length, e.g., $\mathcal{D}^{(m=2)}(n) \approx e^{\beta\varepsilon_I}$ when $N \gtrsim O(10^3)$.

The average bubble size is $\langle n \rangle = Z^{-1} \sum_{m=1} e^{-m\beta\varepsilon_I} \sum_{n=m} n \mathcal{D}^{(m)}(n) e^{-\beta\mathcal{E}(n)} \equiv \sum_{m=1} \langle n \rangle^{(m)}$, where $\langle n \rangle^{(m)}$ is the average bubble size contribution from the m -bubble state. Introducing the probability distribution function (pdf) $\mathcal{P}^{(m)}(n)$ of m -bubble openings with a total of n broken bp, we see that $\langle n \rangle^{(m)} = \sum_{n=m} n \mathcal{P}^{(m)}(n)$. By comparison, we find the pdf for multiple-bubble states, $\mathcal{P}^{(m)}(n) = e^{-m\beta\varepsilon_I} \mathcal{D}^{(m)}(n) e^{-\beta\mathcal{E}(n)} / Z$ with normalization $\sum_{m=0} \sum_{n=m} \mathcal{P}^{(m)}(n) = 1$. The probability of m -bubble occurrence $\mathcal{P}^{(m)} = \sum_{n=m} \mathcal{P}^{(m)}(n)$ and the average number of bubbles becomes

$$\langle m \rangle = - \frac{\partial}{\beta \partial \varepsilon_I} \log Z = \sum_{m=1} m \sum_{n=m} \mathcal{P}^{(m)}(n). \quad (3)$$

Random heteropolymer.—To gain some insight into the effect of the heteropolymer nature of DNA we consider a random sequence of bps and obtain the disorder-averaged partition. Now $e^{-\beta\mathcal{E}(n)}$ becomes a random variable and thus the disorder-averaged partition $\overline{Z^{(m)}}$ is

$$\overline{Z^{(m)}} = e^{-m\beta\varepsilon_I} \sum_{n=m} e^{-\beta\mathcal{E}_{sc}(n)} m! \mathcal{D}^{(m)}(n) e^{-\beta \sum_{i=k}^{k+n-1} \varepsilon_{i,i+1}}; \quad (4)$$

i.e., we need to evaluate $\overline{\exp(-\beta \sum_{i=k}^{k+n-1} \varepsilon_{i,i+1})}$. When $n=1$ we obtain the disorder average in the form $\Lambda = \sum_{\{\alpha, \alpha'\}} \exp(-\beta\varepsilon_{\alpha, \alpha'}) / 16$ with $\alpha, \alpha' \in \{A, T, G, C\}$. For $n \geq 2$, we use a transfer matrix method: Consider the 4×4 matrix $\mathbf{T}_{\alpha, \alpha'} = \sqrt{p_\alpha p_{\alpha'}} \exp(-\beta\varepsilon_{\alpha, \alpha'})$ with equal probability of nucleotide occurrence $p_\alpha = 1/4$ [16]. Then, $\overline{e^{-\beta \sum_{i=k}^{k+n-1} \varepsilon_{i,i+1}}} = \text{tr}\{\mathbf{T}^n\} = \sum_{j=1}^4 \Lambda_j^n$, where Λ_j are the eigenvalues of \mathbf{T} , such that

$$\overline{Z^{(m)}} = e^{-m\beta\varepsilon_I} \sum_{n=m} e^{-\beta\mathcal{E}_{sc}(n)} m! \mathcal{D}^{(m)}(n) \left(\delta_{n,1} \Lambda + \sum_{j=1}^4 \Lambda_j^n \right). \quad (5)$$

$\overline{Z^{(m)}}$ is now calculated quite efficiently. For large N we approximate $\log \overline{Z}$ as $\log \bar{Z}$, and the disorder-averaged multiple-bubble occurrence pdf $\overline{\mathcal{P}^{(m)}(n)}$ is inferred from Eq. (3) via $\sum_{m=1} m \sum_{n=m} \overline{\mathcal{P}^{(m)}(n)} = -\beta^{-1} \partial \log \bar{Z} / \partial \varepsilon_I$:

$$\overline{\mathcal{P}^{(m)}(n)} = \frac{e^{-\beta(m\varepsilon_I + \mathcal{E}_{sc}(n))} m! \mathcal{D}^{(m)}(n) \left(\delta_{n,1} \Lambda + \sum_{j=1}^4 \Lambda_j^n \right)}{\bar{Z}}. \quad (6)$$

We now study the effect of DNA supercoiling σ and DNA length N on the bubble statistics. Figure 2 shows $\overline{P^{(m)}}(n)$ of heterogeneous DNA of lengths 300 and 2000 bp for given values of σ at 310 K and 100 mM salt. The degree of supercoiling significantly impacts on the pdf regardless of DNA length. When slightly supercoiled (here $\sigma = -0.03$), the distribution follows a typical exponential decay with n , showing that bubble formation is unfavorable. In contrast, the distribution becomes bell shaped as $|\sigma|$ increases, i.e., when a stable bubble is expected. There also occurs a dramatic change of the pdf depending on DNA length. For shorter DNA ($N = 300$), the single-bubble occurrence always prevails over the two-bubble event within a reasonable range of σ , although the two-bubble contribution tends to increase with enhanced supercoiling $|\sigma|$. For longer DNA ($N = 2000$), the two-bubble probability can be larger than the single-bubble probability, in particular, as σ approaches the native value $\sigma \approx -0.06$.

From the pdfs above we obtain the probability for multiple-bubble occurrence and the average number of bubbles for AT/GC homopolymers and random heteropolymer in Fig. 3. Again a qualitative difference between shorter and longer DNA occurs. In shorter DNA at most one bubble occurs in the biologically relevant range of σ , contrasting longer DNA where the probability to find one bubble is decreased as it is supercoiled, e.g., below $\sigma \approx -0.04$ for AT-DNA: two-bubble states are preferred at large $|\sigma|$ values at physiological temperature.

Supercoiling-induced bubbles are highly sequence dependent. In contrast to AT homogeneous or random heterogeneous DNA, within the physiologically relevant range for σ and N considered here, even a single-bubble event becomes significantly less probable for a GC homopolymer DNA (Fig. 3). Bubbles due to partial twist release will therefore occur almost exclusively in DNA domains rich in AT. Qualitatively, the bubble occurrence in random heterogeneous DNA appears similar to those for AT

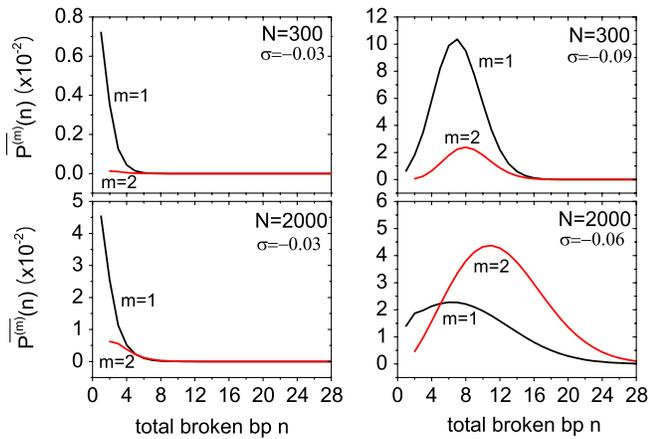


FIG. 2 (color online). Disorder-averaged pdf of multiple-bubble occurrence in random heteropolymer DNA ($N = 300, 2000$) with stability parameters from Ref. [15], at 310 K, 100 mM salt.

homopolymer DNA in Fig. 3. However, compared to the latter case the two-bubble occurrence probability is suppressed, so that at $\sigma \approx -0.05$ (-0.06) the expected number of bubbles is ≈ 0.6 (1.4) in random heterogeneous DNA while it is ≈ 1.6 (1.8) in homogeneous AT-DNA, due to the absence of stable GC bps.

Figure 4 depicts the average number of broken bps $\langle n \rangle$ as a function of σ and salt concentration: bubble formation occurs only when σ exceeds a threshold below which bubbles grow linearly with increasing $|\sigma|$. This linearity indicates that bubbles alleviate the twist strain to an extent proportional to $|\sigma|$, such that the residual twist strain in the DS part becomes almost σ independent. Note that bubbles do not completely remove the twist strain as bp-breaking consumes $\sim 1k_B T$. Less surprisingly, the N dependence of the threshold value shows that a bubble forms easier in longer DNA. In Fig. 4 we also show the dependence of $\langle n \rangle$ on salt concentration at $\sigma = -0.06$. Within the physiologically accessible range, reduced salt concentration favors bubble formation in short DNA and promotes their growth in long DNA. A similar behavior is observed upon temperature increase (see [8]).

The above theory can be extended to include the binding of SSBs on open bubbles, an important experimental trick to corroborate bubble existence [Fig. 1(b)]. For two major binding modes for *E. coli* SSB, i.e., $(SSB)_{M=35}$ and $(SSB)_{M=65}$ modes (see [8]), the statistical weight of SSB binding on a bubble of size n per mode M is

$$Z_M^S(n) = 2(n - \lambda_M + 1)c_S K_M + \sum_{n_S=2}^{2n/\lambda_M} \sum_{j=0}^{n_S-1} \Omega_{M,2n}(n_S, j) \times (c_S K_M)^{n_S} \omega_M^j. \quad (7)$$

K_M is the equilibrium binding constant, ω_M the nearest neighbor cooperativity factor: c_S the SSB concentration, and λ_M the number of nucleotides occluded per SSB in mode M . $\Omega_{M,2n}(n_S, j)$ counts the degeneracy of n_S SSBs binding to $2n$ bases with j adjacencies (see [8] for details).

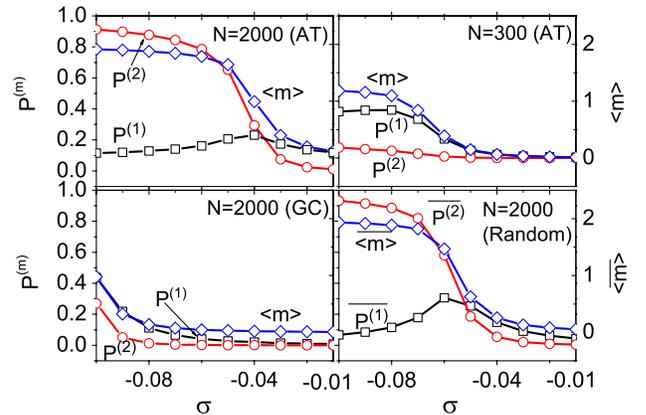


FIG. 3 (color online). Probability of multiple-bubble occurrence and average number of bubbles for pure AT/GC DNA and heteropolymer DNA. Same parameters as Fig. 2.

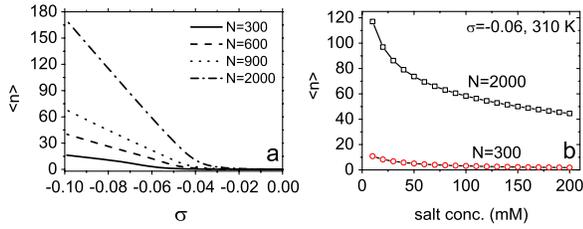


FIG. 4 (color online). (a) Total number of broken bps $\langle n \rangle$ vs σ for DNA length $N = 300, 600, 900,$ and 2000 (bottom to top) at 310 K and 100 mM salt. (b) $\langle n \rangle$ vs salt concentration at $\sigma = -0.06$ and 310 K. Results for AT homopolymer DNA.

The total partition Z_{tot} for a plasmid-SSB complex yields by substitution: $e^{-\beta\mathcal{E}(n)} \rightarrow e^{-\beta\mathcal{E}(n)}[1 + Z_{35}^S(n) + Z_{65}^S(n)]$ in $Z^{(m)}$. The probability for SSB binding to a bubble is

$$\mathcal{P}_S = \sum_{m=1} \sum_{n=m} \frac{e^{-m\beta\epsilon_i} \mathcal{D}^{(m)}(n) e^{\beta\mathcal{E}(n)}}{Z_{\text{tot}}} [Z_{35}^S(n) + Z_{65}^S(n)]. \quad (8)$$

Table I shows the probability to observe an SSB complex in pUC19 plasmids prepared at buffer conditions of 37°C , $\text{pH } 7.8$, and SSB concentration $0.13 \mu\text{M}$ ($10 \mu\text{g/ml}$). We observe excellent agreement with Eq. (8). Without correction for SSB binding, the agreement is still reasonable.

Experiment.—Our analysis demonstrates that denaturation bubbles in both homogeneous and heterogeneous supercoiled circular DNA are stabilized by partial relaxation of the twist. Their lifetime is therefore considerably longer than denaturation bubbles in linear free DNA.

We performed an AFM study of pUC19 plasmids (2686 bp, 95% occurrence in supercoiled form) at low ionic strength and in the presence of *E. coli* SSB [8]. In the case of supercoiled molecules (more than 1000 molecules were analyzed) only one denaturation bubble per molecule was observed [Fig. 1(a)]. This finding is consistent with the theoretical results above for the heteropolymer case. However, no bubble was observed in the case of relaxed or nicked (350 and 420 molecules analyzed) plasmids for $\sigma = 0$. In this case, in contrast to the stable bubbles seen for $\sigma \approx -0.06$, any thermally induced bubbles die out quickly and thus cannot be detected by our AFM method. In the presence of SSB proteins, their complexation with the SS DNA bubble region is clearly visible [Fig. 1(b)]. As the deposition process is quite fast and the binding of SSB proteins on deposited SS expected to be impeded, the experiment strongly suggests that a single long-lived bubble already exists in solution, i.e., the SSB protein-bubble complex is not an effect of the deposition on the mica.

Combining theoretical and single plasmid experimental results we proved that partial twist release stabilizes DNA bubbles in supercoiled DNA, and is thus likely to facilitate molecular processes requiring the opening up of DS DNA in living cells. We note that we reanalyzed all results from

TABLE I. SSB binding probability on pUC19 at varying $[\text{NaCl}]$: experimental results, predictions from Eq. (8), and the model without SSB. >300 plasmids were imaged per $[\text{NaCl}]$.

Salt concentration	50 mM	100 mM	150 mM
SSB binding prob.	98%	78%	49%
\mathcal{P}_S [Eq. (8)]	99%	78%	50%
w/o SSB binding	95%	74%	59%

Figs. 2–4 with a different set of stability parameters from Ref. [17] (see [8]). Despite the difference in specific values we observe remarkable consistency.

- [1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland, New York, 2002).
- [2] M. D. Frank-Kamenetskii, *Phys. Rep.* **288**, 13 (1997).
- [3] R. M. Wartell and A. S. Benight, *Phys. Rep.* **126**, 67 (1985); C. Richard and A. Guttman, *J. Stat. Phys.* **115**, 925 (2004).
- [4] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, *Phys. Rev. Lett.* **90**, 138101 (2003).
- [5] A. Hanke and R. Metzler, *J. Phys. A* **36**, L473 (2003); D. J. Bicout and E. Kats, *Phys. Rev. E* **70**, 010902(R) (2004); T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, *Phys. Rev. Lett.* **97**, 128105 (2006); J.-H. Jeon, W. Sung, and F. H. Ree, *J. Chem. Phys.* **124**, 164905 (2006); T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, *Biophys. J.* **92**, 2674 (2007); H. C. Fogedby and R. Metzler, *Phys. Rev. Lett.* **98**, 070601 (2007).
- [6] T. R. Strick, V. Croquette, and D. Bensimon, *Biophys. J.* **74**, 2016 (1998).
- [7] J.-H. Jeon and W. Sung, *Biophys. J.* **95**, 3600 (2008).
- [8] See supplementary material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.105.208101>.
- [9] C. Brack, T. A. Bickle, and R. Yuan, *J. Mol. Biol.* **96**, 693 (1975); S. Dasgupta *et al.*, *J. Biol. Chem.* **252**, 5916 (1977).
- [10] W. R. Bauer and C. J. Benham, *J. Mol. Biol.* **234**, 1184 (1993).
- [11] S. P. Mielke *et al.*, *J. Chem. Phys.* **123**, 124911 (2005); C. A. Sucato *et al.*, *Biophys. J.* **86**, 3079 (2004); F. Trovato and V. Tozzini, *J. Phys. Chem. B* **112**, 13197 (2008).
- [12] R. M. Fye and C. J. Benham, *Phys. Rev. E* **59**, 3408 (1999); T. B. Liverpool, S. A. Harris, and C. A. Laughton, *Phys. Rev. Lett.* **100**, 238103 (2008).
- [13] A. Kabakçioğlu, E. Orlandini, and D. Mukamel, *Phys. Rev. E* **80**, 010903(R) (2009).
- [14] A. D. Bates and A. Maxwell, *DNA Topology* (Oxford University Press, Oxford, 2005).
- [15] D. Jost and E. Everaers, *Biophys. J.* **96**, 1056 (2009).
- [16] T. Hwa, E. Marinari, K. Sneppen, and L. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4411 (2003).
- [17] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, *Biophys. J.* **90**, 3091 (2006).