Maxwell relations for single-DNA experiments: Monitoring protein binding and double-helix torque with force-extension measurements

Houyin Zhang
Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA

John F. Marko
Department of Physics and Astronomy and Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

(Received 30 November 2007; published 18 March 2008)

Single-DNA stretching and twisting experiments provide a sensitive means to detect binding of proteins, via detection of their modification of DNA mechanical properties. However, it is often difficult or impossible to determine the numbers of proteins bound in such experiments, especially when the proteins interact nonspecifically (bind stably at any sequence position) with DNA. Here we discuss how analogs of the Maxwell relations of classical thermodynamics may be defined and used to determine changes in numbers of bound proteins, from measurements of extension as a function of bulk protein concentration. We include DNA twisting in our analysis, which allows us to show how changes in torque along single DNA molecules may be determined from measurements of extension as a function of DNA linking number. We focus on relations relevant to common experimental situations (e.g., magnetic and optical tweezers with or without controlled torque or linking number). The relation of our results to Gibbs adsorption is discussed.

DOI: 10.1103/PhysRevE.77.031916 PACS number(s): 87.14.gk, 82.35.Rs, 87.15.kj, 61.30.Hn

I. INTRODUCTION

Micromanipulation of individual DNA molecules provides a powerful tool for the study of protein-DNA interactions [1–4]. DNA in living cells is to a first approximation covered by proteins which provide the context for functions of the double helix, including gene expression, physical chromosome organization, chromosome replication, and genetic recombination. Therefore, development of quantitative tools for analysis of protein-DNA interactions is essential to mechanistic understanding of the cell. In the rather common case where a protein of interest deforms DNA (e.g., by bending, twisting, or otherwise deforming the double helix), mechanical experiments can directly detect its binding [1,5–10].

If there are only a few binding sites along a DNA strand under micromechanical study, then one might be fortunate enough to be able to simply count the number of binding events, e.g., by observation of jumps in apparent DNA extension at some fixed force [1,7,11]. This is especially relevant to proteins which bind specific DNA targets of defined sequence. Alternatively, one might directly observe the number of proteins bound, e.g., by detection of fluorescently labeled proteins [12]. However, both of these strategies start to run into severe limitations when one starts to consider an assembly of arrays of proteins along a long DNA strand, in experiments where one is trying to analyze proteins that are capable of binding to essentially any sequence position. Examples of such proteins include most of the DNA-packaging proteins found in high numbers inside cells, such as the histone proteins of eukaryote nucleosomes [13], eukaryote HMG-type proteins which nonspecifically bend DNA [9], or similar DNA-bending proteins found in high numbers in prokaryote cells such as the *Escherichia coli* proteins HU [8,9], IHF [5], and FIS [14,15].

The proteins mentioned above have all been studied in single-DNA micromanipulation experiments, but with the limitation that in each case no strategy has allowed convincing determination of the numbers of proteins bound. In some cases experimentalists have estimated numbers of bound proteins by carrying out parallel solution-phase bulk experiments [5,15], but the different thermodynamic conditions make such comparisons problematic. Theorists have addressed this problem through the approach of analyzing microscopic statistical-mechanical models in order to match their predictions to experimental data [2–4]. While valuable in generating qualitative predictions for broad classes of experiment, solid quantitative results for numbers of proteins bound are difficult to obtain from a microscopic model due to the large number of unknown microscopic parameters typically involved in such models (binding site size, bending angle, etc.), not to mention the further complication of sequence variation of affinities that occurs even in the case of sequence-nonspecific DNA-binding proteins.

This paper develops the idea that measurements of thermodynamic work done against an externally controlled tension can be used as a tool to analyze binding of such proteins, without the need for theoretical assumptions beyond the existence of thermodynamic equilibrium. Analogous ideas allow analysis of torques generated in single-DNA-twisting experiments via measurement of extension changes with changes in DNA linking number. Our approach uses single-DNA analogs of the Maxwell relations of classical thermodynamics. We present definitions of free energies applicable to various experimental situations starting from a statistical-mechanical perspective, and then we analyze the resulting free energies from a thermodynamic point of view. Our approach has some similarities to thermodynamic methods used to analyze binding of water to biomacromolecules as a function of osmotic pressure, which is usually varied through changing concentration of an "osmolyte" molecule which itself adsorbs water [16,17].

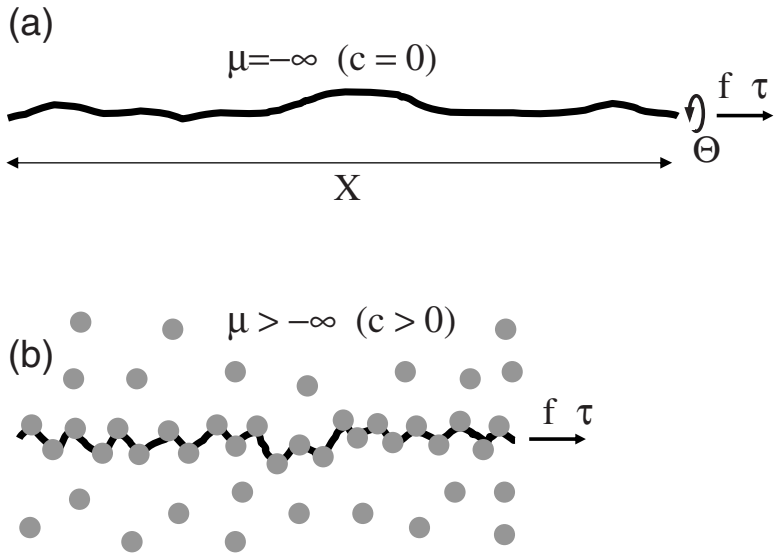

FIG. 1. Cartoon of DNA molecule held at constant force and torque (a). When a DNA-binding protein is introduced at some finite concentration (b), it will bind and change the average extension and linking number of the molecule.

Section II treats the case of a DNA molecule subject to fixed force, chemical potential for binding, and DNA torque. We show how the Maxwell relations arise, and how they allow measurements of extension as a function of chemical potential and force to be used to determine changes in the number of proteins bound as force is changed. Section II also illustrates the use of Maxwell relations via analysis of a simple, exactly soluble model of proteins that generate "folding" of DNA when they bind.

Section III discusses the case corresponding to force-cantilever experiments with a relatively large cantilever spring constant, where one has fixed end-to-end extension, chemical potential, and torque. This case corresponds to common optical tweezer experiments with single-strand DNA attachments (where the DNA torque is zero). The general case of nonzero torque corresponds to a new class of optical tweezer experiments carried out at fixed extension and controlled torque [18,19].

Section IV discusses the case where a DNA molecule is again held at fixed force and protein chemical potential, but where the DNA linking number is controlled. This case corresponds to magnetic-tweezer DNA pulling-twisting experiments [20–22]. In the thermodynamic limit, it would correspond to the torque-to-linking-number Legendre-transformed version of the case of Sec. II. However, note that for short DNA molecules (<1 kb) that are sometimes used experimentally, finite-size effects may generate significant differences between different ensembles.

A notable result in the fixed-force case is a method to compute changes in torque using measurements of extension changes with linking number, which is potentially useful even for studies of "naked DNA" with no protein present [23].

II. DNA MOLECULE CONTROLLED BY FORCE, TORQUE, AND PROTEIN CONCENTRATION

We begin by discussing the case of a single DNA held at fixed force and constant torque [Fig. 1(a)]. In the special case of zero torque, this corresponds to the situation where a "nicked" DNA (or a DNA held down by single-strand end attachments) is studied, which is unable to support any internal static torsional stress. Constant DNA tension is usual in magnetic-tweezer experiments, but may also be obtained using a force-clamped feedback system for optical tweezers, or even with specialized optical field gradients [24].

We also suppose that a species of protein (or other molecules that might bind or adsorb to the DNA contour) can then be introduced at concentration c [Fig. 1(b)]. When these small molecules bind to the double helix, they will generate a change in molecule extension, or linking number that can be observed experimentally. It is important to recognize that the discussion of this paper presupposes thermodynamic equilibration of protein binding, i.e., reversibility of the accessibility and distribution of microstates as protein concentration is cycled up and down. This is not always the case experimentally: many protein-DNA complexes have been observed to assemble readily, but to disassemble only sluggishly or not at all, apparently not reaching equilibrium on experimentally accessible time scales [9,14,15,25–27].

If binding equilibrium can be achieved, the situation discussed above is described using Boltzmann statistical mechanics starting from a microscopic energy function for a finite-length DNA molecule interacting with protein, of the form

$$E = -\tau\Theta - fX - \sum_i n_i k_B T \ln(c/K_i) + E_{\text{int}}. \qquad (1)$$

Here τ is the external constant torque, coupled to the DNA linking number expressed as an angle; the conventional linking number of molecular biology is $\text{Lk}=\text{Lk}_0+\Theta/(2\pi)$. A constant external force f is coupled to one space component of the end-to-end extension (X).

Binding of proteins is accounted for by a (finite) sum over all possible binding positions i, with (in general variable) binding affinities K_i, coupled to binding-position-occupation variables n_i, which are either 1 or 0 depending on whether a protein is bound there or not. This type of model is widely

used to describe binding of proteins and other small molecules to DNA, and is suitable so long as the DNA concentration is low enough that it cannot perturb the bulk protein concentration c (valid for single-DNA studies), and so long as the bulk concentration of protein is low enough for its solution thermodynamics to be treated as a dilute ideal gas (valid for DNA-binding proteins usually studied at micromolar or lower concentrations). The "chemical potential"–like parameter $\mu \equiv k_B T \ln c$ will be used below; note that this quantity can be additively decoupled from the site-variable affinities K_i.

Finally Eq. (1) contains an internal (possibly free) energy E_{int} which describes the conformational fluctuations of the DNA molecule and bound proteins, including the elastic properties of the double helix, interactions between proteins at nearby binding positions, and effects of DNA deformation mediated by the bound protein. The results of this paper do not rest on the detailed form of E_{int} (a variety of models can be found in the literature [2,4,28–30]), apart from the essential property that the externally controlled thermodynamic fields f, c (equivalently μ), and τ do not appear within E_{int}.

It should be recognized that the end-to-end extension in the force direction X, the average total linking number Θ, and the average number of proteins bound N are quantities that are dependent on the microscopic degrees of freedom, e.g., $N \equiv \Sigma_i n_i$. These quantities undergo microscopic fluctuations, and their averages correspond either to measurable quantities (e.g., $\langle X \rangle$), or things we would like to know about (e.g., $\langle N \rangle$).

A. Expectation values

Given this rather general model for a finite micromanipulated DNA molecule interacting with proteins or other molecules in solution, we suppose that we compute the partition function by summing the Boltzmann weights over all conformations of the n_i and other microscopic variables. For a finite-length molecule, the resulting partition function $\ln Z(f,\mu,\tau) \equiv \ln \Sigma \exp(-\beta E)$ is an analytic function of thermodynamic fields f, μ, and τ. Equation (1) indicates that the partial derivatives of $\ln Z$ generate the average end-to-end extension in the force direction, the average number of bound proteins, and the average linking number:

$$\langle X \rangle = \left(\frac{\partial k_B T \ln Z}{\partial f} \right)_{\mu,\tau}, \quad \langle N \rangle = \left(\frac{\partial k_B T \ln Z}{\partial \mu} \right)_{f,\tau},$$

$$\langle \Theta \rangle = \left(\frac{\partial k_B T \ln Z}{\partial \tau} \right)_{f,\mu}. \tag{2}$$

The expectation values emphasize that these are statistical-mechanical relations, following directly from the coupling of fields to quantities dependent on microscopic degrees of freedom in Eq. (1), and also that their values must be determined statistically in an experiment.

B. Maxwell relations

One may write the "perfect differential" for the thermodynamic potential,

$$d(k_B T \ln Z) = \langle X \rangle df + \langle N \rangle d\mu + \langle \Theta \rangle d\tau. \tag{3}$$

Barring any singularities (none should be present for finite DNA length), the mixed second derivatives are independent of the order of partial differentiation, i.e., $\partial^2 \ln Z / \partial f \, \partial \mu = \partial^2 \ln Z / \partial \mu \, \partial f$. This gives rise to the three Maxwell relations

$$\left(\frac{\partial \langle X \rangle}{\partial \mu} \right)_{f,\tau} = \left(\frac{\partial \langle N \rangle}{\partial f} \right)_{\mu,\tau}, \quad \left(\frac{\partial \langle N \rangle}{\partial \tau} \right)_{f,\mu} = \left(\frac{\partial \langle \Theta \rangle}{\partial \mu} \right)_{f,\tau},$$

$$\left(\frac{\partial \langle \Theta \rangle}{\partial f} \right)_{\mu,\tau} = \left(\frac{\partial \langle X \rangle}{\partial \tau} \right)_{\mu,f}. \tag{4}$$

Equation (4) suggests strategies to measure quantities which would otherwise be difficult to determine. For example, the first relation indicates that the relatively straightforward determination of how DNA extension changes with protein chemical potential immediately indicates how the number of bound proteins changes as force is changed. The second relation shows how changes in linking number with protein concentration acting against a constant torque, expected for any protein that either twists or chirally bends ("writhes") DNA, determine how protein binding changes with torque at fixed concentration.

The final relation in (4) indicates how changes in linking number with force are related to changes in extension with torque; note that this relation applies to DNA in the absence of any protein (i.e., the case $\mu = -\infty$) and provides a route to measure torque changes in naked DNA. This type of measurement will be discussed in more detail in Sec. IV.

C. Determination of numbers of proteins bound from force-extension measurements

We now return to the first Maxwell relation of (4) that relates extension changes with chemical potential to changes in numbers of bound protein with force. The requirement that this be done at a fixed torque may be achievable experimentally, but also corresponds to the special case $\tau = 0$ corresponding to a torsionally unconstrained DNA (either nicked, or tethered by a single-strand connection, a rather widely studied case). This Maxwell relation indicates that, if one measures extension as a function of force and protein concentration, the change in number of proteins bound as force is changed from f_0 to f, at fixed μ and τ, can be obtained by integration:

$$\langle N \rangle_{f,\mu,\tau} = \langle N \rangle_{f_0,\mu,\tau} + \int_{f_0}^{f} df' \left(\frac{\partial \langle X \rangle}{\partial \mu} \right)_{f',\tau}. \tag{5}$$

Equation (5) shows how to use force-extension data to directly determine the change in absolute numbers of bound molecules, for binding positions with possibly widely varying affinities (generally the case even for "nonspecific" protein-DNA interactions), with no theoretical presumptions beyond existence of thermodynamic equilibrium.

For a protein that bends or otherwise contracts DNA, extension will be reduced as μ is increased, making the integral of (5) negative for $f > f_0$. Thus, as force is increased, DNA-

contracting proteins will thus be driven off the double helix. In the opposite case where DNA is extended by a protein or other ligand, one can expect increased binding as the force is increased. In a favorable experimental situation (possibly achieved by careful choice of protein concentration) one might be able to drive off the majority of a DNA-contracting protein by increasing the force sufficiently. In this case, we would have $f_0=\infty$ (or large) and $\langle N\rangle_{f_0}=0$, and therefore

$$\langle N\rangle_{f,\mu,\tau}=-\int_f^\infty df'\left(\frac{\partial\langle X\rangle}{\partial\mu}\right)_{f',\tau}. \tag{6}$$

The identification of this type of favorable "reference state" allows determination of absolute numbers of bound proteins.

D. Analysis of a simple model of DNA-distorting proteins

To concretely illustrate Eqs. (5) and (6) we analyze a simple and exactly solvable statistical-mechanical model of DNA-compacting proteins. This will give some idea of what experimental data might look like and how they can be treated to extract numbers of bound proteins. We suppose that our DNA molecule is of length L; we consider the easily soluble case where there are distinct binding sites each of length d, for a total of L/d noninteracting binding sites (the more general case of proteins which interact along the DNA is included in the case to which our general thermodynamic formalism applies but for simplicity we consider a "noninteracting" model here). For simplicity we also suppress discussion of torque; one may generalize the model presented below to the torsionally constrained case [3,4].

For a naked double helix, previous work has established that the semiflexible polymer model gives a good account of low-force (<10 pN) DNA elasticity. As a high-force expansion, this model has a partition function that behaves as

$$k_BT\ln Z=Lg(f,A)=L(f-\sqrt{k_BTf/A}+\cdots), \tag{7}$$

where $A\approx 50$ nm is the persistence length of the double helix. The extension as a fraction of the total polymer length is just $\langle X\rangle/L=\partial g/\partial f=1-\sqrt{k_BT/(4Af)}+\cdots$. In order to obtain a globally reasonable thermodynamic potential for the semiflexible polymer at fixed force $g(f,A)$, it is convenient to use the Legendre transformation from the ensemble of fixed extension X [2,43]. Given the force as a function of extension ($f(X)$), the free energy at fixed extension is just the work done extending the polymer, $W(X)=\int_0^X dX' f(X')$. In turn, $g(f)=fX-W(X)$, where $f=\partial W/\partial X$. Below we use a free energy $g(f)$ computed in this way, starting with the approximate fixed-extension formula $f(X)=(k_BT/A)[X/L+(1/4)(1-X/L)^{-2}-1/4]$. If one wishes to start from an expression more closely approximating that associated with the idealized semiflexible polymer model, one may use the numerically complete solution discussed in Ref. [43].

We suppose that, when a protein binds to one of the L/d binding sites, it modifies the contribution to $k_BT\ln Z$ for that region of the molecule from its naked value gd to a modified value hd', where $h\equiv g(f,A')$. This is a model where binding of a protein changes the maximum extension of the binding site from d to d', and changes the persistence length of the binding site from A to A'. This provides a simple and analytically tractable way to modify the effective persistence length and maximum extended length with protein binding.

Since each binding site is treated independently, the partition function for the whole molecule is readily calculated from the two possible states for each binding site:

$$Z=(e^{\beta gd}+e^{\beta(\mu+hd')})^{L/d}. \tag{8}$$

Torque effects could also be included in this model, but for this discussion we limit ourselves to the torsionally unconstrained case $\tau=0$. Other elaborations such as binding-site inhomogeneity and interactions between proteins at adjacent binding sites may be added to this model.

The extension is readily calculated:

$$\langle X\rangle=\frac{\partial k_BT\ln Z}{\partial f}=L\frac{\partial g}{\partial f}+L\frac{[(d'/d)(\partial h/\partial f)-(\partial g/\partial f)]}{1+\exp[-\beta(\mu-gd+hd')]}, \tag{9}$$

as is the average number of bound proteins:

$$\langle N\rangle=\frac{\partial k_BT\ln Z}{\partial\mu}=\frac{L/d}{1+\exp[-\beta(\mu-gd+hd')]}. \tag{10}$$

Figure 2(a) shows extension (normalized to total DNA contour length L) versus force for this model, for the case $A=50$ nm, $A'=25$ nm, and $d'/d=0.9$. Curves from left to right show the cases $\beta\mu=-\infty$ (naked DNA), -8, -4, -6, -2, 0, and ∞ (protein-saturated DNA). Successively more positive $\beta\mu$ values lead to successively more protein binding, and consequent compaction against the applied force. The parameters chosen here generate a slight length contraction and a reduction in effective persistence length with protein binding, corresponding to effects of DNA-bending proteins seen experimentally [5,8,9,14,15] and studied using more detailed bending models [4].

Figure 2(b) shows the corresponding protein occupancies (bound protein number per total binding sites), which drop as force is applied. Successively larger μ values delay reduction in bound protein to successively larger forces (the sigmoidal curves from left to right show $\beta\mu=-8$, -4, -6, -2, and 0; the case $\mu=-\infty$ gives zero occupancy for all forces, while $\mu=\infty$ gives unit occupancy for all forces).

The mixed derivatives satisfy the Maxwell relation

$$\frac{\partial\langle X\rangle}{\partial\mu}=\frac{\partial\langle N\rangle}{\partial f}=\frac{L}{k_BT}\left(\frac{d'}{d}\frac{\partial h}{\partial f}-\frac{\partial g}{\partial f}\right)\times\frac{\exp[-\beta(\mu-gd+hd')]}{\{1+\exp[-\beta(\mu-gd+hd')]\}^2}. \tag{11}$$

This formula is a product of two factors. The latter ratio of exponentials is the usual peaked response function familiar from the theory of thermally excited two-state systems, and will always contribute. However, the preceding term in large parentheses controls the overall magnitude of the response function, and is proportional to the difference in extension of the naked polymer and the protein-coated polymer as a fraction of the naked polymer length. If this difference is zero or small, i.e., if the mechanical properties of naked and protein-coated DNA are nearly the same, then it will be difficult to

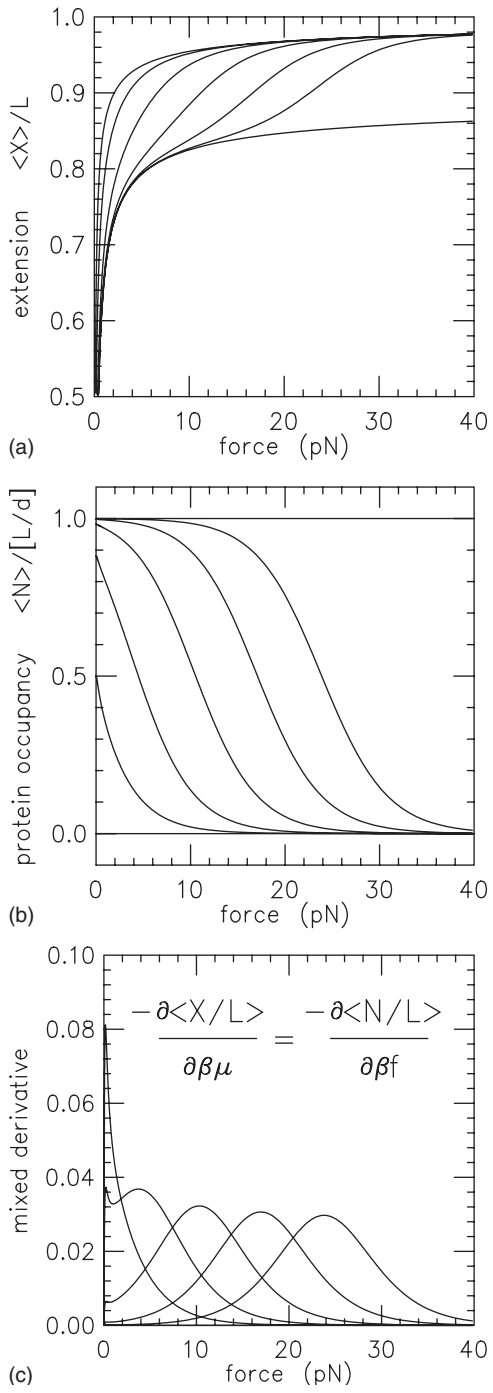

FIG. 2. Behavior of (a) extension versus force, (b) protein occupation versus force, and (c) the mixed derivatives for the simple model analyzed in Sec. II C. Parameters are chosen for a protein that contracts its DNA substrate against applied force when it binds. Successively lower extensions in (a), larger values of protein occupancy in (b), or rightward-shifted peak functions in (c) correspond to successively larger values of μ (see text for details).

impossible to measure a nonzero value of Eq. (11). On the other hand, this formula indicates that, if there are mechanical differences between naked and protein-coated DNA, and of course if equilibrium can be achieved, then changes in protein binding can be thermodynamically monitored.

Figure 2(c) shows these mixed derivatives (multiplied by $-k_BT/L$ so as to be converted to positive dimensionless form) corresponding to the cases shown in Figs. 2(a) and 2(b). Peaked curves from left to right show results for $\mu=-8$, -6, -4, -2, and 0; the maxima in Fig. 2(c) correspond to the drops in the curves of Fig. 2(b). In each case, since the protein occupancy drops to zero at large forces ($\lim_{f\to\infty}\langle N\rangle\to 0$), the integrals of these pulse-shaped functions from a given force up to $f=\infty$ give the numbers of proteins bound at that force.

Thus, in an experiment where thermodynamic equilibrium of protein binding occurs, one can obtain data such as those shown in Fig. 2(a), by measuring the force-extension response at a series of solution protein concentrations. One normally cannot accurately determine the number of bound proteins [Fig. 2(b)], but by calculating the chemical-potential derivative of the curves of Fig. 2(a) followed by integration over force, one can thus estimate the change in numbers of bound proteins using the Maxwell relation Eq. (5). We emphasize that the model discussed in this section simply provides an illustration of this general procedure. For the f-μ-τ ensemble, the second and third Maxwell relations of Eq. (4) can be used in an analogous way.

III. DNA MOLECULE CONTROLLED BY EXTENSION, TORQUE, AND PROTEIN CONCENTRATION

We now briefly describe the Maxwell relations relevant to the different ensemble where extension X is fixed and force fluctuates. We imagine dropping the term fX from Eq. (1), and then carrying out the partition sum at fixed extension X, to obtain $Z_O(X,\mu,\tau)$. The complete differential of the corresponding thermodynamic potential is

$$d(k_BT\ln Z_O) = -\langle f\rangle dX + \langle N\rangle d\mu + \langle\Theta\rangle d\tau. \tag{12}$$

In the thermodynamic limit $L\to\infty$ this potential can be obtained by Legendre transformation of Eq. (3); however, for finite L the relation $\langle X\rangle f = X\langle f\rangle$ is only approximate.

This set of thermodynamic control parameters is essentially that used in optical-tweezer or other stiff-cantilever experiments; we continue to work at fixed μ and fixed τ. For molecules that are torsionally unconstrained, $\tau=0$; recently a new type of optical-tweezer method has been developed which allows constant nonzero torques to be maintained, providing a complete realization of this ensemble [19].

Following the same approach as in the previous section, we obtain expectation values

$$\langle f\rangle = -\left(\frac{\partial k_BT\ln Z}{\partial X}\right)_{\mu,\tau}, \quad \langle N\rangle = \left(\frac{\partial k_BT\ln Z}{\partial \mu}\right)_{X,\tau},$$

$$\langle\Theta\rangle = \left(\frac{\partial k_BT\ln Z}{\partial \tau}\right)_{X,\mu}. \tag{13}$$

The corresponding Maxwell relations are

$$-\left(\frac{\partial\langle f\rangle}{\partial\mu}\right)_{X,\tau} = \left(\frac{\partial\langle N\rangle}{\partial X}\right)_{\mu,\tau}, \quad \left(\frac{\partial\langle N\rangle}{\partial\tau}\right)_{X,\mu} = \left(\frac{\partial\langle\Theta\rangle}{\partial\mu}\right)_{X,\tau},$$

$$\left(\frac{\partial\langle\Theta\rangle}{\partial X}\right)_{\mu,\tau} = -\left(\frac{\partial\langle f\rangle}{\partial\tau}\right)_{\mu,X}. \quad (14)$$

The first relation indicates that measurements of force as a function of concentration can be used to determine changes in bound protein number with extension. In the majority of optical-tweezer experiments done on torsionally unconstrained molecules ($\tau=0$) interacting with proteins, this relation allows changes in the number of proteins bound to be computed:

$$\Delta N \equiv \langle N\rangle_{X,\mu,\tau} - \langle N\rangle_{X_0,\mu,\tau} = -\int_{X_0}^{X} dX' \left(\frac{\partial\langle f\rangle}{\partial\mu}\right)_{X',\tau}. \quad (15)$$

For DNA-compacting (e.g., DNA-bending) proteins, the most useful reference state is likely to be high extension (maximal X_0), where the bound protein concentration might be driven to near zero.

The latter two Maxwell relations of (14) may become useful in newly developed optical-tweezer experiments where nonzero constant torques can be applied and measured. In such a case, (14) shows how linking number changes with chemical potential at constant torque indicate changes in protein binding with torque, and finally how linking number changes with extension are related to force changes with torque.

A. Work-adsorption relation

If we exchange the order of integration and differentiation Eq. (15) can be rewritten as

$$\Delta N = -\frac{\partial}{\partial\mu}\int_{X_0}^{X} dX'\langle f\rangle_{X',\tau} = -\frac{\partial}{\partial\mu}\Delta W, \quad (16)$$

where $\Delta W \equiv \int_{X_0}^{X} dX'\langle f(X')\rangle$ is the mechanical work done quasistatically in extending the polymer from X_0 to X, at constant torque and chemical potential. This relation connects changes in the number of molecules bound to the change in mechanical work that must be done to stretch the polymer with change in chemical potential.

Equation (16) has a form roughly similar to the "Gibbs adsorption" formula familiar from the thermodynamics of surface adsorption. However, the physical meaning of Eq. (16) is distinct from that of the classical Gibbs formula. The Gibbs formula asserts that the number of molecules N_s adsorbed to a surface of area A is proportional to the derivative of the surface tension γ with bulk chemical potential μ [31]: $N_s = -A\,\partial\gamma/\partial\mu$. This formula indicates that free energy lost by removing a particle from solution to an adsorbing surface must be balanced by the interactions between that particle and the surface, which generate the surface tension. Gibbs adsorption has been discussed in the context of binding of small molecules at the surface of large macromolecules, for example in osmotic stress studies of hydration of DNA [17].

However, unlike the classical Gibbs adsorption formula, Eq. (16) does not involve the total surface area (total chemical length for the present polymer case), but instead is based on the mechanical-force-dependent end-to-end extension of that polymer, in the presence of a solute which modifies that force dependence when adsorbed. While classical Gibbs adsorption is based on the surface tension contribution to the thermodynamic potential $\gamma\,dA$ for an area-independent surface tension, Eq. (16) is based on the mechanical work contribution $f\,dX$, with a strongly extension-dependent force. Therefore, Eq. (16) is not precisely analogous to the classical Gibbs adsorption formula, and is more closely related to thermodynamic relations used to make measurements of changes in the number of ions bound to bulk DNA phases as a function of their osmotic compression [16] than to classical Gibbs adsorption.

Rather than Eq. (16), the relation in this paper that most closely corresponds to the Gibbs adsorption formula is the expectation value for bound proteins $\langle N\rangle$ in Eqs. (2) and (13). The thermodynamic potential can be identified as a line tension associated with the total length L of the polymer, $\psi \equiv -(k_BT \ln Z)/L$, and the differential (3) can be generalized to include a term $-\psi\,dL$. Equilibrium between adsorbed and free molecules demands

$$\frac{\langle N\rangle}{L} = -\left(\frac{\partial\psi}{\partial\mu}\right)_{f,\tau}. \quad (17)$$

Equation (17) is the one-dimensional version of Gibbs' adsorption formula, modified to describe adsorption of particles onto a one-dimensional "surface" characterized by a free energy per length or line tension ψ.

One might measure ψ using a DNA molecule passed through a nanopore that permits sliding transfer of the double-helix contour, but which cannot pass proteins. If DNA-binding proteins were placed in a solution on only one side of such a setup, the DNA contour would be pulled toward the protein-containing region. Measurement of this protein-generated translocation force would give the protein-concentration-dependent part of ψ.

Equations (16) and (17) could have been obtained via purely thermodynamic arguments, by analyzing the total adsorbed excess obtained by integrating the difference between the protein concentration and the bulk concentration over a volume enclosing the entire DNA molecule (see Ref. [17] for an example of such a derivation involving adsorption onto molecules free in solution). Microscopic models of the form (1) can describe the general adsorption case where the nearby "atmosphere" of enriched concentration is counted as adsorbed; the only constraint is that the bulk concentration c (i.e., μ) should not appear within $E_{\rm int}$. Consideration of an atmosphere of adsorbed molecules is essential to analysis of water or ions near biomolecule surfaces [17]; for binding of larger biomolecules like proteins at low bulk concentrations, the question of the extent of the corresponding adsorption atmosphere is unknown.

B. Force "plateaus" during pseudo-phase-coexistence transitions

Given sufficient cooperativity, it is possible for even a one-dimensional and finite DNA molecule to show phase-coexistence-like behavior [2]. Examples from studies of stretching of single DNA molecules include structural changes in DNA structure such as the "overstretching" tran-

sition [32,33], binding of proteins such as RecA which essentially polymerize onto the double helix [34], and chemical condensation of DNA [35]. These types of highly cooperative reactions are often observed to proceed at nearly constant force.

Specializing to the case of drugs or proteins binding in a highly cooperative manner to DNA, it may be reasonable to describe the molecule as partitioned into two "phases," e.g., protein-bound and protein-free regions, with the length fractions of the two states variable with extension X. During conversion from one state to the other by variation of X, force must remain constant, since it is an intensive quantity determined by the intensive free energies of the two "coexisting states:" $(\partial\langle f\rangle/\partial X)_{\mu,\tau}=0$. We immediately obtain $\partial^2\langle f\rangle/(\partial X\partial\mu)=0$ and $\partial^2\langle f\rangle/(\partial X\partial\tau)=0$. Therefore, during this conversion process, the first and third relations of (14) are X independent. Integration of the first relation across the force plateau leads to

$$\frac{\Delta N}{\Delta X}=-\left(\frac{\partial f_{\text{plateau}}}{\partial\mu}\right)_{\tau}. \tag{18}$$

Thus, measurement of how a force plateau changes with chemical potential can provide an estimate of the total number of proteins bound or released as one moves across it. Alternatively, when one observes a nearly constant force with extension $(\partial f/\partial X\approx 0)$, one can immediately conclude that N is varying linearly with X over that interval.

This general idea can also be used to analyze highly cooperative structural transitions in naked DNA such as the "overstretching" transition [32,33]. When two thermodynamic fields are available, one can map out a "phase diagram," consisting of first-order-like transition lines. Force and torque provide a prototypical example of such a pair of fields, allowing structural transition lines to be determined in the force-torque plane [18,23,36,37]. Across these transitions there will be rapid variations of extension and linking number. Perhaps the best known of these is the overstretching transition at ≈65 pN and zero torque, across which DNA length changes by about 70% over only a few piconewtons force range [32,33]; the DNA linking number also changes dramatically during this transition [36]. Integration of the third relation in (14) across such a transition gives a relation between the ratio of extension and linking number shifts to the slope of the state boundary in the force-torque plane:

$$\frac{\Delta X}{\Delta\Theta}=-\left(\frac{\partial f_{\text{plateau}}}{\partial\tau}\right)_{\mu}. \tag{19}$$

Equations (18) and (19) are analogous to the Clausius-Clapeyron relation $dP/dT=\Delta S/\Delta V$ familiar from the study of solid-liquid-gas phase transitions.

Relations like this in the osmotic-pressure–temperature plane have been used to analyze changes in entropy and hydration of DNA during abrupt first-order-like structural transitions in DNA solutions driven by osmotic stress [16,38]. Single-DNA overstretching has been discussed as a sharp, first-order phase transition involving two coexisting phases, leading to formulas analogous to Eq. (18) defined in the force-temperature plane. This phase-equilibrium approach has been used to measure total entropy changes during the overstretching transition [39–42]. A similar approach in the force–ion-concentration plane has been employed to measure changes in ion binding during DNA overstretching [40,42].

Of course, in the case of one-dimensional stretched DNA molecules, transitions such as overstretching will be smoothed by thermal fluctuations and by quenched sequence disorder. Consequently application of phase equilibrium ideas can be only approximate, i.e., there will always be some "width" to the force or torque plateaus observed in experiments. However, the Maxwell-relation-integration procedure outlined above is always valid. In fact, Eqs. (18) and (19) may *always* be used even when there is an ill-defined plateau, or no plateau at all, provided that f_{plateau} is defined to be the average value of the force over the extension interval of interest, i.e., $f_{\text{plateau}}\equiv(\int dX\, f)/\Delta X$.

Differential Maxwell relations are straightforward to obtain for other thermodynamic fields; for example, since temperature changes contribute $S\,dT$ to the thermodynamic potentials discussed here, one can immediately obtain Maxwell relations such as $(\partial S/\partial X)_{\mu,\tau}=-(\partial f/\partial T)_{\tau,X}$. Integration of this relation leads to the Clausius-Clapeyron-like relations discussed in Refs. [39–42].

IV. DNA MOLECULE CONTROLLED BY FORCE, LINKING NUMBER, AND PROTEIN CONCENTRATION

In magnetic-tweezer experiments, the force, chemical potential, and linking number are all controlled and can be constant. The partition sum for this ensemble should be carried out subject to the constraint of fixed Θ, to obtain $Z_M(f,\mu,\Theta)$. The expectation values of extension, bound protein number, and torque follow as

$$\langle X\rangle=\left(\frac{\partial k_BT\ln Z}{\partial f}\right)_{\mu,\Theta},\quad \langle N\rangle=\left(\frac{\partial k_BT\ln Z}{\partial\mu}\right)_{f,\Theta},$$

$$\langle\tau\rangle=-\left(\frac{\partial k_BT\ln Z}{\partial\Theta}\right)_{f,\mu}. \tag{20}$$

These expectation values are summarized by the differential

$$d(k_BT\ln Z_M)=\langle X\rangle df+\langle N\rangle d\mu-\langle\tau\rangle d\Theta. \tag{21}$$

The Maxwell relations in this case are:

$$\left(\frac{\partial\langle X\rangle}{\partial\mu}\right)_{f,\Theta}=\left(\frac{\partial\langle N\rangle}{\partial f}\right)_{\mu,\tau},\quad \left(\frac{\partial\langle N\rangle}{\partial\Theta}\right)_{f,\mu}=-\left(\frac{\partial\langle\tau\rangle}{\partial\mu}\right)_{f,\Theta},$$

$$-\left(\frac{\partial\langle\tau\rangle}{\partial f}\right)_{\Theta,\mu}=\left(\frac{\partial\langle X\rangle}{\partial\Theta}\right)_{f,\mu}. \tag{22}$$

The first relation of (22) can be used to measure changes in numbers of proteins bound with force, by performing an integral of the extension rate with chemical potential via a fixed-Θ version of Eq. (5). The second relation of (22) tells us the connection between bound protein number changes with linking number and torque changes with chemical potential. If torque changes with protein concentration can be measured, one would therefore know the linking number change contributed per protein bound.

We note the last relation of (22), which is usable at constant protein concentration, and also can be used to measure torque changes in naked DNA in the absence of any protein. This final relation allows extension changes with linking number, easily measured in magnetic-tweezer experiments, to be used to obtain torque differences:

$$\tau(f,\mu,\Theta) = \tau(f_0,\mu,\Theta) - \int_{f_0}^{f} df' \left(\frac{\partial\langle X\rangle}{\partial\Theta}\right)_{\mu,f'}. \qquad (23)$$

It is a challenge to find a reference state with absolute known torque; the zero-force state has remnant torque at all values of Θ except for the not terribly useful fully relaxed case where $\tau=0$. Nevertheless, the ability to accurately measure absolute torque changes should allow one to use magnetic-tweezer experiments to test specific microscopic models of double-helix statistical mechanics [22,23].

Finally, we also note that, when naked DNA is stretched and twisted, it can undergo a kind of internal phase separation into extended and "plectonemically supercoiled" domains. This domain formation can be analyzed analogously to the cooperative protein binding discussed in Sec. III B. During formation of plectonemic domains by changing linking number Θ at fixed force, the torque must be constant [22,23], and therefore by (22) $\langle X\rangle$ must change linearly with linking number [22].

V. CONCLUSIONS

This paper has discussed Maxwell relations for the general problem of a polymer which is stretched and twisted, and to which other smaller molecules are able to bind. More generally, these results apply to any elastic object with additional Ising-like internal degrees of freedom (the binding variables n_i), to which an external field (μ) is somehow coupled. The main application that we have discussed is analysis of mechanically constrained double-helix DNA in solution with protein molecules, which distort the double helix when they bind to it. We have shown how Maxwell relations can be used to measure changes in the total number of proteins that bind to a DNA molecule, by measuring how the polymer elasticity changes as bulk protein concentration is changed; the most useful relations in this regard are likely to be (5) and its fixed-extension counterpart (15). Given that direct measurements of numbers of proteins binding to a DNA molecule range from difficult to inaccurate to impossible, the procedure outlined here might prove valuable.

In experiments where DNA twisting is controlled, we have also shown how measurements of extension as a function of linking number can be used to measure torque changes with force. Over the past few years, magnetic-tweezer experiments have become quite widely used to twist and stretch DNA, but direct measurement of torque is quite difficult in such experiments [18,19]. One approach that has been quite widely used is *estimation* of the torque in DNA under a given force, which requires some degree of microscopic theory [22,23]. The method described here can be highly complementary to modeling approaches, since (23) allows measurement of absolute torque changes with almost no theoretical suppositions beyond existence of an equilibrium ensemble.

If the relationship between force, torque, extension, and linking number can be determined at a series of protein concentrations, then the extension and linking number changes per protein that binds are determined. Equation (14) immediately tells us the extension change per protein bound at constant protein concentration and torque, $(\partial X/\partial N)_{\mu,\tau} = -(\partial\mu/\partial f)_{X,\tau}$. In the same way, Eq. (22) determines the linking number change per protein bound at constant concentration and force, $(\partial\Theta/\partial N)_{f,\mu} = -(\partial\tau/\partial\mu)_{f,\Theta}$.

All these applications to study protein-DNA interactions presuppose experiments where the binding of the protein (or ligand) being studied is reversible over the range of concentrations of interest. However, in some cases it has been observed that, for sufficiently large concentrations, the binding of proteins to DNA can become irreversible, possibly because of cooperative protein-protein interactions [9,15]. In such situations, thermodynamic measurements involving different protein binding states are not possible.

However, in situations where proteins cease to dissociate, an alternative approach may be possible, based on the observation that DNA fragments in solution are able to remove proteins from complexes that are entirely stable in DNA-free solution [9,15]. Given this, small DNA segments in solution with protein may allow protein binding to equilibrate, by providing efficient protein-transfer kinetic pathways. In this case, the thermodynamic methods discussed here are usable, as long as reversibility of the observable quantities as a function of protein (and DNA fragment) solution is observed. This type of medium may in fact be more biologically relevant than DNA-fragment-free solutions; the situation found inside a cell nucleus, or a bacterial nucleoid, is one where both DNA and many species of protein are present at high concentrations.

This paper has considered only one species of protein or ligand binding to the double helix. It is straightforward to generalize the results presented here to the case where two or more species of molecules that bind to DNA are present in solution. In such a situation, experiments varying the bulk concentrations of each molecule species could be used to measure bound amounts, through relations of the form (5) defined for each species. It is also worth noting that the response functions in the Maxwell relations of this paper all have statistical interpretations as correlators of microscopic variables. For example, the first Maxwell relation of (22) is equal to $\beta(\langle XN\rangle - \langle X\rangle\langle N\rangle)$.

Although the thermodynamic approach outlined in this paper is useful in the absence of microscopic models, it is likely that the use of thermodynamic relations combined with microscopic statistical models will be desirable. The simple noninteracting model presented in Sec. II D provides a starting point for such studies, with many elaborations of it possible. Energetic cooperativity or sequence inhomogeneity effects could easily be added, as could a more realistic description of protein-driven DNA deformations such as DNA bending and DNA twisting. All such additions are within the scope of the general binding model presented in Eq. (1).

The Maxwell relations and their applications described here may have applications to other types of soft matter or

biophysical situations where stress response and molecular binding degrees of freedom are coupled. More closely related to this paper, force-modulated protein-binding effects may occur along cytoskeletal filaments, and could provide stress-sensitive regulatory responses; a thermodynamic formalism similar to that used here could be applicable to analysis of such a situation. Similarly, proteins adsorbed to cell membrane surfaces that affect that membrane's mechanical properties could have their binding modulated through bilayer stress; their behavior could be analyzed using methods similar to those described in this paper.

ACKNOWLEDGMENTS

We thank Professor Mark Schlossman for helpful discussions. This research was supported by NSF Grant No. DMR-0715099.

[1] L. Finzi and J. Gelles, Science **267**, 378 (1995).
[2] J. F. Marko and E. D. Siggia, Biophys. J. **73**, 2173 (1997).
[3] S. Cocco, J. F. Marko, R. Monasson, A. Sarkar, and J. Yan, Eur. Phys. J. E **10**, 249 (2003).
[4] J. Yan and J. F. Marko, Phys. Rev. E **68**, 011905 (2003).
[5] B. M. Ali, R. Amit, I. Braslavsky, A. B. Oppenheim, O. Gileadi, and J. Stavans, Proc. Natl. Acad. Sci. U.S.A. **98**, 10658 (2001).
[6] R. Amit, A. B. Oppenheim, and J. Stavans, Biophys. J. **84**, 2467 (2003).
[7] G. Lia, D. Bensimon, V. Croquette, J. F. Allemand, D. Dunlap, D. E. Lewis, S. Adhya, and L. Finzi, Proc. Natl. Acad. Sci. U.S.A. **100**, 11373 (2003).
[8] J. van Noort, S. Verbrugge, N. Goosen, C. Dekker, and R. T. Dame, Proc. Natl. Acad. Sci. U.S.A. **101**, 6969 (2004).
[9] D. Skoko, B. Wong, R. C. Johnson, and J. F. Marko, Biochemistry **43**, 13867 (2004).
[10] B. Schnurr, C. Vorgias, and J. Stavans, Biophys. Rev. Lett. **1**, 29 (2006).
[11] S. F. Tolic-Norrelykke, M. B. Rasmussen, F. S. Pavone, K. Berg-Sorensen, and L. B. Oddershede, Biophys. J. **90**, 3694 (2006).
[12] J. Mameren, M. Modesti, R. Kanaar, C. Wyman, G. Wuite, and E. J. Peterman, Biophys. J. **91**, L78 (2006).
[13] M. L. Bennink, S. H. Leuba, G. H. Leno, J. Zlatanova, B. G. de Grooth, and J. Greve, Nat. Struct. Biol. **8**, 606 (2001).
[14] D. Skoko, J. Yan, R. C. Johnson, and J. F. Marko, Phys. Rev. Lett. **95**, 208101 (2005).
[15] D. Skoko, D. Yoo, H. Bai, B. Schnurr, J. Yan, S. M. McLeod, J. F. Marko, and R. C. Johnson, J. Mol. Biol. **364**, 777 (2006).
[16] D. C. Rau and V. A. Parsegian, Biophys. J. **61**, 260 (1992).
[17] V. A. Parsegian, R. P. Rand, and D. C. Rau, Proc. Natl. Acad. Sci. U.S.A. **97**, 3987 (2000).
[18] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, Nature (London) **424**, 338 (2003).
[19] C. Deufel, S. Forth, C. R. Simmons, S. Dejgosha, and M. D. Wang, Nat. Methods **4**, 223 (2007).
[20] T. R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996).
[21] T. R. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, Biophys. J. **74**, 2016 (1998).
[22] T. R. Strick, J.-F. Allemand, V. Croquette, and D. Bensimon, Prog. Biophys. Mol. Biol. **74**, 115 (2000).
[23] J. F. Marko, Phys. Rev. E **76**, 021926 (2007).
[24] R. Nambiar, A. Gajraj, and J.-C. Meiners, Biophys. J. **87**, 1972 (2004).
[25] M. Steiniger, C. D. Adams, J. F. Marko, and W. S. Reznikoff, Nucleic Acids Res. **34**, 2820 (2006).
[26] C. D. Adams, B. Schnurr, D. Skoko, J. F. Marko, and W. S. Reznikoff, Mol. Microbiol. **62**, 1558 (2006).
[27] C. D. Adams, B. Schnurr, J. F. Marko, and W. S. Reznikoff, J. Mol. Biol. **367**, 319 (2007).
[28] A. Ahsan, J. Rudnick, and R. Bruinsma, Biophys. J. **74**, 132 (1998).
[29] H. Diamant and D. Andelman, Phys. Rev. E **61**, 6740 (2000).
[30] J. Yan, R. Kawamura, and J. F. Marko, Phys. Rev. E **71**, 061905 (2005).
[31] L. D. Landau and E. M. Lifshitz, *Statistical Mechanics* (Pergamon, New York, 1978), Sec. 145.
[32] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).
[33] S. B. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[34] J.-F. Léger, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, Proc. Natl. Acad. Sci. U.S.A. **95**, 12295 (1998).
[35] M. Ueda and K. Yoshikawa, Phys. Rev. Lett. **77**, 2133 (1996).
[36] J. F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, Phys. Rev. Lett. **83**, 1066 (1999).
[37] A. Sarkar, J. F. Léger, D. Chatenay, and J. F. Marko, Phys. Rev. E **63**, 051903 (2001).
[38] S. Leikin, D. C. Rau, and V. A. Parsegian, Phys. Rev. A **44**, 5272 (1991).
[39] I. Rouzina and V. A. Bloomfield, Biophys. J. **80**, 882 (2001).
[40] I. Rouzina and V. A. Bloomfield, Biophys. J. **80**, 894 (2001).
[41] M. C. Williams, J. R. Wenner, I. Rouzina, and V. A. Bloomfield, Biophys. J. **80**, 1932 (2001).
[42] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, Biophys. J. **82**, 3160 (2002).
[43] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).