Eur. Phys. J. E **10**, 249–263 (2003)
DOI 10.1140/epje/i2002-10113-2

THE EUROPEAN PHYSICAL JOURNAL E

Force-extension behavior of folding polymers

S. Cocco[1], J.F. Marko[2,a], R. Monasson[3], A. Sarkar[2], and J. Yan[2]

[1] LDFC CNRS-ULP, Institut de Physique, 3 rue de l'Université, 67084 Strasbourg, France
[2] Department of Physics, University of Ilinois at Chicago, 845 West Taylor Street, Chicago IL 60607-7059, USA
[3] CNRS-Laboratoire de Physique Theorique de l'ENS, 24 rue Lhomond, 75231 Paris 05, France

Received 7 August 2002 and Received in final form 7 March 2003 /
Published online: 15 April 2003 –

Abstract. The elastic response of flexible polymers made of elements which can be either folded or unfolded, having different lengths in these two states, is discussed. These situations are common for biopolymers as a result of folding interactions intrinsic to the monomers, or as a result of binding of other smaller molecules along the polymer length. Using simple flexible-chain models, we show that even when the energy ϵ associated with maintaining the folded state is comparable to $k_{\rm B}T$, the elastic response of such a chain can mimic usual polymer linear elasticity, but with a force scale enhanced above that expected from the flexibility of the chain backbone. We discuss recent experiments on single-stranded DNA, chromatin fiber and double-stranded DNA with proteins weakly absorbed along its length which show this effect. Effects of polymer semiflexiblity and torsional stiffness relevant to experiments on proteins binding to dsDNA are analyzed. We finally discuss the competition between electrostatic self-repulsion and folding interactions responsible for the complex elastic response of single-stranded DNA.

PACS. 87.14.Gg DNA, RNA – 87.37.Rs Single molecule manipulation of proteins and other biological molecules – 81.16.Fg Supramolecular and biochemical assembly – 36.20.Ey Conformation (statistics and dynamics)

1 Introduction

The initial linear elastic response of many polymer materials is due to the work done against reduction of flexible polymer entropy [1]. For a single chain, the characteristic force associated with this "entropic elasticity" is $k_{\rm B}T/b$, where b is the statistical segment length of the polymer in question. For the usual polymers of chemical physics, with a highly flexible singly bonded backbone and under solvated conditions, $b \approx 0.4$ nm, giving an entropic elasticity force scale ≈ 10 pN.

Many biological polymers and fibers have properties which undermine simple entropic elasticity as the cause of their basic elastic response. Many biopolymers have long statistical segment lengths (relative to flexible chemical polymers), and are characterized as "semiflexible" polymers with a backbone bending modulus B and a persistence length $A = B/k_{\rm B}T$. Familiar examples include dsDNA with $b = 2A \approx 100$ nm [2], chromatin fiber with $b \approx 60$ nm [3], and actin filaments with $b \approx 20$ microns [4]. Such stiff filaments have a much lower scale for entropic elasticity < 0.1 pN. For these stiff polymers, other physical factors can easily replace this weak force scale.

In this paper we consider one such factor, the "folding" of a polymer by self-interactions along its contour. Such interactions might be intrinsic to the polymer itself, as is the case for single strands of nucleic acids [5–10] or polypeptide chains (Fig. 1a-b). Alternately, the folding effects we consider might be a result of interactions between additional molecules absorbed onto the polymer contour (Fig. 1c), as is the case for certain proteins which bind onto double-stranded DNA [11–15]. For either of these cases, we focus on interactions which are specific enough to lead to local folding of the polymer without leading to overall chain collapse. While rare in chemical physics where interactions are relatively generic (*e.g.*, hydrophobicity of entire monomers), this is typical for the spatially specific interactions found in molecular biology (*e.g.*, hydrogen bonding at specific points on amino acid or nucleic acid monomers).

In Section 2 we introduce a simple, exactly solvable model of a completely flexible ("freely jointed") polymer [16] which is composed of elements which can be in a folded (short) or unfolded (long) state. We show how, even when the folding interactions are weak, an elastic response can be observed which is at a larger force scale than the flexible-polymer entropic elasticity intrinsic to the backbone. This model includes cases where there is

[a] e-mail: jmarko@uic.edu

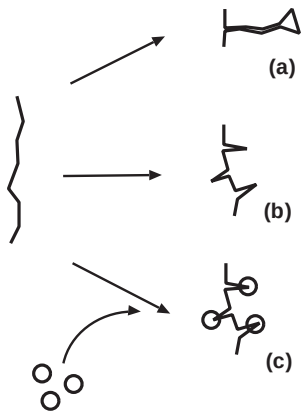

Fig. 1. Folding of polymers. A polymer may be folded by specific interactions between monomers (*e.g.*, base-pairing along nucleic acids) so as to form "zippered" domains as in (a); alternately, the folds may be localized bends that independently bind adjacent segments as in (b). (c) Local bends may alternately be produced by binding by an additional molecule which distorts the polymer backbone (*e.g.*, proteins which bind to and bend dsDNA).

more than one folding element per statistical segment, and where the folding elements in a statistical segment length "unzip" instead of unfolding independently. We then discuss how unfolding elastic response appears in recent experiments on single-stranded DNA (ssDNA) [5,7,17] and on "chromatin fibers", the DNA-protein composite fibers found in eukaryote chromosomes [3,18,19]. Although less fundamental than detailed statistical-mechanical models which consider all possible base-paired structures [9,10], our theory applied to single-stranded nucleic acids provides a simple analytical description of their elasticity.

In Section 3 we present a model incorporating semiflexible polymer behavior, which can be quantitatively applied to experiments where the underlying polymer is double-stranded DNA (dsDNA). Qualitatively, the results match those of the freely jointed model of Section 2, but semiflexible chain elasticity is necessary to make quantitative contact with recent protein+DNA experiments by Ali *et al.* [14].

Two further developments of the semiflexible polymer model are discussed at the end of Section 3. First (Sect. 3.5), for the case of proteins binding to and folding dsDNA, the effect of twisting the DNA is discussed. By monitoring the stretching elasticity as a function of twisting, experiments can study how DNA linking number is altered by the proteins which fold it. Second (Sect. 3.6), Debye-Huckel–type electrostatic interactions are added, providing a model for ssDNA over a wide range of salt conditions. Due to the high degree of flexibility of the ssDNA backbone, modification of screening can drastically affect the chain elasticity. We show how the strongly salt-dependent elastic response observed experimentally [5,7, 17] and recently studied using computer simulations [8] can be analytically understood.

2 Freely jointed polymer with folding interactions

We consider a flexible polymer containing "folding elements" which can be organized into either a folded form of length d_0, or unfolded into a longer form of length d_1. The free-energy difference between the unfolded and folded states of an element is taken to be $\epsilon k_{\mathrm{B}}T$. We also suppose that a statistical segment of the polymer contains M folding elements, and that the folding elements in the different statistical segments along the chain are able to fold and unfold independently. The whole chain is supposed to contain N statistical segments, and thus NM folding elements.

The folding may be the result of direct polymer-polymer interactions, or the result of interactions moderated by an additional type of molecule adsorbed from the surrounding solution (*e.g.*, proteins binding to dsDNA). In the case where folding is mediated by a binding event, the folded state will be assumed to include a bound molecule, with the factor ϵ representing the total chemical potential difference between bound and unbound molecules, *i.e.* $\epsilon = \ln(c/K_{\mathrm{d}})$ where K_{d} is the dissociation constant of the protein to the folded polymer and c is the bulk concentration of the adsorbing molecule (Mol/litre units are usually used for K_{d} and c in the biochemical literature). Increasing ϵ corresponds to stronger binding.

The degrees of freedom per statistical segment are therefore M folding element variables n_i, $i = 1, \cdots, M$ which are 0 or 1 for folded or unfolded states, respectively, the elastic stretching of each segment, plus the segment orientation. For our purposes, the orientation may be described by the angle θ between the segment and the z-axis along which an external force acts to stretch the chain. Elastic stretching of the segment is described by an additional length change y, giving a total statistical segment length of $Md_0 + [d_1 - d_0](n_1 + n_2 + \cdots + n_M) + y$.

In this model, the number of statistical segments along the chain is fixed. This amounts to the approximation that we consider no change of flexibility at the boundary between statistical segments. Although, in general, one expects a change in chain flexibility as the segments unfold (note that Ref. [13] discusses effects of changes in chain bending driven by local protein binding), we want to focus on just the effect of segment length reduction. In the calculation to follow, only a single segment needs to be considered. The total contour length, and the length of each segment, depend on the unfolding degrees of freedom. We take the force to be applied in the z-direction.

The Boltzmann factor for one statistical segment is

$$\exp\Big\{\beta M d_0 t f + (\beta t f[d_1 - d_0] - \epsilon)\,[n_1 + \cdots + n_M] - \frac{\beta f_0}{2Md_1} y^2 + \beta f t y\Big\}, \tag{1}$$

where $t = \cos\theta$ (θ is the polar angle between the segment and the z-axis), f is the externally applied force, and $\beta = (k_{\mathrm{B}}T)^{-1}$. The elastic constant f_0 has the dimensions of a force. In general, the folded and unfolded elements

will contribute differently to the net segment stretching, but again to focus on the segment length-change effect we simplify our calculation by considering the case where the segment stretching is independent of its folding state. Both this, and the assumption of no change in the total number of statistical segments during unfolding, greatly simplify the analysis of the model, while preserving the essential feature of segment unfolding. Thus, we will show how the unfolding —the lengthening of the segments alone— affects the overall chain elasticity.

Integration of the length fluctuations ($\int \mathrm{d}y$) reduces the partition function per segment to

$$Z = \sqrt{\frac{2\pi M d_1}{\beta f_0}} \sum_{\{n\}} \int_{-1}^{+1} \frac{\mathrm{d}t}{2} \exp\left\{ \frac{1}{2} \frac{\beta M d_1}{f_0} f^2 t^2 + \beta M f d_0 t + (\beta t f [d_1 - d_0] - \epsilon)\,[n_1 + \cdots + n_M] \right\}. \quad (2)$$

Next, we carry out the sum on the unfolding variables n_i. In this paper we consider two models. In the first one, the "independent" model, each unfolding variable is independent of the others, and the sum is over 2^M states. In the second, "zipper" model, the unfolding variables open and close in a definite sequence, and therefore the sum is over only $M+1$ states. The "independent" model is suitable for considering a series or string of folded elements, while the "zipper" describes unfolding of hairpin structures such as those formed by nucleic acids, or the unfolding of protein domains where there is a defined pathway for chain unfolding.

After carrying out these summations, the partition functions for the independent and zipper cases are:

$$Z_{\mathrm{ind}} = \sqrt{\frac{2\pi M d_1}{\beta f_0}} \int_{-1}^{1} \frac{\mathrm{d}t}{2} \exp\left[\frac{1}{2} \beta M d_1 f^2 t^2 / f_0 + \beta M d_0 f t \right] \left(1 + e^{\beta f t [d_1 - d_0] - \epsilon} \right)^M ,$$

$$Z_{\mathrm{zip}} = \sqrt{\frac{2\pi M d_1}{\beta f_0}} \int_{-1}^{1} \frac{\mathrm{d}t}{2} \exp\left[\frac{1}{2} \beta M d_1 f^2 t^2 / f_0 + \beta M d_0 f t \right] \left(\frac{1 - e^{(M+1)[\beta f t [d_1 - d_0] - \epsilon]}}{1 - e^{\beta f t [d_1 - d_0] - \epsilon}} \right) . \quad (3)$$

The extension (projection along the force direction) *versus* force of each segment of M unfolding elements can be computed using

$$\langle z_{\mathrm{segment}} \rangle = \langle (y + M d_0 + [d_1 - d_0][n_1 + \cdots + n_M]) t \rangle = \frac{\partial \ln Z}{\partial (\beta f)} . \quad (4)$$

The total extension of a chain of multiple segments is N times this result, since the segments are independent. We will also report extension per segment, as a fraction of the extended segment length $M d_1$ (equivalently the total chain end-to-end extension as a fraction of the maximum extended length $N M d_1$).

This model shows initial linear response of extension to applied force, and the proportionality "elastic" constant combines usual flexible-polymer entropic elasticity, with contributions from chain folding. The end-to-end extension is therefore $\langle z \rangle = N \langle z_{\mathrm{segment}} \rangle = N \beta d^2 f/3 + \mathcal{O}(f^2)$, where the coefficient d is a length, and for a freely jointed chain without folding interactions would be the statistical segment length. With folding interactions, the effective segment length is in between the two segment lengths d_0 and d_1: for independent and zipper folding we have

$$d^2_{\mathrm{ind}} = M^2 d_0^2 + \frac{M d_1 k_{\mathrm{B}} T}{f_0} + \frac{M (d_1 - d_0)(d_1 + [2M-1] d_0)}{1 + e^{\epsilon}} + \frac{M(M-1)[d_1 - d_0]^2}{(1 + e^{\epsilon})^2},$$

$$d^2_{\mathrm{zip}} = M^2 d_0^2 + \frac{M d_1 k_{\mathrm{B}} T}{f_0} + 2 M d_0 [d_1 - d_0] \times \left[\frac{1}{e^{\epsilon} - 1} - \frac{M+1}{e^{[M+1]\epsilon} - 1} \right] + [d_1 - d_0]^2 \left[\frac{2}{(e^{\epsilon} - 1)^2} + \frac{e^{[M+1]\epsilon} - 2M - 3}{(e^{\epsilon} - 1)(e^{[M+1]\epsilon} - 1)} - \frac{(M+1)^2}{e^{[M+1]\epsilon} - 1} \right] . \quad (5)$$

For $M = 1$ (one unfolding unit per segment) the independent and zipper models are the same, with the form

$$d^2 = d_0^2 + \frac{d_1 k_{\mathrm{B}} T}{f_0} + \frac{d_1^2 - d_0^2}{1 + e^{\epsilon}} . \quad (6)$$

Note that d is the apparent segment length that would be inferred from a low-force elastic-response measurement, *i.e.* where the initial elastic response was interpreted in terms of the standard Gaussian polymer elastic response $f = (3 k_{\mathrm{B}} T / [N d^2]) z + \mathcal{O}(z^2)$.

Equations (5) and (6) show that the apparent segment length squared, d^2, is a sum of terms with different physical origins. First, the d_0^2-term is the direct flexible-polymer contribution associated with the folded segment length. Second, the segment stretching elasticity adds a simple elastic contribution $\propto k_{\mathrm{B}} T d_1 / f_0$. The final contributions are from the unfolding degrees of freedom giving a length change weighted by the Boltzmann factor for unfolding at zero force. When d_0 is not much smaller than d_1, then the low-force linear response will be mainly due to usual flexible-polymer elasticity, plus corrections from the other stretching and unfolding degrees of freedom. In this case, the additional degrees of freedom simply reduce the effective spring constant $3 k_{\mathrm{B}} T / (N M d^2)$ of the polymer.

However, in the situation where d_0 is relatively small or zero (*e.g.*, when the segments fold into compact structures such as nucleic acid hairpins), and ϵ is not too large, then the linear elasticity may be determined mainly by the unfolding terms. In such a situation, at low forces, the effective segment length inferred from a force experiment (*i.e.* d) may be inconsistent with the actual structural segment length. The apparent d from low-force data may also be inconsistent with the apparent segment length

measured at higher forces; for example, beyond the unfolding point (for $f \gg k_{\rm B}T\epsilon/[d_1 - d_0]$) the apparent segment length in the $M = 1$ model would be d_1.

We now consider three experimental systems to which this model is relevant.

2.1 Chromatin fiber

First, we consider recent experiments of Cui and Bustamante [3] on chromatin fiber, the basic structure underlying eukaryote chromosomes. This fiber is made up of double-stranded DNA (dsDNA) wrapped around complexes of histone proteins, to form a "beads-on-a-string" structure. These $\approx$ 10 nm diameter protein-DNA beads ("nucleosomes") are separated by $\approx$ 8 nm long stretches of dsDNA. With no tension applied to this chain, and under physiological solution conditions (pH 7.5, $>$ 30 mM Na^+ or K^+ the nucleosomes associate with one another to form a "condensed" fiber $\approx$ 30 nm in diameter.

Experiments [3] at moderate ionic strength show two low-force elastic regimes, a very-low-force entropic regime ($f < 0.5$ pN), where thermal bending fluctuations of the condensed fiber are removed, and then an almost linear increase in force to 5 pN over which the fiber is about doubled in length (see inset, Fig. 2). The low-force (< 0.1 pN) behavior is similar to that of a flexible polymer with statistical segment length $\approx$ 60 nm, containing perhaps 20 nucleosomes. The length doubling at larger forces is most simply explained by the opening of adjacent nucleosome-nucleosome contacts [3].

The "independent unfolding" model shows a behavior similar to the experimental data, using parameters $\epsilon = 4$, $d_0 = 5$ nm, $d_1 = 10$ nm, $f_0 = 50$ pN and $M = 20$ (main Fig. 2, solid line). In this case, there is an initial low-force regime where the orientational degrees of freedom are polarized, followed by a gradual opening of the nucleosome fiber. The initial elastic response is in the regime where the first term of (6) is dominant; at larger forces $\approx$ 5 pN, a nearly linear elastic response is observed as the chromatin fiber opens up, with an effective force constant roughly given by the final term of (6). This interpretation is consistent with that discussed by Cui and Bustamante [3] based on a previously published model [11]. Finally, at large forces $>$ 5 pN, the maximum extensibility and intrinsic elasticity (second term of (6)) of the fiber is seen; in this regime, the experimental force curves show appreciable hysteresis, but have the approximate shape given by the reversible model. It is possible that a reversible force response would be obtained given a sufficiently slow force-relaxation cycle.

The dashed curve of Figure 2 shows the force-extension curve using the same parameters, but now for the "zipper" model. The zipper model effectively introduces a strong coupling between the unfolding degrees of freedom, resulting in a sigmoidal transition. This type of transition is not consistent with the chromatin unfolding data.

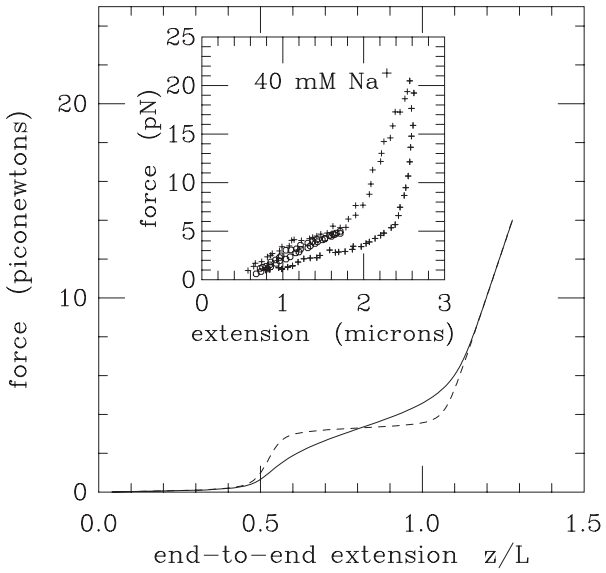

Fig. 2. Force-extension curves for chromatin fiber. The main figure shows theoretical curves; the inset shows experimental data from Figure 4A and C of reference [3]. In experiment (inset), for small extensions (○) a nearly reversible elastic response is observed, while for large extensions (+) a hysteresis loop is observed. The reversible "independent" segment-unfolding theory of Section 2 can generate the reversible response and the qualitative shape of the irreversible response (solid curve). The "zipper" model for the same parameters (dashed curve) generates a force "plateau" not observed experimentally.

2.2 Single-stranded DNA (ssDNA)

The previous example showed an initial, weak entropic elasticity regime, followed by a stronger elastic response associated with opening of weakly folded structures, which will be typical for polymers which in their initially folded state, still have an extended fiber structure. However, some biological polymers will fold in such a way as to reduce their end-to-end length essentially to zero. An important example of this are single-stranded nucleic acids, which can form "hairpin" or "helix-loop" secondary structure.

Recent studies of elastic response of long ssDNAs have indicated that under conditions where the double helix can base-pair (*i.e.* moderate ionic strength, physiological pH), the force needed to extend the polymer goes to a nearly constant value $\approx$ 2 pN at zero extension (Fig. 3, circles) [7,17]. This effect is thought to be due to local association of the hydrophobic and hydrogen-bonding bases along the ssDNA contour [8,20]. The (weakly sequence-dependent) 2 pN force needed to open up the ssDNA is far less than the $\approx$ 15 pN needed to open $>$ 10 bp DNA helix-loop structures [17] indicating that the self-associated regions are relatively weakly bound. This is reasonable since long self-complementary regions, which again require $\approx$ 15 pN for opening [21], are not present along large genomic ssDNAs [7,17].

The main features of the low-force experimental data can be qualitatively understood with nothing more than

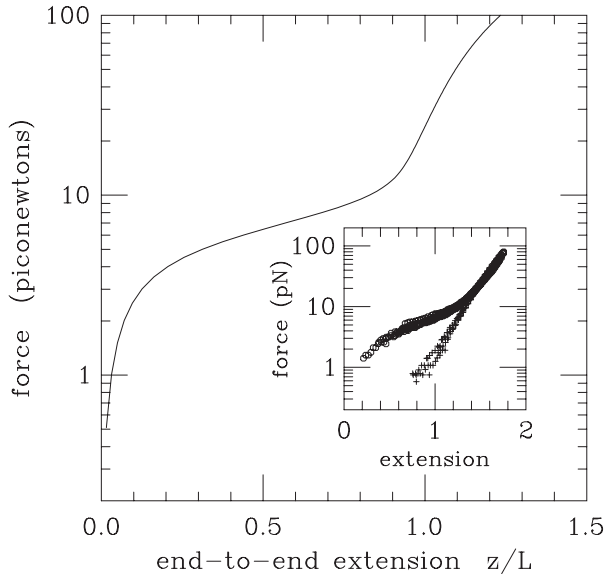

Fig. 3. Force-extension behavior of single-stranded DNA. The main figure shows the polymer-folding model of Section 2, applied to hairpin formation expected for ssDNA at high ionic strength and low force. Extension is relative to fully extended length L of ssDNA backbone in the model, but since there is linear stretching elasticity, the ssDNA backbone can be appreciably extended beyond $z/L = 1$. The solid line shows the zipper model (see text). At low force, the folding interactions reduce extension to zero; extension of the chain requires ≈ 2 pN of force. The inset shows experimental data from reference [17], for ssDNA extension at high ionic strength ($\circ$, 150 mM Na^+) where hairpin formation occurs at $\approx$ pN forces as in the folding model. Data for low ionic strength (+, 2.5 mM Na^+) show a different low-force elasticity with no tendency for the chain to fold up for ≈ 1 pN forces; the behavior at low ionic strength is discussed in Section 3.4. Note that these experimental data show extension relative to B-DNA length of 3.4 Å per bp; this corresponds to an extension of about 0.75 in the main figure.

some weak local self-attraction, plus main-chain flexibility and elasticity. The "zipper" model, with $\epsilon = 0.2$, $d_0 = 0$, $d_1 = 0.7$ nm, $f_0 = 220$ pN and $M = 3$ produces the solid curve of Figure 3. This simple model reproduces two important features of the experimental data: the force $\approx$ 2 pN needed to extend the chain and the slow subsequent increase of force with extension.

The choice of a zipper model with $M = 3$ is needed to generate the lower-force response; this corresponds to formation of small hairpin structures. These hairpins are well bound enough that the initial elasticity is dominated by the 3rd term of (6). The interplay between the folding and the chain extension is responsible for the wide variation in "segment lengths" of ssDNA when simple flexible polymer models are fit to experimental data [6,5]; this point is well emphasized in reference [8].

A microscopic approach to the local self-association described in our simple model is the formulation of a detailed theory of all possible base-pairing interactions. Work of Gerland *et al.* [9] showed that even for rather short genomic-sequence single-stranded nucleic acids, a smooth (self-averaged) elastic response occurs similar to that obtained here. Other, recent work of Mezard and co-workers, showed that a model of sequence-nonspecific base-pairing can describe the lowest-force response [10]. Our simple model describes this behavior by assuming small hairpin structures to be distributed along a partially extended polymer, which is the situation for ssDNA once it is partially extended (*i.e.* over the range of the experimental data shown in Fig. 3).

Interestingly, ssDNA elasticity has strong salt, or screening length, dependence [17]. For example, the low-force threshold behavior seen at high salt vanishes as salt concentration is reduced. It has been recently shown using Monte Carlo computer calculations that the extended-chain elastic response is in part due to Coulomb self-repulsion along the chain [8]. In Section 3 we will add such interactions to a simple polymer model to analytically describe these effects.

2.3 Mitotic chromosome

Another example of polymer folding comes from recent experiments on the elastic response of whole mitotic chromosomes, in the fully folded "mitotic" form that occurs during metaphase of eukaryote cell division. The elastic response of these objects is peculiar: while their spring constants are too strong to easily attribute to simple entropic elasticity of chromatin fibers [19,22], reversible extensible response occurs over a fivefold range of stretching. This large range of extensibility, plus a slow stress relaxation behavior, suggests that their elastic response is due to gradual unfolding of compacted chromatin fiber domains.

We therefore consider the above model with parameters chosen so that entropic elasticity plays no role in the initial elastic response (*i.e.* the folding is to a completely compacted state): $\epsilon = 2$, $d_0 = 0$, $d_1 = 10$ nm, $f_0 = 50$ pN and $M = 1$. Thus, extension is relative to the initial dense state in which usual polymer elasticity is strongly supressed. As the chromatin domains open, a gradual elastic response is observed which is solely due to unfolding, rather like the ssDNA case (Fig. 4). The "S" shape of the force response is observed experimentally (Ref. [10], Fig. 4, inset). The experimental data involve extension of a large number ($\approx 10^3$) of folded chromatin fibers in parallel, and show a proportionally larger force. It must be noted that the experimental data certainly reflect, in addition to chromatin fiber unfolding, the effect of three-dimensional interfiber interactions, not included in this approach.

3 Folding of semiflexible polymers

The previous sections presented results for a simple completely flexible (freely jointed) model. Its advantage is its simplicity, but its disadvantages include the commensuration of folding (or binding) sites with the flexible segments, the completely freely jointed nature of the segment

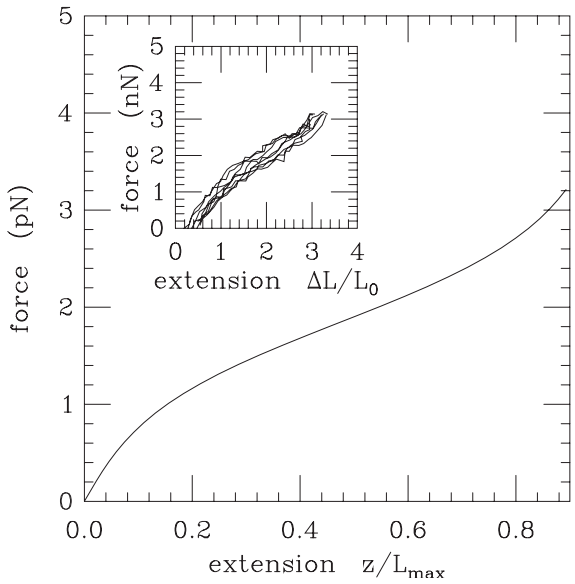

Fig. 4. Force-extension behavior similar to that of mitotic chromosome, which can be thought of as a compacted network of folded chromatin domains. The inset shows experimental data of Poirier *et al.* [19], which have a reversible elastic response over a fivefold range of extension (extension here is the change in length, relative to the native or unstretched length). The large ≈ nN force scale of the experiment comes from the simultaneous extension of many domains together in a bulk chromosome. The main figure shows elastic response of the independent folding model (see text), which displays a similar large range of elastic response. In the model, extension corresponds to the change in length relative to the maximum extended length. The force scale in the main figure refers to a single chromatin domain, and has a characteristic force $\approx 10^{-3}$ of that in the experiment on a whole chromosome.

joints, and the fixed total number of segments. For applications where dsDNA is the basic underlying polymer, it would be advantageous to obtain the semiflexible polymer behavior characteristic of dsDNA [23] when folding is disrupted. Here a simple semiflexible-polymer–based model is presented which can describe some of the situations discussed above.

A long semiflexible polymer of total (fixed) length L whose ends are separated by a force f (taken to be in the z-direction) is described by the partition function

$$Z(L,f)=\int \left\{\mathcal{D}\hat{\mathbf{t}}\right\} \exp\left\{-\int_0^L \mathrm{d}s \left[\frac{A}{2}\left(\frac{\mathrm{d}\hat{\mathbf{t}}}{\mathrm{d}s}\right)^2-\beta f\hat{\mathbf{t}}\cdot\hat{\mathbf{z}}\right]\right\}, \tag{7}$$

where s is the contour length, A is the persistence length, and $\hat{\mathbf{t}}(s)$ is the tangent to the polymer at point s along its contour. In the thermodynamic limit $L/A \to \infty$, this partition function has the form

$$\ln Z(L,f) = \frac{L}{A} g(\beta A f). \tag{8}$$

The end-to-end extension along the force direction is

$$\langle z\rangle = \frac{\partial \ln Z}{\partial(\beta f)} = L g'(\beta A f). \tag{9}$$

The potential $g(x)$ can be computed to sufficient numerical precision for all practical problems as discussed in [23]. The asymptotic behaviors of $g(x)$ are known exactly [23]:

$$\begin{aligned} \lim_{x\to 0} g(x) &= \frac{x^2}{3}, \\ \lim_{x\to\infty} g(x) &= x - \sqrt{x}. \end{aligned} \tag{10}$$

To describe the folding of this type of polymer, we suppose that each folded unit simply removes a certain amount of contour length ℓ from the polymer. If the maximum number of folded units that can occur along the chain is K, and assuming that the folded units simply reduce the chain length L without other effects, the partition function of the chain now with folding degrees of freedom is

$$\begin{aligned} Z_{\mathrm{fold}} = {} & Z(L,f) + W_1 e^{\epsilon} Z(L-\ell,f) \\ & + W_2 e^{2\epsilon} Z(L-2\ell,f) + \cdots \\ & + W_K e^{K\epsilon} Z(L-K\ell,f). \end{aligned} \tag{11}$$

Here, the binding free energy is $\epsilon k_{\mathrm{B}}T$ as in Section 2. The first term describes the "clean" chain, the second term the chain with one fold, the third term with two folds, and so on, up to the maximum number of folds K. The factor W_k is the combinatoric factor equal to the number of ways that the k folded units may be distributed along the chain. The factors W_k can also include effects of interactions (cooperativity) between folding units.

Equation (11) assumes no coupling between polymer bending degrees of freedom and the folding degrees of freedom. Thus, k folds are considered to leave behind a "seamless" piece of semiflexible chain, reduced in length by $k\ell$. It is presumed that not only no permanent bends are induced at fold boundaries, but also that no inhomogeneities in deformability occur at the fold boundaries. Such effects can be accounted for only with substantial complication; as in Section 2, our aim is to obtain as much information as possible by analytical calculation. However, in single-molecule DNA stretching experiments, the gradual reduction in bending fluctuation correlation length ($\xi \approx (A/\beta f)^{1/2}$ [23]) means that such an approach is accurate at low enough concentrations, *i.e.* when the mean distance between proteins exceeds ξ.

Using (8), (11) may be written as

$$Z_{\mathrm{fold}} = \sum_{k=0}^{K} W_k \exp\left\{\left[\frac{L-k\ell}{A}\right] g(\beta A f) + k\epsilon\right\}, \tag{12}$$

where the index k is the number of folded units along the chain.

The main limitation of the model (11) is that the folding and bending degrees of freedom are decoupled (*e.g.*, the bending modes of the semiflexible chain are taken to be those of an infinite chain, even when parts of it are folded up). Folding simply reduces the total length available for the semiflexible chain. This is not to say that interactions between the folding units cannot be taken into account; below, a simple case with interactions is described. Thus,

many situations can already be described by (11), and this model has the virtue of being completely solvable. This model is a simplified version of a theory of compaction of DNA by proteins previously studied using a mean-field approximation [11] and thereby provides a way to check the validity of that theory in some limits. The model of this paper is also related to a model of the B-DNA to S-DNA transition introduced by Ahsan *et al.* [24].

Here we consider two simple models for W_k which correspond to the cases where: a) there are K folding units at fixed locations along the chain, and b) where the folding units can be located at any point along the chain as long as they do not overlap. The first "fixed" model has a W_k which is simply the number of ways to choose k of the units to fold

$$W_k = \frac{K!}{k!(K-k)!} \approx \exp(-K[\psi \ln \psi + (1-\psi)\ln(1-\psi)]). \tag{13}$$

The final term of (13) indicates the thermodynamical limit ($\psi = k/K = k\Delta/L$ is the site occupation fraction) which is familar as the lattice gas entropy [1]. We suppose that the arc length along the chain between fold points is fixed at $\Delta = L/K$.

In the second "sliding" case we suppose that folding units are restricted to be farther apart from one another by more than arc length Δ (thus $\Delta \geq \ell$). If we suppose no other constraints on the folding unit positions, then W_k is the configurational partition function of a one-dimensional continuum gas of k particles each of length Δ, confined to a total length L, which is exactly solved [25]. The result for $L \gg \Delta$ is

$$W_k = \frac{(K-k)^k}{k!} \approx \exp\left[k \ln\left(\frac{1-\psi}{\psi}\right) + k\right]. \tag{14}$$

Again, $\psi = k/K = k\Delta/L$ represents the folded fraction of the chain. Note that this partition function goes to zero when $\psi \to 1$ since in that fully filled limit the particles are fixed in position. For $\psi \to 0$, the entropy per particle is $\approx -k_{\mathrm{B}} \ln \psi$, the usual dilute gas result.

These two cases are relevant to folding of biopolymers. The first "fixed" case is a situation where some feature of the polymer itself directs the folding, *e.g.*, the sequence of nucleotides along a nucleic acid. One might imagine there to be specific binding sites for a protein which then mediates folding the nucleic acid (inhomogeneity in the binding free energy ϵ associated with the different binding sites may easily be included in (12)). The second "sliding" case corresponds to a sequence-nonspecific folding case, for example where a protein can bind equally well to any site along a nucleic acid, and thus meditate its folding at any point along its length.

3.1 Fixed folding sites

Plugging (13) into (12) and carrying out the sum over K gives the partition function

$$Z_{\mathrm{fold}} = e^{Lg}\left[1 + e^{\epsilon - \ell g/A}\right]^K. \tag{15}$$

The end-to-end extension along the force direction is simply

$$\langle z \rangle = \frac{\partial \ln Z_{\mathrm{fold}}}{\partial(\beta f)}, \tag{16}$$

which gives

$$\frac{\langle z \rangle}{L} = \left[1 - \frac{\ell/\Delta}{1 + e^{\ell g/A - \epsilon}}\right] g'(\beta A f), \tag{17}$$

i.e. just the result (9) for the unfolded chain, multiplied by a factor accounting for the reduction in contour length due to folding. This factor includes the probability of sites being folded,

$$\frac{\langle k \rangle}{K} = \frac{1}{K}\frac{\partial \ln Z_{\mathrm{fold}}}{\partial \epsilon} = \frac{1}{1 + e^{\ell g/A - \epsilon}}. \tag{18}$$

In this case, the folding degrees of freedom act as simple two-state systems. Unfolding by force occurs when $\ell g/A \approx \epsilon$. If this unfolding force is at least $k_{\mathrm{B}}T/A$, then $g \approx \beta A f$ and unfolding occurs at the threshold force $f^* \approx \epsilon k_{\mathrm{B}}T/\ell$, in accord with reference [11]. At low force the chain is fully folded, and we have elastic response:

$$\frac{\langle z \rangle}{L} = \left(1 - \frac{\ell}{\Delta}\right)\frac{2Af}{3k_{\mathrm{B}}T} + \mathcal{O}(f^2), \tag{19}$$

i.e. the apparent persistence length appears smaller compared to its "true" value A which would be observed at large enough forces that unfolding were to occur.

If $\epsilon < \ell g/A$ so that unfolding occurs at low forces, then we may use $g \approx (\beta A f)^2/3$, and therefore find

$$\frac{\langle z \rangle}{L} = \left(1 - \frac{\ell/\Delta}{1 + e^{-\epsilon}}\right)\frac{2Af}{3k_{\mathrm{B}}T} + \mathcal{O}(f^2). \tag{20}$$

As in Section 2, the low-force apparent persistence length appears smaller than that obtained from high-force measurements.

3.2 Sliding folding sites

Combining (14) and (12) yields the partition function for sliding folding sites:

$$Z_{\mathrm{fold}} = K e^{Lg/A} \times \int_0^1 \mathrm{d}\psi \exp\left\{-K\psi\left[\ln\psi - \ln(1-\psi) + \ell g/A - 1 - \epsilon\right]\right\}. \tag{21}$$

The sum from (12) has been converted to an integral, which is accurate for $K \gg 1$, the case we are interested in. The exponent of the integrand is proportional to K and therefore can be accurately evaluated as a Gaussian integral by expansion of the exponent around its stationary point.

The stationary point of the exponent occurs at $\psi = \psi_0$ as determined by

$$\ln(\psi_0) - \ln(1-\psi_0) + \frac{\psi_0}{1-\psi_0} + \ell g/A = \epsilon. \tag{22}$$

This equation always has a solution for ψ between 0 and 1, always corresponding to the maximum of the exponent. If the RHS of (22) is large and positive (the binding energy ϵ dominates the chain elastic free energy), then ψ_0 will be close to 1, and the chain will be folded. If the RHS of (22) is large and negative (chain elastic free energy dominant), then ψ_0 will be near zero, and the chain will unfold. The solution of (22) also gives the thermodynamical average of k, $\langle k\rangle / K = \psi_0$.

The extensive part of the partition function is therefore

$$\frac{\ln Z_{\rm fold}}{L} = \frac{g}{A} - \frac{\psi_0}{\Delta}\left[\ln\psi_0 - \ln(1-\psi_0) + \ell g/A - 1 - \epsilon\right]. \quad (23)$$

The end-to-end distance follows via (16) as

$$\frac{\langle z\rangle}{L} = \left[1 - \frac{\ell}{\Delta}\psi_0\right] g'(\beta A f). \quad (24)$$

This is the same form as (17), the only change being the different dependence of $\langle k\rangle$ on applied force.

In the strong-binding limit $\epsilon \gg 1$ and at low forces $f \ll k_{\rm B}T\epsilon/\ell$, (22) indicates that $\epsilon \approx \psi_0/(1-\psi_0)$, and therefore that

$$\frac{\langle z\rangle}{L} = \left(1 - \left[\frac{\epsilon}{1+\epsilon}\right]\frac{\ell}{\Delta}\right)\frac{2Af}{3k_{\rm B}T} + \mathcal{O}(f^2). \quad (25)$$

Again, low-force measurements indicate a smaller value of A than would be inferred from high-force measurements where the chain unfolds.

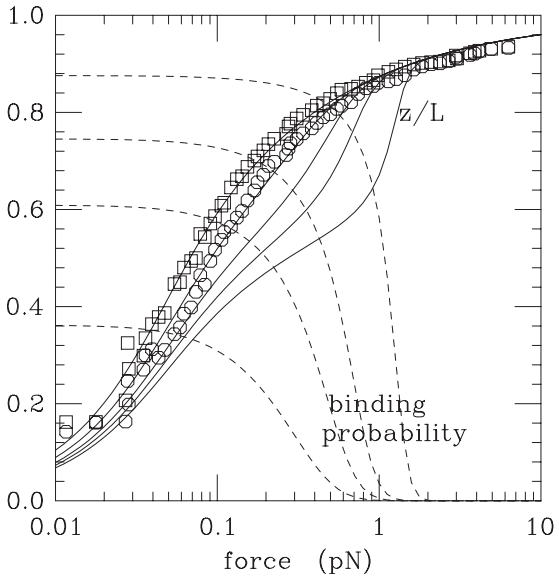

Fig. 5. Effect of DNA-binding proteins on elastic response of dsDNA. The solid curves from top to bottom show extension for the worm-like chain with folding interactions, for $\epsilon = -\infty$ (bare DNA), and $\epsilon = 0$, 2, 4 and 9. As ϵ is increased, successively more force is required to fully extend the DNA, which requires unbinding of the protein. Data for $\epsilon = 0$ and bare DNA fit well to data of reference [14] (□ bare DNA; ○ 1.25 μM IHF). The dashed curves show the IHF binding as a fraction of the maximum possible; from bottom to top, curves for $\epsilon = 0$, 2, 4 and 9 are shown. At a force $\approx \epsilon/(\Delta - \ell)$ the proteins unbind and the chain extends.

3.3 Comparison of freely jointed and semiflexible models

One can compare the results for this model to those of Section 2 as follows. First, the relation between statistical segment lengths is $b \leftrightarrow 2A$. The corresponding lengths of unfolded and folded units are $d_1 \leftrightarrow \Delta$ and $d_0 \leftrightarrow \Delta - \ell$ (similarly, the reductions in chain length per folding unit are related by $d_1 - d_0 \leftrightarrow \ell$). The relation between the total number of folding units per segment of underlying polymer is $M \leftrightarrow 2A/\Delta$. Similarly, the relation between the number of statistical segments along the unfolded chains in the two models is $N \leftrightarrow L/[2A]$.

Although these correspondences allow us to compare results of the freely jointed and semiflexible models, there are two important differences between them. First, in the model of Section 2, there were always a fixed number of statistical segments, while in the model of this section the persistence length (segment length) stays fixed while the total length and therefore the number of statistical segments are reduced by folding. Second, in Section 2, there was at least one folding unit per statistical segment length, while the model of this section can have arbitrary spacing of folding units.

3.4 dsDNA-folding proteins

A recent experiment of Ali *et al.* [14] carried out the force-distance measurement corresponding to the theory of the previous section. IHF protein was put into solution with a single DNA molecule that was under mechanical control, via a magnetic bead attached to one end. By measurement of the fluctuations of the bead, the force applied to it by a nearby permanent magnet could be continuously monitored. Ali *et al.* observed that the IHF reduced the extension of the DNA (Fig. 5, circles) relative to "bare" DNA (without IHF present, Fig. 5, squares). The shift of the +IHF data on the log force axis relative to the "bare" DNA data is the basic elasticity-enhancing (persistence-length-reducing) effect discussed a few times above.

Based on a previously developed theory [11] which is an approximate version of the calculations discussed in the previous subsections, Ali *et al.* determined that in their buffer without IHF, dsDNA had a persistence length of 65 nm. From fitting to the theory of reference [11], the effective size of a binding site was determined to be $\Delta =$ 98 nm. This site size is larger than the actual "footprint" of IHF on dsDNA, which is thought to be roughly five helix repeats (52 bp), or about 20 nm.

The length reduction per binding site was found to be $\ell = 39$ nm [14]. This reduction in length is in accord with the fact that IHF is known to bend dsDNA through greater than 90° (in its sequence-specific binding mode, IHF can bend DNA through nearly 180°). It is to be noted that in this experiment on a large inhomogeneous-sequence dsDNA, IHF most likely is binding in a mixture of modes, with a few strong-binding sequence-specific

interactions on top of a background of more numerous but less energetic nonspecific binding interactions [26]. This experiment is at present the only experimental realization of nonspecifically binding, DNA-bending proteins studied micromechanically with binding near equilibrium, providing a concrete example for application of the theory.

Using the same structural parameters, the force-distance of the "sliding" model of this section is shown in Figure 5, for few different binding free energies ϵ (equivalently, a few different bulk protein concentrations, see Sect. 2). From left to right, the solid curves show theoretical z/L for $\epsilon = -\infty$ (*i.e.* bare DNA), 0, 2, 4 and 9. The dashed curves show the protein binding as a fraction of the maximum possible, for the $\epsilon = 0, 2, 4$ and 9 (higher occupations correspond to higher values of ϵ). For a net binding free energy of $\epsilon = 0$, the protein partially binds the DNA at low force, reducing the extension relative to bare dsDNA. The curves for larger ϵ correspond to thermal equilibrium of a protein with a higher binding free-energy parameter ϵ. As ϵ is increased, the force required to remove the protein and fully extend the dsDNA increases. The results of this model agree closely with the corresponding approximate calculations of reference [11].

To apply the theory to the IHF data (symbols, Fig. 5), we note that $\epsilon = 0$ fits the experimental data for 1.25 μM protein concentration, indicating a dissociation constant $K_{\rm d} = 1.25\,\mu{\rm M} \times e^{-\epsilon} = 460$ nM. Assuming ideal solution behavior, $\epsilon = \ln(c/K_{\rm d})$, the solid curves correspond to IHF concentrations of 0, 1.25 μM, 9.2 μM, 68 μM and 10 mM. Given the large concentrations, these later predictions for IHF are likely of limited value, and are mainly a qualitative guide.

However, this kind of model for the concentration dependence of the force curves should apply well to DNA-folding proteins which bind to dsDNA nonspecifically (to all sequences with nearly equal binding enthalpy), and with a smaller value of $K_{\rm d}$ in the 10 nM range. Examples of such proteins are HMGB1, a DNA-bending protein which helps to compact the DNA in eukaryote chromosomes, and HU, a protein with similar function in bacteria. We note that for strong binding ($\epsilon \geq 5$) it is likely that equilibration of binding will be very slow. Achieving equilibrium may not be straightforward; it may instead be possible to observe the kinetics of approach to equilibrium, *i.e.* "coarsening" of "domains" of protein-bound DNA. Also, at high occupations, interactions between adjacent proteins will most likely occur, giving rise to cooperativity effects. These effects have been to some extent observed in the essentially irreversibly binding RecA-dsDNA system where essentially the kinetics of a low-temperature one-dimensional system going through a first-order phase transition was observed [27] (in that system, note that ds-DNA is extended longer than B-form as the proteins bind to it).

3.5 Effects of twisting on dsDNA-protein composites

The recent work of Ali *et al.* on equilibrated IHF-dsDNA binding [14] plus the fact that many proteins which bind to dsDNA in some way constrain or alter DNA twisting, suggests the study of extensibility of twisted dsDNAs in the presence of DNA-binding proteins. Our theory can be extended to the case where DNA linking number is fixed, since the problem of determining the generalization of (7) to the ensemble of fixed DNA linking number is in large part solved in the regime where the stress on the double helix is small enough that the double helix is not denatured [28–31]. We will assume that the double-helix twist is small enough that effects of the chirality of DNA winding are not very strong; effects associated with unwinding [32–37] or other DNA structural transitions [5,38–40] are not included here.

The linking number Lk of the two DNA strands is the control parameter of DNA twisting. In a relaxed ds-DNA, the two strands wrap around one another once every 10.5 bp or 3.5 nm, or with an angular "rate" $\omega_0 = 2\pi/(3.5\ {\rm nm})$. The linking number of a relaxed molecule is therefore ${\rm Lk}_0 = 2\pi\omega_0 L$. It is conventional in molecular biology to use the change in linking, ${\rm Lk} - {\rm Lk}_0$, and the fractional change $\sigma = {\rm Lk}/{\rm Lk}_0 - 1 = [{\rm Lk} - {\rm Lk}_0]/(2\pi\omega_0 L)$. The linking number relative to relaxed DNA, ${\rm Lk} - {\rm Lk}_0 = \Delta{\rm Lk}$, has been used as a control parameter for all DNA twisting experiments done to date, since it corresponds to the number of rotations of a bead at one end of the molecule when the other end of the molecule is fixed.

Moroz and Nelson [30] have shown that for torques less than about $k_{\rm B}T$ and for tensions between about 0.2 pN and 2 pN, dsDNA is described by the partition function for a semiflexible polymer with twist stiffness. For a chain of total contour length L under tension f and with linkage density σ, this is $Z = e^{L\tilde{g}(x,y)/A}$, where

$$\tilde{g}(x,y) = x - \sqrt{x} + \frac{1}{8}\frac{y^2}{\sqrt{x}} - \frac{A}{2C}y^2 + \cdots \qquad (26)$$

using an expansion for large force (x) and small twisting (y). Here $x = \beta A f$ is the rescaled force as before, $y = C\omega_0\sigma$ is a dimensionless measure of the linking-number change, and C is the twist stiffness in $k_{\rm B}T$ units with dimensions of length. Since linear twisting elasticity is assumed in (26), y is the mechanical torque in the dsDNA molecule in $k_{\rm B}T$ units.

To apply (26) to DNA, some limitations must be recognized. Equation (26) ignores any asymmetry between overwinding and underwinding, being an even function of the reduced torque y. At first, the asymmetric-y effects are small and take the form of a twist-stretch coupling which is easily added to (26) [41,42]. This type of correction does not strongly affect results presented here. However, the torque response of dsDNA becomes highly asymmetric once the DNA begins to unwind. The threshold unwinding torque for this is roughly $-2k_{\rm B}T$ [34,35]. For overwinding, denaturation of DNA also occurs, but at a larger torque $\approx 6k_{\rm B}T$ [38].

Beyond these torque limits, severe alterations of DNA secondary structure occur [5,40,36,37], and the theory of this section cannot apply. Among other complications, there is the question of the affinity of the protein in question for single-stranded *versus* double-stranded DNA.

Thus we only show results for which the DNA torque is within $-2k_BT$ and $6k_BT$. We also do not study forces above ≈ 1 pN, where stretch-driven modification of DNA structure can occur [5,39,40,36,37]. Between these limits and at forces between about 0.1 and 2 pN, the harmonic twist-elastic model of (26) is a reasonable description of dsDNA, using C between 75 and 100 nm. Finally, we note that (26) does not include folding of DNA into plectonemic supercoils, which occurs at low forces for sufficient twisting [32].

For proteins which bind to, fold, but also untwist the double helix, the parameters K, ℓ and Δ are still relevant. However, if Lk of the DNA is fixed, we now must also specify the linking-number change induced in the DNA by each bound protein, $\Theta/(2\pi)$. For example, a protein which twists the double helix when it binds would have a Θ equal to the excess twist angle; $\Theta < 0$ would correspond to unwinding. In general, protein-binding might also introduce chiral bending (writhing) which will also contibute to Θ.

When k proteins bind, the linking number of the protein-free region of the DNA will therefore be Lk $- k\Theta/(2\pi)$. Combining this with the free DNA length $L - k\ell$ indicates that in (26) y should be replaced with

$$\tilde{y} = \frac{C\omega_0\left(\sigma - \frac{\Theta}{\omega_0\Delta}\psi\right)}{1 - \frac{\ell}{\Delta}\psi}, \tag{27}$$

where, as before, $\psi \equiv k/K$. The resulting partition function for the twisted and stretched molecule in the presence of proteins is therefore

$$Z_{\mathrm{fold,tw}} = \sum_{k=0}^{K} W_k \exp\left[(L - k\ell)\tilde{g}(x,\tilde{y})/A + k\epsilon\right]. \tag{28}$$

For large K, this sum can as above be converted to an integral over $\psi \equiv k/K$ which can be evaluated using the stationary point of the exponent, as done above. The expectation value $\langle\psi\rangle = \psi_0$ is given by the stationary point equation

$$0 = \frac{\partial}{\partial_\psi}\left[\ln W_k + (L - K\psi\ell)\tilde{g}(x,\tilde{y})/A + K\psi\epsilon\right]_{\psi=\psi_0}. \tag{29}$$

The extensive part of the partition function follows as

$$\ln Z_{\mathrm{fold,tw}} = \ln W_k + (L - K\psi_0\ell)\tilde{g}(x,\tilde{y})/A + K\psi_0\epsilon. \tag{30}$$

The end-to-end extension can be computed again using (16), and has a form similar to (17):

$$\begin{aligned}\frac{\langle z\rangle}{L} &= \frac{k_BT}{L}\frac{\partial \ln Z_{\mathrm{fold,tw}}}{\partial f} \\ &= \left[1 - \frac{\ell}{\Delta}\psi_0\right]\frac{\partial\tilde{g}(x,\tilde{y})}{\partial x} \\ &= \left[1 - \frac{\ell}{\Delta}\psi_0\right]\left[1 - \frac{1}{2\sqrt{x}} - \frac{\tilde{y}^2}{16x^{3/2}} + \cdots\right].\end{aligned} \tag{31}$$

The only difference between these results in the "fixed" and "sliding" cases discussed above is the form of W_k, and in turn the precise form of (29) which determines ψ_0. Plugging in either (13) or (14) converts (29) to the forms

$$\begin{aligned}0 = \frac{\partial}{\partial\psi}\Bigg[&\psi\ln\psi + (1-\psi)\ln(1-\psi) \\ &+ \frac{\ell\psi - \Delta}{A}\tilde{g}(x,\tilde{y}) - \psi\epsilon\Bigg]_{\psi=\psi_0}\end{aligned} \tag{32}$$

for the fixed case, and

$$\begin{aligned}0 = \frac{\partial}{\partial\psi}\Bigg[&\psi\ln\psi - \psi\ln(1-\psi) \\ &+ \frac{\ell\psi - \Delta}{A}\tilde{g}(x,\tilde{y}) - (1+\epsilon)\psi\Bigg]_{\psi=\psi_0}\end{aligned} \tag{33}$$

in the sliding case.

We now describe the application of this model to an IHF-like protein binding to dsDNA with fixed linking number. The change in DNA linkage induced by nonspecific IHF binding is not well understood, providing a motivation for a micromanipulation experiment. Here, we must simply guess the value of Θ, the change in linkage (angle) per bound protein. IHF is known to severely bend dsDNA, and therefore it should force the double helix to unwind near the sharp bend. Other DNA-bending proteins such as HMGB1 or HU should similarly drive helix unwinding in the bend region. Per bound protein, we will assume that this unwinding is a half-link, *i.e.* $\Theta = \pi$. Then, low-force binding of IHF at an occupation fraction of 0.50 requires a half-link of unwinding of the DNA for every $2\Delta = 150$ nm ≈ 500 bp, or a whole link every 1000 bp.

Figure 6 shows results, assuming (as in the previous section) $\epsilon = 0$, $A = 65$ nm, $\ell = 98$ nm and $\Delta = 39$ nm, using the "sliding" entropy model. The twisting elastic constant is taken to be $C = 100$ nm. Figure 6(a) shows extension *versus* force, at fixed linking numbers $\sigma = -0.01, 0$ and $+0.01$. At low forces, the symmetry between positive and negative linking number that would normally hold for $\sigma = \pm 0.01$ is broken by the presence of IHF, the binding of which is stimulated by undertwisting. This binding drives the $\sigma = -0.01$ force curve to be well below that for $\sigma = +0.01$ at low forces. This asymmetry is generated by the protein binding, since the bare DNA response (26) is symmetric under $\sigma \to -\sigma$.

Figure 6(b) shows extension *versus* linkage number, at fixed forces $f = 0.125$ (bottom curve) and 0.4 pN (top). The binding of the protein again generates asymmetry in σ not present in (26), relative to the bare DNA result (dotted curves). As in Figure 6(a), $\sigma < 0$ drives protein binding, which in turn reduces extension relative to $\sigma > 0$. The IHF occupation fraction corresponding to these force cases is shown in Figure 6(c). Note that the lower occupation curves correspond to higher forces; high forces force DNA extension and therefore protein unbinding. As the DNA is underwound, the protein gradually binds.

The theory presented here has not included a description of the sharp transitions that bare DNA has been observed to undergo when it is strongly twisted and stretched [5,39,40,36,37]. For this reason we have shown results only

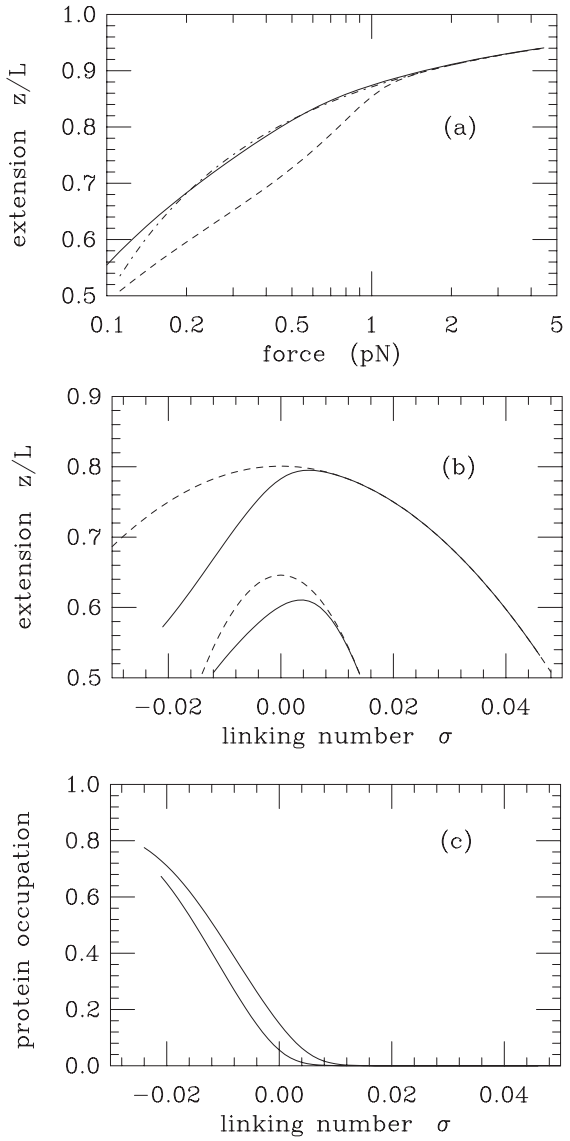

Fig. 6. Effect of linking-number constraint on elastic response of dsDNA, in the presence of an IHF-like protein, under the assumption that each bound IHF introduces an untwisting of the double helix by $\Theta = -\pi$ radians. Results for the "sliding" model with a net binding free energy $\epsilon = 0.0 k_{\rm B}T$ are shown. Results are plotted only for DNA twisting torque in the range $-2k_{\rm B}T$ to $6k_{\rm B}T$, where the double helix is known to be stable. (a) Extension *versus* force, at fixed linking numbers $\sigma = -0.01$ (dashed), 0 (solid) and +0.01 (dot-dashed). At high forces, all curves converge to the "bare" dsDNA elastic response. At low forces, the usual symmetry between positive and negative linking number is broken by the presence of the bound protein, the binding of which is stimulated by undertwisting. Since protein binding removes length from the chain, the $\sigma = -0.01$ curve is appreciably below the others. (b) Extension *versus* linkage number, at fixed forces $f = 0.125$ (bottom) and 0.4 pN (top). For bare DNA extension is an even function of linkage number change (dashed) for these low forces. Addition of protein (solid lines) breaks this symmetry; undertwisting stimulates protein binding, and reduction of extension. (c) IHF occupation fraction corresponding to part (b), $f = 0.125$ (top), 0.4 pN (bottom). Undertwisting drives binding of the protein.

in the regime where DNA is not denatured, and where there is essentially no asymmetry between undertwisting and overtwisting [32,33]. Against this background, the chiral asymmetry induced by the binding of the protein (Fig. 6a,b) provides a clear experimental signal.

For undertwisting beyond $\sigma \approx -0.01$, one can expect bare dsDNA to strand-separate when it is fully extended; in this situation the linking number is forced into twist, and when the unwinding torque reaches about $-2k_{\rm B}T$ dsDNA becomes unstable [40]. As remarked previously, these DNA structural transitions can be treated by models similar to that of this section [24,42]. Structural transitions of the DNA and protein binding could be therefore treated by a generalization of the model of this section to one with multiple states [40]. However, we warn of the complication that the affinity of the protein may vary greatly with DNA structural state. On the other hand, if one stays in the regime of low forces < 1 pN and low supercoiling ($|\sigma| < 0.01$) where the model (26) is appropriate, our theory can provide quantitative information about DNA untwisting by protein.

The calculation of this subsection was done for fixed linkage change σ, which is the control parameter for all DNA twisting experiments to date. It would be extremely interesting to carry out DNA twisting experiments as a function of fixed torque, and the above calculation can be redone in that ensemble. Roughly, when a length and angle constraining protein dissociates (at enthalpic cost $k_{\rm B}T\epsilon$), force-distance work $\approx f\ell$ will be done, and at fixed torque, twisting work $\tau\Theta + \tau^2\ell/(2C)$ is also done. The first term is due to protein-driven unwinding, and the second term is the twisting free energy of the released DNA. The protein-dissociated state therefore has lower free energy when

$$f > \frac{k_{\rm B}T\epsilon - \tau\Theta}{\ell} - \frac{\tau^2}{2k_{\rm B}TC}. \qquad (34)$$

Here we have used the harmonic DNA twist energy [11, 28] which applies only when the molecule is not stress-denatured. Although this condition is only semiquantitative, it indicates how unwinding generated by protein binding could be measured using the shift in the coexistence force of folded and unfolded states. In the nondenatured regime the linear torque-twist relation $\tau = C\omega_0\sigma$ [11,28] can be used to express (34) in terms of linking number. This brings together results for force-removal [11] and torque-removal [43] of DNA-binding proteins, including the effect of protein-induced unwinding.

3.6 ssDNA —folding and electrostatic effects

If we include screened electrostatic interactions, the model of Section 3.1 can be used to understand the salt-dependent elastic response of single-stranded DNA (ssDNA). At relatively high univalent ionic strengths $>$ 100 mM, ssDNA collapses at low force, requiring about 2 pN to start to extend [7,17]. As discussed in Section 2.2, this relatively low-force threshold is most likely due to opening of short "hairpin" structures [9,10].

At low univalent ionic strength ≈ 1 mM, ssDNA has a peculiar extension apparently varying $\approx \ln f$ (Fig. 2, experimental data), and the 2 pN force threshold seen at low ionic strength is absent [17]. Recent computer simulation work has indicated that these effects can be taken into account by a combination of hairpin formation (*i.e.* chain folding) and electrostatic self-repulsion. The basic idea is that at high ionic strength, screening of electrostatic self-repulsion allows the chain to fold up at low force.

The electrostatic effects can be taken into account using Debye-Huckel interactions, and at the harmonic level of description for a worm-like chain, following Barrat and Joanny [44]. These interactions have been discussed in connection with force-distance experiments on dsDNA [23]. Barrat and Joanny's calculation gives rise to a scale-dependent persistence length,

$$A(q) = A_0 + \ell_{\rm OSF} K(q),$$
$$K(q) = \frac{2}{\lambda_{\rm D}^4 q^4}\left([1+\lambda_{\rm D}^2 q^2]\ln[1+\lambda_{\rm D}^2 q^2] - \lambda_{\rm D}^2 q^2\right), \quad (35)$$

where q is the wave number of bending modes, $\lambda_{\rm D}$ is the Debye screening length, $\ell_{\rm OSF} = \ell_{\rm B}(\lambda_{\rm D}\nu)^2/4$ is the Odijk-Skolnick-Fixman length, for chain charge density ν in electron charges per contour length, and $\ell_{\rm B}$ is the Bjerrum length, about 0.7 nm in water at room temperature. The basic behavior of $K(q)$ is that it goes to 1 for $q \to 0$, and decays to zero as $(\ln q)/q^2$ for large q. Thus, the effective persistence length increases from its intrinsic elastic value A_0 at small wavelength, to an electrostatically enhanced value $A_0 + \ell_{\rm OSF}$ at large wavelengths.

The main point is that $\ell_{\rm OSF}$ can be quite large (> 10 nm) at low ionic strength; we recall that $\lambda_{\rm D} \approx 0.3$ nm$/c^{1/2}$, where c is the univalent ion concentration in Mol/ℓ. For ssDNA, $\nu \approx 1.6$ nm^{-1} (this is near the structural charge, and the Manning condensation limit). Thus for 1 mM NaCl concentration, we have $\ell_{\rm OSF} \approx 40$ nm. This is a large enhancement over the intrinsic ssDNA backbone persistence length $A_0 \approx 1$ nm, and thus as force (and hence the characteristic q for bending modes) is increased, $A(q)$ will decrease. This will make the polymer progressively more difficult to extend as it is stretched out, which is qualitatively the effect observed in experiments at low ionic strength. This effect is far more striking for ssDNA than for dsDNA [23], due to the two polymers having about the same charge density, but very different intrinsic persistence lengths.

All we need to apply the formalism of the previous sections is to compute the free energy per length of the unfolded chain as a function of force f. For this problem, analytical progress is practical only for the harmonic approximation to the bending modes, where

$$g(\beta f A_0) = \beta f A_0 + \frac{\beta A_0 f^2}{2 f_0} - A_0 \int \frac{{\rm d}q}{2\pi} \ln\left(\frac{[A_0 + \ell_{\rm OSF} K(q)]\, q^2 + \beta f}{A_0 q^2 + 1/A_0}\right). \quad (36)$$

The denominator of the logarithm is chosen to set the free energy to be roughly near zero for forces on the order of $k_{\rm B}T/A_0 \approx 4$ pN. Note that the intrinsic persistence length $A_0 \approx 1$ nm is used as the reference length scale. The final term is a harmonic stretch modulus for the chain, where f_0 is the characteristic force (elastic) constant. The chain extension, in units of the unperturbed contour length, is obtained by differentiation with respect to $\beta A_0 f$,

$$\frac{z}{L} = 1 - \int \frac{{\rm d}q}{2\pi} \frac{1}{[A_0 + \ell_{\rm OSF} K(q)]\, q^2 + \beta f} + \frac{f}{f_0}, \quad (37)$$

which is equation (19) of reference [23]. This model will be accurate as long as the bending fluctuations are reasonably small, *i.e.* as long as the extension $z/L > 0.5$.

Although there are a number of parameters, there is little choice as to what they are. The charge density along the chain will be the Manning limit $\nu = 1/\ell_{\rm B}$ since the spacing of structural charges exceeds this value (there is one PO_4^- every 0.7 nm along the fully extended polynucleotide backbone). The intrinsic segment length has been measured in the high-force regime to be roughly $b = 2A = 1.5$ nm, and so we adopt $A = 0.75$ nm. The Debye length for sodium chloride solution is $\lambda_{\rm D} = 0.3 M_{\rm NaCl}^{-1/2}$ nm. Finally, the elastic constant f_0 should be large, since the relevant elasticity is that of the chemically bonded backbone. We take $f_0 = 2000$ pN, which means that elastic backbone stretching will play a minor role at the forces < 200 pN relevant to the single-molecule experiments.

The folding interactions require specification of the parameters ℓ and Δ; we take the values $\ell = \Delta = 3$ nm to describe the folding up of a few bases of ssDNA into compact structures. We also use the "fixed" folding model since we are describing sequence-directed folding of the ssDNA. The remaining parameter is the folding energy, which we take to be of the form $\epsilon = \epsilon_0 - \epsilon_1 \exp[-R_0/\lambda_{\rm D}]$. This form accounts for the intrinsic self-attraction of the folded structure, competing with its Coulomb repulsion. We take $\epsilon_0 = 3$, $\epsilon_1 = 30$, and $R_0 = 2$ nm. At relatively low ionic strength ($\lambda_{\rm D} > 1$ nm) Coulomb replusion destabilizes the folded structures. The calculation then just requires numerical evaluation of the q-integrals, and the use of g and z/L in equation (16).

Results for a few NaCl concentrations are shown in Figure 7. The solid curves indicate the region where the unfolded portion of the chain has extension greater than 0.4, and therefore is where the Gaussian calculation (37) is most accurate. At high ionic strength (1.5 M) the chain folds up, and requires a force ≈ 1 pN to be extended, as in experiment. For low ionic strength (1.5 mM and 0.5 mM curves), the Coulomb interactions eliminate the folded states, and also give rise to the scale-dependent elastic response discussed in reference [23].

Figure 7 includes experimental data for 0.15 M and 2.5 mM Na^+ from reference [17] The data for 0.15 M show the "collapse" behavior and the ≈ 1 pN zero-extension force associated with self-adhesion of the single-stranded molecule. However, by the time the salt concentration is reduced to 2.5 mM, the self-adhesion is overwhelmed by the Coulomb repulsion, and a very slow variation of extension with force ($\approx \ln f$) is observed.

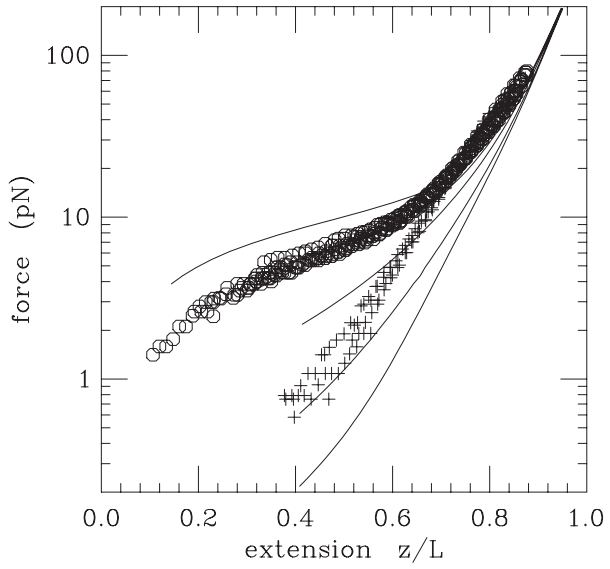

Fig. 7. Force *versus* extension for ssDNA at various ionic strengths. Curves from top to bottom show theory of semiflexible chain with folding and electrostatic interactions, for 1.5 M, 150 mM, 15 mM, 1.5 mM and 0.5 mM NaCl concentrations. At high salt, the folding of the chain dominates the low-force elasticity; at low salt, Coulomb repulsion gives rise to a nearly exponential dependence of force on extension. Extension is relative to the maximum possible extension of the ssDNA, corresponding to about 7 Å per base pair. Relative to this, B-DNA (3.4 Å per bp) has an extension of just slightly less than 0.5. Data for 2.5 mM (+) and 150 mM Na^+ (∘) from reference [17] are shown for comparison. The experimental data are the same as those of the inset of Figure 3, but with the extension coordinate scaled by 0.5.

In this theory, the peculiar force response seen at low ionic strength is due to the Barrat-Joanny scale-dependent elasticity rather than to the formation of secondary structure. This is reasonable since base-pairing is destabilized at 1 mM ionic strength, while there will certainly be strong perturbation of chain elasticity by electrostatic interactions. In addition, these results are in good accord with the conclusions of Zhang *et al.* [8], who have carried out Monte Carlo simulation of a semiflexible chain with folding and Debye-Huckel interactions. Very recently, Dessinges *et al.* [45] came to similar conclusions, again using a comparison of Monte Carlo simulation of a polymer with Debye-Huckel and folding interactions, with single-molecule experiment data.

The weak (log-like) dependence of extension on applied force can be roughly understood by examination of the integrand of (37), which has approximately the form $[f + A_0q^2 + \ell_{\rm OSF}q^2/(1 + \lambda_{\rm D}q^2/4)]^{-1}$. The peak value of the integrand is always $\approx k_{\rm B}T/f$ (the $q \to 0$ limit). The integrand goes to zero for large wave number $\sim 1/q^2$ (neglecting the $\log q$ of the large-q limit of $K(q)$), and the integral can be estimated from the width of the integrand. For high forces $> k_{\rm B}T/A_0$, the width approaches the larger "bare" value $\approx \sqrt{\beta f/A_0}$, causing (33) to tend to the usual WLC form $\approx (\beta A_0 f)^{-1/2}$. At low forces $< k_{\rm B}T/\ell_{\rm OSF}$, the width is reduced by the electrostatic stiffening to $\approx \sqrt{\beta f/(A_0 + \ell_{\rm OSF})}$, which in turn makes (37) $\approx (\beta[A_0 + \ell_{\rm OSF}]f)^{-1/2}$. The result is that for low ionic strength, the value of integral (37) changes only by about a factor of 5 from 1 to 100 pN.

4 Conclusion

In this paper we have studied the effect of "folding" interactions on polymer elastic response. These interactions are presumed to cause nearby monomers along a chain to stick to one another, but without the overall chain self-adhesion that causes collapse of a polymer into a compact "globule". Thus, instead of applying to overall hydrophobicity, the results of this paper apply to polymers which have localized interactions, as occurs for many biopolymers.

We showed that for simple models of a folding polymer, the self-interactions lead to a reduction in the apparent persistence length for low forces ($\leq k_{\rm B}T$ per persistence length). This might lead to a disagreement between persistence lengths measured from coil sizes or low-force polymer elastic response, and measured from high-extension force response. For polymers with long persistence lengths, as is typical for biopolymers, the "folding" interactions do not have to be strong to lead to large effects.

We have applied our model to single-stranded DNA (ssDNA), measurements of the persistence length of which vary over a wide range [46]. The peculiar low-force elasticity observed in single-molecule experiments can be largely explained using a simple folding model describing the folding of the chain into numerous small "hairpin" structures at low force [10]. While a detailed sequence-based model [9,10] may be necessary to understand the lowest-force response, a more generic self-interaction model captures the basic features of experiment. By combining a folding model with electrostatic self-interaction, we are able to describe ssDNA elastic response over a wide range of forces and ionic strengths, providing an analytic theory for effects previously studied using computer simulation [8].

The same model has been used to describe unfolding elastic response of chromatin fiber, which occurs on the few-pN scale. This should not be confused with nucleosome removal which has been observed near to 20 pN by a few groups [3,18]. Currently, only Cui and Bustamante [3] have explored the entropic elasticity of chromatin fiber. It would be of great interest to see further sub-piconewton experiments that would provide better determination of chromatin fiber persistence length, and indeed, stronger evidence that a flexible-polymer elasticity picture is appropriate for chromatin at low forces.

We have also developed a model where the folding is on scales large compared to the polymer persistence length. This is applicable to loops or other folds of dsDNA that might be formed by proteins which interact with the double helix. Such loop-forming and DNA-bending proteins have already been studied using single-molecule micromechanical techniques [15,14]. Examples of other loop-forming proteins that would be good candidates for this

type of experimental study include site-specific recombination enzymes [47] and certain restriction enzymes which bind two target sites [48], under conditions where loops form, but where enzyme catalytic activity is impaired.

In the model, length is effectively lost from the dsDNA in a manner similar to a model for DNA structural changes introduced by Ahsan *et al.* [24]. Essentially, two phenomena occur in this situation. For low extensions, force enhancement occurs due to the effective persistence length being forced down by the folding interactions. However, at large extensions, there can be an unfolding "transition" where the folding proteins are dissociated. This transition can be observed at moderate forces below those that will change double-helix secondary structure. To some extent these effects have already been observed for the protein IHF in single-molecule experiments of Ali *et al.* [14]. We have also predicted force curves for the situation where DNA linking number is constrained; this is likely relevant for DNA-bending proteins which are likely to untwist the double helix.

There are a number of extensions to this model of DNA folding which are practical. First, cooperativity effects could be included in our "length loss" model, in order to model the interaction of nearby DNA-fold structures. Second, DNA-folding proteins may alter DNA flexibility, either stiffening or greatly weakening the double helix; this kind of effect could also be added. Finally, for the case of DNA bending, it would be desirable to include a better geometrical model for the actual double-helix bending. All these effects might be incorporated using discrete transfer matrix techniques.

We thank J. Stavans, R. Amit, C. Bustamante and S. Smith for providing us their experimental data, and we thank V. Croquette for helpful discussions. This research was supported in part by the NSF through grants DMR-9734178 and DMR-0203963, by the Whitaker Foundation through a BMERG grant, by the Johnson and Johnson Focused Giving Grant Program, the American Chemical Society Petroleum Research Foundation, and the Research Corporation. S.C. was partly supported by the Della Riccia Foundation, and R.M. was partly supported by the NSF through grant DMR-9808595.

References

1. P.G. de Gennes, *Scaling Theory of Polymer Physics* (Cornell University Press, 1985) Chapt. 1.
2. P. Hagerman, Ann. Rev. Biophys. Biophys. Chem. **17**, 265 (1988).
3. Y. Cui, S. Smith, C. Bustamante, Proc. Natl. Acad. Sci. USA **97**, 127 (2000).
4. F. Gittes, B. Mickey, J. Nettleton, J. Howard, J. Cell Biol. **120**, 923 (1993).
5. S.B. Smith, Y. Cui, C. Bustamante, Science **271**, 795 (1996).
6. M. Rief, H. Clausen-Schaumann, H.E. Gaub, Nature Struct. Biol. **6**, 346 (1999).
7. B. Maier, D. Bensimon, V. Croquette, Proc. Natl. Acad. Sci. USA **97**, 12002 (2000).
8. Y. Zhang, H.J. Zhou, Z.C. Ou-Yang, Biophys. J. **81**, 1133 (2001).
9. U. Gerland, R. Bundschuh, T. Hwa, Biophys. J. **81**, 1324 (2001).
10. A. Montanari, M. Mezard, Phys. Rev. Lett. **86**, 2178 (2001).
11. J.F. Marko, E.D. Siggia, Biophys. J. **73**, 2173 (1997).
12. J. Rudnick, R. Bruinsma, Biophys. J. **76**, 1725 (1999).
13. H. Diamant, D. Andelman, Phys. Rev. E (2000).
14. B.M.J. Ali, R. Amit, I. Braslavsky, A.B. Oppenheim, O. Gileadi, J. Stavans, Proc. Natl. Acad. Sci. USA **98**, 10658 (2001).
15. L. Finzi, J. Gelles, Science **267**, 378 (1995).
16. M.V. Volḱenshtein, *Configurational Statistics of Polymeric Chains* (Interscience, New York, 1963).
17. C. Bustamante, S.B. Smith, J. Liphardt, D. Smith, Curr. Opin. Struct. Biol. **10**, 279 (2000).
18. M.L. Bennink, S.H. Leuba, G.H. Leno, J. Zlatanova, B.G. de Grooth, J. Greve, Nature Struct. Biol. **8**, 606 (2001).
19. M. Poirier, S. Eroglu, D. Chatenay, J.F. Marko, Mol. Biol. Cell **11**, 269 (2000).
20. C. Bouchiat, condmat/0201445 (2002).
21. U. Bockelmann, P. Thomen, B. Essevaz-Roulet, V. Viasnoff, F. Heslot, Biophys. J. **82**, 1537 (2002).
22. B. Houchmandzadeh, J.F. Marko, D. Chatenay, A. Libchaber, J. Cell Biol. **139**, 1 (1997).
23. J.F. Marko, E.D. Siggia, Macromolecules **28**, 8759 (1995).
24. A. Ahsan, J. Rudnick, R. Bruinsma, Biophys. J. **74**, 132 (1998).
25. L. Tonks, Phys. Rev. **50**, 955 (1936). Note that the one-dimensional partition function of this paper is obtained for k distinguishable cores with definite ordering along the axis of motion: the result for indistinguishable particles with arbitrary ordering is obtained by multiplying the result by $k!$ to take into account the arbitrary ordering, and then dividing it by $k!$ to apply the indistinguishability.
26. J. Stavans, private communication (2002).
27. J.-F. Leger *et al.*, Proc. Natl. Acad. Sci. USA **95**, 12295 (1998).
28. J.F. Marko, E.D. Siggia, Phys. Rev. E **52**, 2912 (1995).
29. A. Vologodskii, J.F. Marko, Biophys. J. **73**, 123 (1997).
30. D. Moroz, P. Nelson, Proc. Natl. Acad. Sci. USA **94**, 14418 (1997); Macromolecules **31**, 6333 (1998).
31. C. Bouchiat, M. Mezard, Phys. Rev. Lett. **80**, 1556 (1998); Eur. Phys. J. E **2**, 377 (2000).
32. T. Strick, V. Croquette, D. Bensimon, Biophys. J. **74**, 2016 (1998).
33. T. Strick, J.-F. Allemand, V. Croquette, D. Bensimon, Prog. Biophys. Mol. Biol. **74**, 115 (2000).
34. T.R. Strick, D. Bensimon, V. Croquette, Genetica **106**, 57 (1999).
35. S. Cocco, R. Monasson, Phys. Rev. Lett. **83**, 5178 (1999).
36. H.J. Zhou, Y. Zhang, Z.C. Ou-Yang, Phys. Rev. Lett. **82**, 4560 (1999); Phys. Rev. E **62**, 1045 (2000).
37. S. Panyukov, Y. Rabin, Europhys. Lett. **57**, 512 (2002).
38. T.R. Strick, D. Bensimon, V. Croquette, in *Structural Biology and Functional Genomics*, edited by E.M. Bradbury, S. Pongor, *NATO Sci. Ser.*, Vol. **3** (Kluwer Academic Publishers, Boston, 1999).
39. P. Cluzel *et al.*, Science **271**, 791 (1996).
40. J.-F. Leger *et al.*, Phys. Rev. Lett. **83**, 1066 (1999); A. Sarkar *et al.*, Phys. Rev. E **63**, 051903 (2001).
41. R. Kamien *et al.*, Europhys. Lett. **38**, 237 (1997).

42. J.F. Marko, *Phys. Rev. E* **57**, 2134 (1998).
43. A. Sarkar, J.F. Marko, Phys. Rev. E **64**, 061909 (2001).
44. J.-L. Barrat, J.-F. Joanny, Europhys. Lett. **24**, 333 (1993).
45. M.-N. Dessinges, B. Maier, Y. Zhang, M. Peliti, D. Bensimon, V. Croquette, Phys. Rev. Lett. **89**, 248102 (2002).
46. S.V. Kuznetsov, Y. Shen, A.S. Benight, A. Ansari, Biophys. J. **81** 2864 (2001).
47. I.Y. Goryshin, W.S. Reznikoff, J. Biol. Chem. **273**, 7367 (1998).
48. M. Mücke, R. Lurz, P. Mackeldanz, J. Behlke, D.H. Krüger, M. Reuter, J. Biol. Chem. **275**, 30631 (2000).