Effects of DNA-distorting proteins on DNA elastic response

Jie Yan and John F. Marko
Department of Physics, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607-7059, USA
(Received 15 January 2003; published 15 July 2003; publisher error corrected 5 August 2003)

Many DNA-binding proteins distort the double helix, and therefore can be studied using single-molecule experiments to investigate how they modify double-helix polymer elasticity. We study this problem theoretically using discrete wormlike-chain models to describe the mechanics of protein-DNA composites. We consider the cases of nonspecific and specific (sequence-targeted) binding. We find that, in general, proteins which bend DNA can be described in terms of a reduction of bending persistence length as long as the binding strength is relatively weak (well below the dissociation point). For strong binding, the force response depends strongly on the bending stiffness of the DNA-protein complex. Since most DNA-bending proteins will cause local DNA untwisting, we also show how the constraint of DNA linking number modifies the observed elastic response. We also show how essentially the same model may be used to describe the binding of proteins and drugs which stiffen and stretch the double helix.

DOI: 10.1103/PhysRevE.68.011905 PACS number(s): 87.14.Gg, 82.35.Lr, 82.37.Rs, 87.15.He

I. INTRODUCTION

The DNA double helix is distorted during its processing inside cells. For example, the millimeter-long DNAs inside bacterial cells and the centimeter-long ones inside eukaryote cells must be folded up just to fit inside the roughly micron-size compartments that contain them. Folding and looping of DNA also plays a key role in the reading of DNA sequence, such as in the regulation of transcription by distant sequences [1–3], transposition of DNA segments [4], and activation of certain restriction enzymes [5].

To facilitate packaging and looping processes, there are many proteins inside cells whose function is to distort the double helix. In eubacteria (e.g., *E. coli*) there are a number of ''architectural'' DNA-bending proteins, including IHF, HU, and H-NS [6]. IHF and HU can introduce severe bends into the double helix [7–9], while H-NS is able to ''coat'' DNA, changing its polymer properties in the process [10]. A well-studied instance where one of these types of proteins mediates DNA loop formation is IHF involvement in λ-phage excision [7].

In eukaryote cells, there exist similar proteins, such as the DNA-bending protein HMGB1. This protein can generate severe DNA bends, and is present at a concentration of one per every few nucleosomes [11]. HMGB1 is considered to play a role in the assembly of nucleoprotein complexes during recombination, transcription, and DNA repair processes [12]. Nucleosomes, the basic genome packaging units of eukaryotes, contain eight complexed ''core histone'' proteins around which DNA wraps in 1.75 tight turns of bending radius ≈ 5 nm [13]. The structural studies mentioned above (see also Ref. [14] for a review of DNA-bending proteins) show that the double helix can be severely distorted by DNA-protein interactions.

A feature of the DNA-distorting architectural proteins mentioned above is that they are able to bind to DNA in a non-sequence-specific manner, i.e., to any DNA sequence. IHF can bind tightly to certain sequences, but it also binds quite strongly to other DNA sequences [15], while HU, H-NS, HMGB1, and histones are classified as non-sequence-specific proteins [11]. The packaging and looping functions of these proteins probably require their action at arbitrary locations along DNA *in vivo*.

The biochemical study of DNA-protein interactions has been focused on characterizing sequence-specific interactions, since those can be studied using short, defined DNAs. A standard experiment comprises measuring the shift of the distance moved by a DNA-protein complex on an electrophoretic gel, relative to DNA alone. If a protein and its DNA target form homogeneous enough complexes, they sometimes can be crystallized together, leading via x-ray diffraction to a complete picture of the DNA distortions introduced by the DNA-protein interactions [7]. These techniques are difficult to apply to study non-sequence-specific DNA-protein interactions.

An alternate approach to studying non-sequence-specific interactions is to study their effects on large DNA molecules. Single-molecule manipulation provides a way to study this via the force-extension response of DNA in the presence of a given protein [15,10]. The aim of this paper is to analyze statistical-mechanical models for these kinds of experiments, taking into account changes in local double-helix structure driven by binding of proteins. Our previous work on this problem [16] considered a model of ''lost length'' resulting from proteins binding to DNA, which occur when a DNA loop is formed [17]; here we introduce models for local changes in DNA bending, as might be generated by the proteins mentioned above. The phenomena of this paper are also related to work of Rudnick and Bruinsma [18] on tension modification of DNA-bending protein binding, and of Diamant and Andelman [19] on proteins which cause coupling of bending modes.

In Sec. II, we discuss the basic method of calculation used, namely, the transfer matrix applied to the orientational degrees of freedom of a DNA. We use a discrete transfer matrix approach based on considering a long DNA to be a semiflexible chain of finite-length segments of equilibrium length b; in most of the calculations in this paper we will use $b \approx 5$ nm. This segment length is short enough to provide an

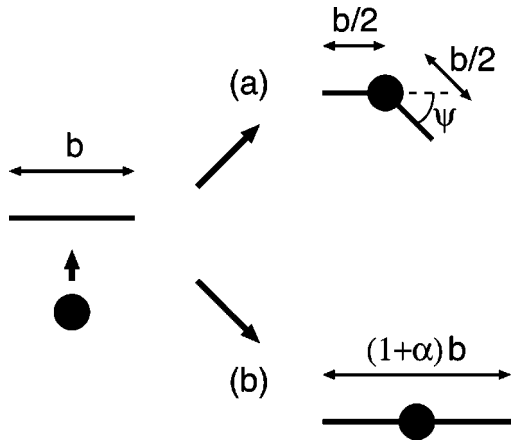

FIG. 1. Geometry of the model for DNA-binding proteins. (a) DNA-bending protein. A protein binds to a DNA segment and severely distorts it, effectively forcing a permanent bend by an angle ψ. Thermal bending fluctuations then occur centered around this permanent bend angle. In this paper, the double helix is divided up into segments of equilibrium length b. (b) DNA-lengthening protein. Many proteins and drugs which bind to DNA insert a ligand between the bases of the double helix, forcing it to lengthen. Here, such binding is considered to increase the length of a DNA segment by an amount αb.

accurate representation of the semiflexible polymer behavior of the double helix.

The discrete-segment model and calculations used here are similar to those used in calculations of force-induced conformational change (''overstretching'') by Storm and Nelson [20]. In this work, we use the segment length to set the size of protein-binding sites. A discrete model is used to describe local changes in DNA properties such as spontaneous bending, or bending rigidity, that would be difficult to analyze using continuum models.

In Sec. II A, we discuss proteins which are in chemical equilibrium between DNA-bound and free-solution states, introducing a source of severe bending disorder along the double helix. Figure 1(a) shows the DNA-bending geometry considered in this paper. We will assume a protein which binds to locally force a bend by an angle ψ; our calculations will often use the parameter $\cos\psi \equiv \gamma$. The model discussed in Sec. II A will be the most applicable to proteins such as HU and HMGB1.

In Sec. IIB, the action of a locally DNA-stiffening protein will be considered. Instead of introducing a bend, a segment will have its bending stiffness increased by the model protein, corresponding to the action of H-NS [10]. Then, in Sec. II C we will discuss the effect of proteins or drugs which lengthen the double helix, for example, via insertion of a ligand between bases. Such binding is considered to increase the length of a DNA segment by an amount αb, as shown in Fig. 1(b).

In Sec. III, we consider the effects of bends at fixed locations along long DNAs. This is relevant to DNA-bending proteins which bind so strongly to specific sequences that they are essentially always bound to their targets, once they are present, even at low concentrations. Examples of such proteins, which would be interesting to study using micromanipulation techniques, include transcription factors [21,22] and restriction enzymes [14]. In Sec. III A, we consider long DNAs with such proteins bound with fixed spacing, and examine the question of how often such proteins must be bound in order for their presence to be detectable in a force-extension curve. Section III B analyzes the case where the binding locations are random but with a given frequency, again focusing on the question of how often proteins must bind in order for their effect to be evident in the polymer elasticity of a long DNA.

Section IV considers the effect of DNA-bending proteins on force-extension response, in the case where the DNA linking number is fixed or equivalently, where the double helix is subject to torsional stress. In some cases, a severe protein-induced DNA bend is known to be accompanied by local unwinding [21,23]. We show how this unwinding can be studied using number force-extension-twist experiments.

II. DISCRETE MODEL OF DNA-BENDING PROTEINS

We consider a discrete realization of a semiflexible polymer, made up of a series of segments. The contour length of each segment is b and the orientation of each segment is described by a unit vector $\hat{\mathbf{t}}_i$. Each of the nodes between the segments also carries a protein occupation degree of freedom n_i, which is either 0 for bare DNA or 1 when a protein is bound. The energy that we consider is expressed as a sum over the segments,

$$E = \sum_{i=1}^{N} E_i . \tag{1}$$

Each contribution to this sum depends on the orientations at adjacent segments i and $i-1$, and on the occupation variable associated with them:

$$\beta E_i = \frac{a}{2}|\hat{\mathbf{t}}_i - \hat{\mathbf{t}}_{i-1}|^2(1-n_i) + \left[\frac{a'}{2}(\hat{\mathbf{t}}_i \cdot \hat{\mathbf{t}}_{i-1} - \gamma)^2 - \mu\right] n_i - \frac{bf}{2}(\hat{\mathbf{t}}_i + \hat{\mathbf{t}}_{i-1}) \cdot \hat{\mathbf{z}}, \tag{2}$$

where $\beta = (k_B T)^{-1}$.

The terms in Eq. (2) control, respectively, the bending stiffness of bare DNA (a), the bending of protein-bound and bent DNA (a'), the binding free energy of the DNA-bending proteins (μ), and the straightening effect of the externally applied force f. The applied tension is thus along the z axis. The contribution of the final term to βE can be written as $f\hat{\mathbf{z}} \cdot (\mathbf{r}_N - \mathbf{r}_0)$, where $\mathbf{r}_0$ and $\mathbf{r}_N$ are the two ends of the chain of segments (recall $\mathbf{r}_k = \mathbf{r}_0 + b\Sigma_{i=1}^{k} \hat{\mathbf{t}}_i$). We reduce force by $k_B T$, so that in our formulas f has units of inverse length. At room temperature (the only temperature considered here), $f = 1\ \mathrm{nm}^{-1}$ corresponds to a physical force of $1 k_B T\ \mathrm{nm}^{-1} = 4.1$ pN.

The dimensionless a and a' are dimensionless parameters setting the bending stiffness of bare and protein-bound DNA segments. Comparison with the continuum semiflexible polymer shows that $A=ab$ is the persistence length of bare DNA. In this paper, we will use a lattice cutoff b from 1 to 5 nm, and therefore a will be in the range 10 to 50. Depending on the rigidity of the DNA-protein complex, one could imagine either $a'>a$ or $a'<a$; we will show results for both cases. For small angular distortions, the protein-bound joint bending energy is $\approx(a'/2)(\sin^2\psi)(\theta-\psi)^2$, where θ is the angle between adjacent tangent vectors, i.e., $a'\sin^2\psi$ is the harmonic bending elastic constant analogous to a.

The parameter μ is the free-energy difference between a protein bound to a site on the DNA and the same protein in free solution. Note that μ includes the binding free energy and the unbound protein entropy, i.e., $\mu=\epsilon_b+\ln c$, where c is the bulk solution protein concentration (in units of the inverse molecular volume) and where $\epsilon_b>0$ is the binding (free) energy. Thus, the choice $\mu=0$ approximately corresponds to the dissociation point, i.e., concentration $K_d \approx e^{-\epsilon_b}$. This neglects a free-energy shift due to the change in flexibility of the protein-DNA complex relative to that of bare double helix. Precise determination of the dissociation point requires computation of the value of μ at which a binding site on an unstretched chain is half occupied.

The partition function including DNA bending and protein-binding fluctuations along the chain is, for fixed orientations of the boundary segments,

$$\mathcal{Z}(\hat{\mathbf{t}}_0,\hat{\mathbf{t}}_N)=\int d^2t_1\cdots\int d^2t_{N-1}\sum_{n_1,\ldots,n_N}\exp[-\beta E], \tag{3}$$

where E is the total energy (1), and where the integrals of $\hat{\mathbf{t}}_i$ are over the 4π solid angle of the unit sphere. The cases of ''bare'' and ''saturated'' protein binding ($\mu=-\infty$ and $+\infty$, respectively) may be handled by noting that they correspond to the cases where respectively only the $n_i=0$ or $n_i=1$ states are included in the sum of Eq. (3). For these cases the value of μ is irrelevant since no fluctuations of n_i occur.

The partition function (3) has not yet had the orientations of the boundary segments $\hat{\mathbf{t}}_0$ and $\hat{\mathbf{t}}_N$ integrated. If we take periodic boundary conditions ($\hat{\mathbf{t}}_N=\hat{\mathbf{t}}_0$) and integrate over the boundary orientation, the partition function takes the form of a trace of a matrix product:

$$Z=\int d^2t_0\mathcal{Z}(\hat{\mathbf{t}}_0,\hat{\mathbf{t}}_0)=\mathrm{Tr}(T^N), \tag{4}$$

where the transfer matrix T is

$$T(\hat{\mathbf{t}}_{i-1},\hat{\mathbf{t}}_i)=\sum_{n_i=0}^{1}\exp[-\beta E_i]. \tag{5}$$

The partition function follows as

$$Z=\sum_k \lambda_k^N, \tag{6}$$

where λ_k are the eigenvalues of T. We suppose that the eigenvalues are labeled so that $\lambda_1>\lambda_2>\cdots$. Then,

$$\ln Z=N\ln\lambda_1+O\left(\frac{\lambda_2}{\lambda_1}\right)^N. \tag{7}$$

The largest eigenvalue dominates the free energy, with exponentially small finite-length corrections. Other choices of boundary conditions (e.g., integration over $\hat{\mathbf{t}}_0$ and $\hat{\mathbf{t}}_N$ as independent variables) will not change the form of Eq. (7). In this paper we will consider the thermodynamic limit $N\to\infty$ where the largest eigenvalue λ_1 dominates Eq. (7).

To compute the spectrum of T it is convenient to express it using the basis of spherical harmonics:

$$\langle l,m|T|l',m'\rangle=\int d^2tY^*_{lm}(\hat{\mathbf{t}})\int d^2t'Y_{l'm'}(\hat{\mathbf{t}}')T(\hat{\mathbf{t}},\hat{\mathbf{t}}'). \tag{8}$$

The transfer matrix (5) is invariant under azimuthal rotations $(\phi,\phi'\to\phi+\phi_0,\phi'+\phi_0)$. Thus, it is a function of $\phi-\phi'$, and independent of $\phi+\phi'$.

In addition, energy (5) is invariant under shifts of $\phi-\phi'$ by 2π, as can be seen from Eq. (2). Therefore, Eq. (5) can be expanded using Fourier components $e^{in(\phi-\phi')}$, where $n=0,\ \pm1,\ \pm2,\ldots$. Use of this expansion partially diagonalizes Eq. (8):

$$\langle l,m|T|l',m'\rangle=\delta_{m,m'}\int_{-1}^{1}dxP_{lm}(x)\int_{-1}^{1}dx'P_{l'm}(x')\times\int_{-\pi}^{\pi}d(\phi-\phi')\cos[m(\phi-\phi')]T(\hat{\mathbf{t}},\hat{\mathbf{t}}'), \tag{9}$$

where we have used the invariance of Eq. (2) under $\phi-\phi'\to\phi'-\phi$. The P_{lm} functions are normalized associated Legendre functions (the polar parts of the spherical harmonics), and an overall constant has been omitted for clarity. Equation (9) is diagonal in m, and the $m=0$ block must contain the largest eigenvalue.

The innermost ϕ integral of Eq. (9) was calculated using adaptive-grid integration. The outer double integral was calculated as a sum on an $M\times M$ point grid of points covering the square (x,x') integration domain. In practice, cutoffs on l and M must be chosen. This allows the real symmetric matrix (9) to be computed and diagonalized. The maximum eigenvalue λ_1 and its derivatives must then be checked to be unaffected, to the numerical accuracy desired, by increasing the cutoff value of l. The computations of Secs. II and III were done using $M=101$ and $l=8$. We have verified that the $m=0$ block of Eq. (5) contains the largest eigenvalue.

From the energy function of model (2) the thermally averaged end-to-end extension of the chain in the force (z) direction is

$$\hat{\mathbf{z}}\cdot\langle\mathbf{r}_N-\mathbf{r}_0\rangle=\frac{\partial\ln Z}{\partial f}=N\frac{\partial\ln\lambda_1}{\partial f}. \tag{10}$$

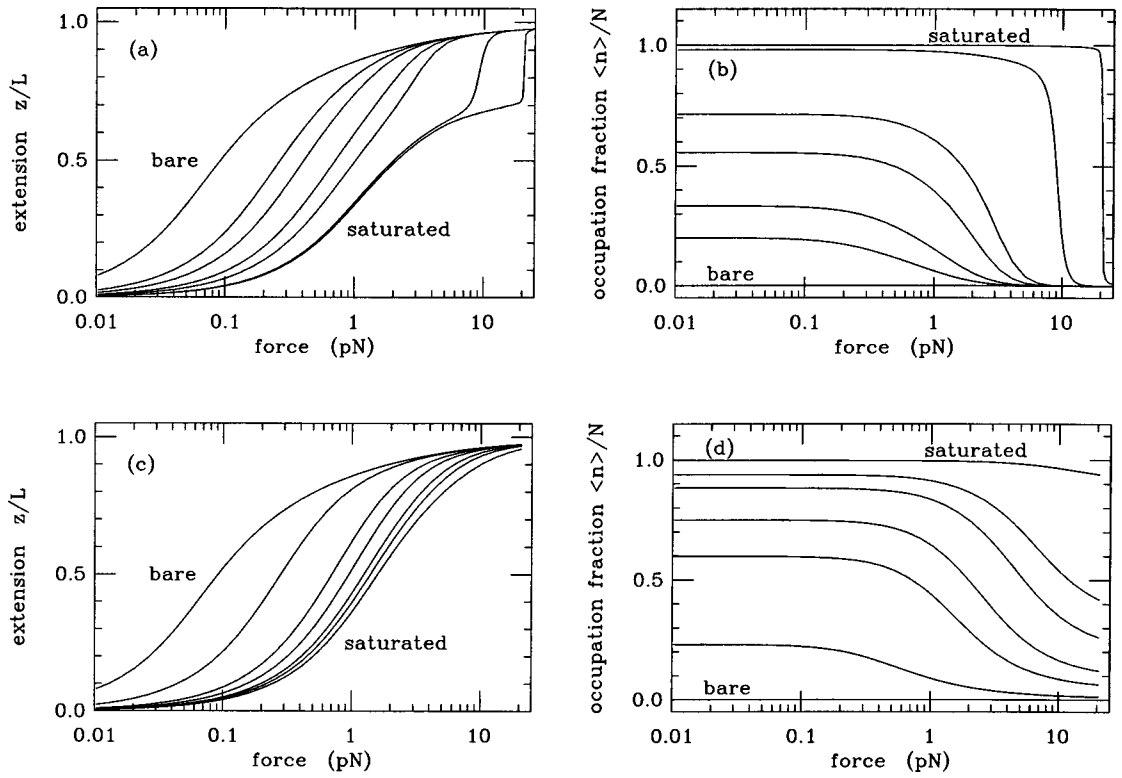

FIG. 2. Nonspecifically binding DNA-bending protein; binding induces a 90° bend ($\gamma=0$). Segment length is $b=5$ nm and bend modulus is $a=10$ (corresponding to bare-DNA persistence length of 50 nm). (a) Force-extension curves for stiff protein-DNA complex ($a'=100$) for binding strengths $\mu=-\infty$ (bare DNA, leftmost curve), $-2.30,-1.61,-0.69$, 0, 3.00, and 6.91. The DNA straightens out and the protein unbinds at a characteristic force. (b) Protein-binding site occupation corresponding to (a). The same set of binding strengths. At high-binding strength, the high stiffness of the DNA-protein complex causes an abrupt unbinding transition, corresponding to the abrupt extension increase observed in the corresponding curves of (a). (c) Force-extension curves for flexible protein-DNA complex ($a'=2$) for binding strengths $\mu=-\infty$ (bare DNA, leftmost curve), $-3.91,-2.30,-1.61,-0.69$, 0, and 3.00. At low-binding strength, the force-extension curves are similar to those of (a), but at high-binding strength no abrupt extension is observed because of the capacity of the DNA-protein complex to deform. (d) Protein-binding site occupation corresponding to (c).

Below, we will give chain extension as a fraction of the total contour length $L=Nb$. Similarly, the total number of proteins bound is

$$\sum_{i=1}^{N}\langle n_i\rangle=\frac{\partial \ln Z}{\partial \mu}=N\frac{\partial \ln \lambda_1}{\partial \mu}. \tag{11}$$

We will give occupation as a fraction of the total number of available binding sites N.

A. DNA-bending protein

The first situation we will describe is where a bending protein can bind at any segment. We suppose that when a protein binds, it stiffens the double helix as well as defines a preferred bend angle. In this case, the protein will tend to unbind under tension, rather than have the protein-DNA composite deform.

Figure 2 shows results for the case where the protein induces a 90° bend, i.e., $\gamma=0$, for a segment length $b=5$ nm. The bare DNA bending modulus is $a=10$ (corresponding to a bare DNA persistence length of $A=ab=50$ nm, and the bending modulus of the DNA-protein complex is $a'=100$. Figure 2(a) shows force-extension curves, while Fig. 2(b) shows protein occupations for the binding strengths $\mu=-\infty$, -2.30, -1.61, -0.69, 0, 3.00, and 6.91. For $\mu=-\infty$ [leftmost curve of Fig. 2(a)], the protein is never bound, and the force response is just that of the discretized wormlike chain model. As μ is made more positive, the protein starts to bind, and has the highest occupation at low force [left part of curves in Fig. 2(b)]. For $\mu\leqslant 0$, the protein gradually unbinds as the DNA is straightened out [compare Figs. 2(a) and 2(b)], and the force curves shift to the right of Fig. 2(a). Although there is a slight distortion of the shape of the force-extension curve, in these relatively weak-binding cases the effect of increasing protein concentration on the force-extension curve can be quite well described as a decrease in the effective persistence length.

For larger $\mu>0$ [rightmost two curves in Figs. 2(a) and 2(b)], the effects become much stronger. Now, the protein does not unbind as the DNA is straightened, and one starts to observe an appreciable reduction in the apparent total contour length, corresponding to the straightening of a nearly saturated and, therefore, ''zigzag'' protein-DNA composite fiber. This fiber shows some stretching elasticity, followed by an abrupt protein unbinding transition. The force-distance curve shows a force 'plateau' similar to the cooperative pla-

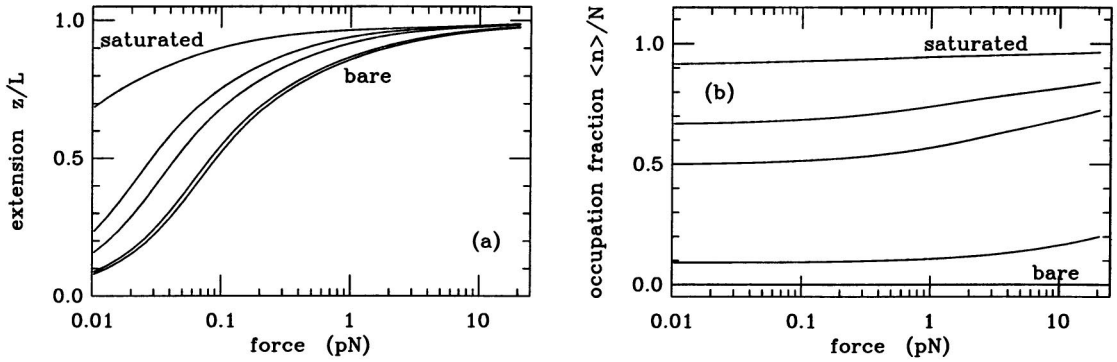

FIG. 3. Effect of DNA-stiffening protein which binds nonspecifically. Bare-DNA parameters are as in Fig. 2 ($b=5$ nm, $a=10$); when protein binds, the preferred joint angle stays straight ($\gamma=1$), but the DNA-protein complex is stiff ($a'=100$). Results are shown for $\mu=-\infty$ (rightmost curves, again corresponding to bare DNA), 0.00, 2.30, 3.00, and 4.60. (a) Force-extension curves gradually shift to lower forces (left) as μ is increased due to an increase in effective persistence length; contrast with results for DNA-bending protein with otherwise the same parameters ($\gamma=0$, Fig. 2). (b) Protein occupation, which gradually increases with μ. This shows much smoother variation with force than when occupation is coupled to chain bending [Fig. 2(b)].

teau observed for the overstretching of bare double-stranded (dsDNA). A sharp transition is observed in this protein unbinding model for $\mu=6.91$ even without direct protein-binding cooperativity; interaction between the binding degrees of freedom is induced by the coupling of binding and bending.

Figures 2(c) and 2(d) show results for parameters as considered above ($a=10, b=5$ nm, $\gamma=0$) but for the case where the protein-DNA complex is flexible ($a'=2$). Results are shown for binding strengths $\mu=-\infty$ (bare DNA, leftmost curve), -3.91, -2.30, -1.61, -0.69, 0, and 3.00. At low binding strength the force-extension curves are similar to those of Fig. 2(a), but at high binding strength no abrupt extension is observed due to the capacity of the DNA-protein complex to deform. There is a shift in μ, corresponding to protein concentration at K_d, due to the different flexibilities and, therefore, free energies of the protein-DNA complexes in the stiff and flexible cases. Figure 2(d) shows the protein binding as a function of force, and again there is no sharp unbinding transition. In the strong-binding cases, the protein can only be partially dissociated by high forces, since the protein-DNA complex can be straightened with only a fraction of the dissociation free energy.

A number of proteins are good candidates for experiments to study these effects. The non-sequence-specific proteins HU from *E. coli* and HMGB1 from eukaryote cells are both able to bend DNA by $>90°$ over ≈ 30 bp. The *E. coli* protein IHF also is known to be able to induce $>90°$ bends of DNA [7], and Ali *et al.* indeed have observed effects of the type shown in Fig. 2(a) [15]. It should be noted that IHF is known to bind to different DNA sequences with a wide range of affinities; in the experiments of Ali *et al.* the maximum number of proteins bound to a λ-DNA was about 300, i.e., a spacing of about 150 bp. The maximum compaction of DNA by IHF was therefore less dramatic than that of Fig. 2 where up to one protein every 15 bp (5 nm) can be bound. This situation can be described by spacing protein-binding segments from ''nonbinding'' segments using the approach of Sec. III.

Recent results of Schnurr [24] indicate that HU, which is thought to bind with less sequence specificity than IHF, causes a large shift of dsDNA force-extension curves which is similar to that of Figs. 2(a) and 2(c). Whether sharp unbinding transitions at a high force can be observed for large μ (i.e., large concentration) as in the stiff DNA-protein complex case [Figs. 2(a) and 2(b)], or alternately whether the complex can be deformed without dissociation [Figs. 2(c) and 2(d)], remains to be seen.

B. DNA-stiffening protein

Another possibility is that when a protein binds, it does not bend the double helix, but instead appreciably stiffens it. To examine this case, we consider the slight modification of Eq. (2);

$$\beta E_i=\frac{a}{2}|\hat{\mathbf{t}}_i-\hat{\mathbf{t}}_{i-1}|^2(1-n_i)+\left(\frac{a'}{2}|\hat{\mathbf{t}}_i-\hat{\mathbf{t}}_{i-1}|^2-\mu\right)n_i - \frac{bf}{2}(\hat{\mathbf{t}}_i+\hat{\mathbf{t}}_{i-1})\cdot\hat{\mathbf{z}}. \tag{12}$$

In the two binding states ($n_i=0,1$), the segment bending elasticity is of the same form, but with different bending constants.

Figure 3 shows results for this model, again for segment length $b=5$ nm, bare DNA bending modulus $a=10$ (again $A=ab=50$ nm for bare DNA), and a protein-bound modulus $a'=100$, corresponding to a protein-DNA composite persistence length of 500 nm. This value is taken to illustrate the effect of a large increase in stiffness caused by protein binding. Results are shown for $\mu=-\infty$ (again corresponding to bare DNA), 0.00, 2.30, 3.00, and 4.60.

The force-extension curves [Fig. 3(a)] show a gradual shift to lower forces (to the left) as the binding strength μ is increased. This is a result of the increase in chain persistence length and consequent reduction of the force $\approx k_BT/A_{\text{eff}}$ needed to extend the chain out of a random-coil conforma-

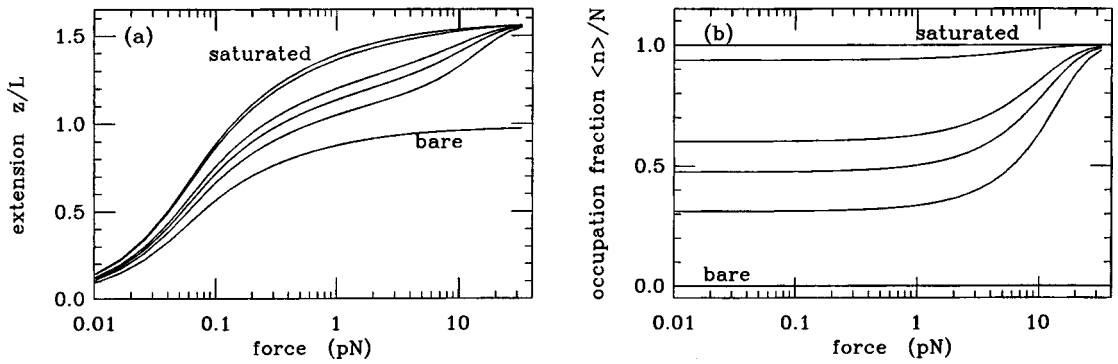

FIG. 4. Effect of intercalating ligands, which stretch the double helix when they bind. A short segment $b=1$ nm is used, corresponding to one ligand binding every three base pairs. For bare DNA, $a=50$; for ligand-bound DNA, $a'=40$ corresponding to a slight reduction in bending stiffness. Results are shown for $\mu=-\infty$ (bare DNA), -1.20, -0.51, 0 (near K_d), 2.3 (about ten times K_d), and ∞ (saturated binding). No permanent bends result from binding ($\gamma=0$). (a) Force-extension curves; increasing μ correspond to longer total extensions. (b) Binding fraction as a function of force. Larger μ corresponds to larger occupations. Force enhances binding; near K_d this enhancement is the largest.

tion [25] as the DNA-stiffening protein binds. The force constant observed after full extension of the protein-DNA complex will likely be larger than that of bare dsDNA.

Two other differences of the trends with increasing μ relative to the results of the preceding section can be observed. First, increasing force weakly increases the protein binding due to a reduction in the larger conformational entropy of the non-protein-bound chain by force. Second, even for large values of μ, no abrupt changes in protein occupation occur; the DNA becomes saturated with protein and has its persistence length shifted up, and increasing force just weakly increases the degree of saturation.

Amit *et al.* [10] have carried out single-DNA experiments that indicate that the *E. coli* protein H-NS behaves along these lines. With increasing protein concentration, Amit *et al.* observed force-extension curve shifts similar to those in Fig. 3(a). For buffer with 50-mM KCl, effective persistence lengths increasing from the bare DNA value of 50 nm to 130 nm were observed over an H-NS concentration range of $0-4$ μM. At higher protein concentrations, even stiffer H-NS-DNA filaments may occur since saturation of the effective persistence length was not reached by Amit *et al.* (see Fig. 4 of Ref. [10]).

Another protein known to stiffen the double helix is RecA, which polymerizes to form a sheath around the double helix. RecA+dsDNA has been studied using single-molecule techniques [26,27] which indicate a persistence length of the RecA-dsDNA complex $\approx$950 nm [27]. The model of this section needs the additional feature of strong cooperativity (a term in the energy $\approx n_i[1-n_{i-1}]$) to be realistically applied to cases such as RecA where the protein essentially polymerizes along the molecule. In this case, a gradual increase in the persistence length with protein concentration (as seen by Amit *et al.* for H-NS [10]) is difficult to observe, since RecA binding occurs by a nucleation-and-growth process [26].

C. DNA-intercalating drugs

Another case of interest is where molecules bind to the double helix and stretch it without introducing kinks. A number of drugs and proteins insert ligands into the double helix to accomplish this; familiar examples include the ''intercalating'' DNA dyes ethidium bromide (EthBr) and YOYO. The binding of such drugs is described by a modification of Eq. (12):

$$\beta E_i=\frac{a}{2}|\hat{\mathbf{t}}_i-\hat{\mathbf{t}}_{i-1}|^2(1-n_i)+\left(\frac{a'}{2}|\hat{\mathbf{t}}_i-\hat{\mathbf{t}}_{i-1}|^2-\mu\right)n_i$$
$$-\frac{bf}{2}(1+\alpha n_i)(\hat{\mathbf{t}}_i+\hat{\mathbf{t}}_{i-1})\cdot\hat{\mathbf{z}}. \qquad (13)$$

Now, binding at site i changes the bending stiffness from a to a' and the segment length from b to $(1+\alpha)b$.

Figure 4 shows results for $b=1$ nm, $a=50$, $a'=40$, and $\alpha=0.5$. This is a model for a drug which can bind every three base pairs, and which when saturated on DNA, stretches the double helix to 1.5 times its *B*-form contour length. Results are shown for $\mu=-\infty$ (bare DNA), -1.20 (weak binding), -0.51, 0 (near the K_d of the drug), 2.3 (concentration ten times the $\mu=0$ case, nearly saturated), and for $\mu=+\infty$ (saturated). The bare and saturated cases reduce to discrete wormlike chains with persistence lengths of 50 and 40 nm [see Fig. 4(a)].

For case $\mu=0$, at zero force the drug binding sites are about 60% occupied [Fig. 4(b)]. As force is increased, the occupation fraction is driven higher. The free-energy drive for this is the mechanical work obtained by extending of the double helix via drug binding. In the force-distance curve, the signature of this force-induced binding is a transition towards a higher degree of stretching. This transition is sharper than the gradual stretching transition in the DNA-stiffening case [Sec. II B, Fig. 3], but is much less sharp than the transitions induced by release of DNA-bending proteins from stiff DNA-protein complexes [Sec. II A, Figs. 2(a) and 2(b)]. For $\mu=2.3$, drug binding is nearly saturated at low force, and is only slightly more concentrated on the DNA as force is increased.

A few single-DNA experiments have been reported where intercalants were used. Saturating concentrations of EthBr have been observed to produce results similar to the intermediate curves of Fig. 4(a), in studies by Cluzel *et al.* [28], and by Hegner *et al.* [29]. In the study of Husale *et al.*, 2.5-μM EthBr was observed to both lengthen and reduce the persistence length of a single dsDNA in a way similar to that observed for the intermediate concentration curves of Fig. 4(a). Rather similar experiments have been performed using YOYO [30], but again at one concentration, and also quite far from binding equilibrium. Experiments for a series of concentrations have not been reported for any intercalator. A series of curves at different concentrations would allow better contact with the present model, and would give some indication whether additional effects such as binding cooperativity are important.

III. FIXED-LOCATION BENDS

The model of the preceding section can be adapted to the situation where bends are positioned at specific points along a DNA. Sequence-specific DNA-bending proteins (e.g., transcription factors such as TATA-binding protein [21], or restriction enzymes [14]) generally bind their targets tightly, and therefore we will not calculate association-dissociation equilibria as in the preceding section. Instead, we will consider the situation where DNA-bending proteins are permanently bound at specific positions. Our main interest is in determining the inter-protein separation at which the mechanical effects of the DNA-bending proteins become negligible.

We consider a region of a large molecule made up of M segments of the preceding section, again each of contour length b. The energy will be taken to be that of $M-1$ unbent segments, terminated by one segment with a preferred bend angle:

$$\beta E_M = \frac{a'}{2}(\hat{\mathbf{t}}_M \cdot \hat{\mathbf{t}}_{M-1} - \gamma)^2 + \frac{a}{2}\sum_{i=1}^{M-1} |\hat{\mathbf{t}}_i - \hat{\mathbf{t}}_{i-1}|^2 - \frac{bf}{2}\sum_{i=1}^{M} (\hat{\mathbf{t}}_i + \hat{\mathbf{t}}_{i-1}) \cdot \hat{\mathbf{z}}. \quad (14)$$

We imagine that a chain is made up of S of these regions, each consisting of $M_j - 1$ unbent segments and one bent segment, where $j = 1, \ldots, S$. Following the approach of the preceding section, the partition function can be expressed in terms of a matrix product as

$$Z = \mathrm{Tr}\left(\prod_{j=1}^{S} T_0^{M_j - 1} T_\gamma \right), \quad (15)$$

where the transfer matrices of unbent and bent segments are

$$T_0(\hat{\mathbf{t}}, \hat{\mathbf{t}}') = \exp\left[-\frac{a}{2}|\hat{\mathbf{t}} - \hat{\mathbf{t}}'|^2 + \frac{bf}{2}(\hat{\mathbf{t}} + \hat{\mathbf{t}}') \cdot \hat{\mathbf{z}} \right],$$

$$T_\gamma(\hat{\mathbf{t}}, \hat{\mathbf{t}}') = \exp\left[-\frac{a'}{2}(\hat{\mathbf{t}} \cdot \hat{\mathbf{t}}' - \gamma)^2 + \frac{bf}{2}(\hat{\mathbf{t}} + \hat{\mathbf{t}}') \cdot \hat{\mathbf{z}} \right], \quad (16)$$

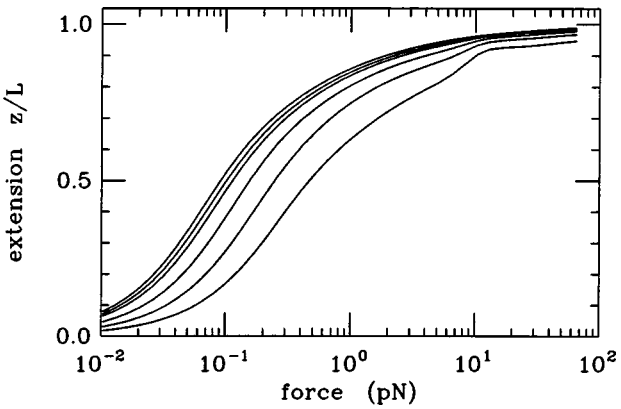

FIG. 5. Permanent bends ($90°, \gamma = 0$) at regularly spaced positions along a dsDNA. Bare DNA has segment length $b = 5$ nm, $a = 10$ (persistence length 50 nm). From left to right, the permanent bend spacings are ∞ (bare DNA), 500 nm (1500 bp), 250 nm (750 bp), 100 nm (300 bp), 50 nm (150 bp), and 25 nm (75 bp). For spacings of 100 nm or less there is a strong effect, but for wider spacings the force-extension curve is essentially indistinguishable from that of bare DNA. For the shortest spacing (rightmost curve, 25 nm), the force-extension curve has a distinct shape with a ''knee'' corresponding to the straightening of the essentially zigzag permanent bends.

respectively. The total number of segments in the polymer is $N = \Sigma_{i=1}^{S} M_i$. Calculation of the extension as a function of force is as in Eq. (10).

A. Detecting widely spaced severe bends along a DNA

We first consider a long uniform DNA, with periodically spaced bends, i.e., $M_i = M$ and $N \gg M$. The partition function is just the largest eigenvalue of the transfer matrix (15). Figure 5 shows force-extension curves for a series of bend spacings compared with uniform DNA. Parameters for the calculation match those used above, i.e., segment length $b = 5$ nm (15 bp), straight-segment persistence length $a = 10$ ($A = 50$ nm), bend elasticity $a' = 100$, and bend angle $\gamma = 0$ (90°). This situation corresponds to proteins bound irreversibly to specific sites that are spaced by DNA contour length Mb. The bend angle is rather severe, but similar to bends introduced by transcription factors, restriction enzymes, and other sequence-specific DNA-binding proteins [14].

The spacings shown in Fig. 5 are, from left to right, $M = \infty$ (bare DNA), $M = 100$ (500 nm = 1500 bp), $M = 50$ (250 nm = 750 bp), $M = 20$ (100 nm = 300 bp), $M = 10$ (50 nm = 150 bp), and $M = 5$ (25 nm = 75 bp). As the spacing of the fixed bends is decreased, the elastic response shifts to larger forces, since the permanent bends act to fold up the molecule. This effect is weak until the bends are about 300 bp apart, corresponding to one bend per statistical segment length (recall that for the semiflexible polymer the statistical segment length is double the persistence length). The 1500 and 750 bp curves of Fig. 5 will be indistinguishable from that of bare DNA in an experiment. Therefore, the binding of even strong DNA-bending proteins causes an experimentally negligible effect on DNA polymer elasticity unless

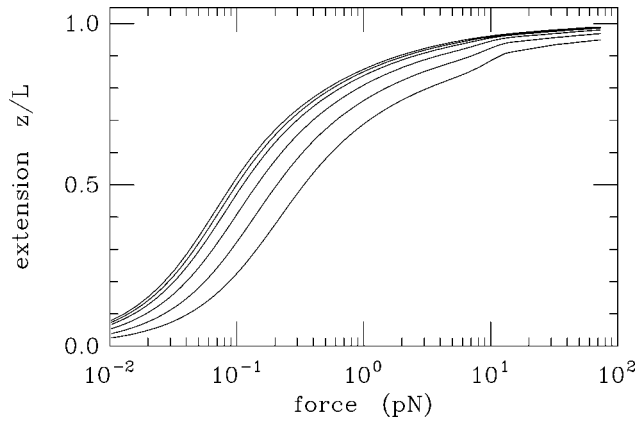

FIG. 6. Permanent bends $(90°, \gamma=0)$ at random positions along a dsDNA. This calculation is done for a 45-kb molecule, consisting of 3000 segments each of length $b=5$ nm. Bends are inserted randomly into segments with probability p; from left to right, $p=0$ (bare DNA), 0.01, 0.02, 0.05, 0.1, and 0.2. The curves thus correspond to mean spacings of bends equal to the fixed spacings of Fig. 5. As the bend spacing is reduced, the curves essentially shift to larger forces, corresponding to a gradual reduction in effective persistence length. As in Fig. 5, this shift is negligible until there is one permanent bend introduced per 100 nm (third curve from right).

the bends are less than $2A=300$ bp apart. For short interbend spacings (Fig. 5, 75 bp), the elastic response becomes strongly shifted to larger forces. We note that for bends spaced by less than about 300 bp, the ''opening'' of the bend-protein-DNA complex leads to a cusplike feature near 10 pN, $\approx k_BTa'/b$, which is the ratio of the work needed to straighten the complex to the $\approx b$ lengthening thus obtained.

B. Bends with fixed random location

We now consider randomly located bends. This is the only finite-N calculation of this paper, and is done for a chain corresponding to a 45-kb DNA, of contour length ≈ 15 μm, comparable to molecule lengths studied experimentally. The model is essentially that of the preceding section, but with M_i's resulting from the random assignment of segments to be bent or straight. The partition function of this model is thus based on a random product of matrices:

$$Z=\mathrm{Tr}\left[\prod_{i=1}^{N} T_i\right]. \tag{17}$$

The matrices T_i are just those of Eq. (16), assigned according to the ''random sequence'' of bent vs unbent monomers. The sequence is fixed at the start of the calculation; bent monomers occur with fixed probability p.

We set a segment length $b=5$ nm (15 bp), and thus for $\approx 45\,000$ bp we take $N=3000$. Straight segments are considered to have bending constant $a=10$, corresponding to straight-DNA thermal persistence length $A=50$ nm. Bent segments are taken to have a fixed value of $\gamma=0$ as in the preceding section, and a bending rigidity $a'=10$. Figure 6 shows results for bent-site probability $p=0$ (bare DNA), $p=0.01$ (1500 bp on an average between bends, leftmost curve), $p=0.02$ (750 bp), $p=0.05$ (300 bp), $p=0.1$ (150 bp), and $p=0.2$ (75 bp, rightmost curve). Similar to the results for regularly spaced DNAs, the effect of random bends is experimentally undetectable until one reaches roughly one bend per statistical segment ≈ 300 bp. For the closest spacing of bends (75 bp), there is a slight difference between the shapes of the curves in the equal- and random-spacing cases, but for wider spacings, there is little difference between the results of the two cases.

IV. CONSTRAINT OF DNA LINKING NUMBER

Many DNA-bending proteins are known to produce DNA unwinding [7,14] in the vicinity of the bend. On purely theoretical grounds the expulsion of twist from a sharp bend can be argued to be a generic result of the different bending rigidities of the ''ladder'' of base pairs parallel and perpendicular to the base-pair ''rungs'' [31]. In the simplest type of single-DNA experiment where only one strand of the double helix is tethered at each end, or where there are ''nicks'' or breaks in one strand, untwisting induced by protein binding will not lead to any buildup of twisting stress in the DNA molecule, and thus the chain elasticity will have no dependence on the chain twisting.

Alternately, both strands of the double helix can be anchored at both ends. In such situations, the topological linking number (Lk) of the two single strands of the double helix becomes experimentally controllable. The torsional stress and the elasticity of the molecule become dependent on the double-helix linking number set experimentally [32], as observed in experiments of Strick *et al.* [33,34]. If the double-helix linking number is held fixed while allowing DNA-unwinding proteins to bind, there will be a similar change in the DNA elastic response.

In this section, we modify the model of the previous sections to include the effect of the fixed linking number. We do this by using an applied torque (below denoted τ, in units of k_BT) which we use to keep the double-helix linking number fixed. Our approach is a discretized version of the continuum elastic analysis of Fain *et al.* [35], and is also similar to the statistical-mechanical analyses of Moroz and Nelson [36] and Bouchiat and Mezard [37], but with additional protein-binding degrees of freedom.

We imagine that each protein that binds to a dsDNA generates an untwisting of $\Theta/(2\pi)$. The appropriate energy including torque and linking number contributions is that of Eq. (1), plus

$$\frac{c}{2}\sum_i (\Omega_i+\Theta n_i)^2-2\pi\tau\Delta\,\mathrm{Lk}. \tag{18}$$

The first term is the twist energy; the dimensionless parameter c is the twist modulus, and is related to the continuum twist modulus $C=bc$ for $b\to 0$. The variable Ω_i is the twist distortion of the double helix at segment i, and is expressed in terms of Euler angles θ_i, ϕ_i, and χ describing the orientation of segment i as [35]

$$\Omega_i=\chi_i-\chi_{i-1}+\tfrac{1}{2}(\phi_i-\phi_{i-1})(\cos\theta_i+\cos\theta_{i-1}) \tag{19}$$

to linear order in small differences of the angles (this expression reduces to the continuum result for $b \to 0$). Here, θ_i and ϕ_i are the polar and azimuthal angles describing $\hat{\mathbf{t}}_i$, while χ is the third Euler angle specifying a segment's rotation about its axis. The energetically preferred twist distortion in a segment is shifted to $-\Theta$ when a protein is present.

The second term in Eq. (18) describes the coupling of torque to the double-helix linking number. For a thin-ribbon model of the double helix, the linking number can be broken into a double-helix twist, and backbone ''writhe'' contributions, $\Delta\,\mathrm{Lk} = \Delta\,\mathrm{Tw} + \mathrm{Wr}$. The writhe measures the chiral bending of the backbone, and is zero for an entirely straightened molecule [32]. The total change in the double-helix twist away from its equilibrium value of one twist per 10.5 bp is just the sum of the individual segment twists, i.e.,

$$2\pi\Delta\,\mathrm{Tw} = \sum_i \Omega_i . \tag{20}$$

The writhe is problematic, being a nonlocal function of the $\hat{\mathbf{t}}_i$. However, for small distortions away from a straight conformation (i.e., $|\hat{\mathbf{t}}_i - \hat{\mathbf{z}}| \ll 1$) the writhe is [35]

$$4\pi\,\mathrm{Wr} = \sum_i (\phi_i - \phi_{i-1})(2 - \cos\theta_i - \cos\theta_{i-1}). \tag{21}$$

This formula restricts our theory to the case where the end-to-end extension is close to the total molecular length, i.e., to a sufficiently large force such that $\hat{\mathbf{t}} \approx \hat{\mathbf{z}}$. Description of cases where the molecule is distorted strongly away from a straight configuration (e.g., to form ''plectonemic'' supercoils or other self-wrapped configurations) requires exact accounting of Wr, and a numerical Monte Carlo calculation.

Plugging Eqs. (19)–(21) into Eq. (18) allows the twisting part of the energy to be written as

$$\sum_i \left\{ \frac{c}{2}\left[\chi_i - \chi_{i-1} + \frac{(\phi_i - \phi_{i-1})(\cos\theta_i + \cos\theta_{i-1})}{2} + \Theta n_i \right]^2 - \tau[\phi_i - \phi_{i-1} + \chi_i - \chi_{i-1}] \right\}. \tag{22}$$

The variables χ_i may be integrated out after completion of squares, leaving the effective energy associated with the applied torque, for segment i:

$$\beta F_{\tau,i} = \frac{1}{2}\ln c - \frac{\tau^2}{2c} + \tau\Theta n_i - \tau(\phi_i - \phi_{i-1}) \times \left[1 - \frac{1}{2}(\cos\theta_i + \cos\theta_{i-1}) \right]. \tag{23}$$

It should be noted that although here the twist modulus is the same for bare and protein-bound DNA (i.e., for $n_i = 0$ or 1), all that needs to be done is to make c dependent on n_i in Eq. (23). In this paper, this effect is not considered.

The torque effective energy (23) can be added to the bending-stretching energy (2). The transfer matrix for a DNA subject to binding of proteins which both bend and unwind the double helix follows as

$$T_\tau(\hat{\mathbf{t}}_{i-1}, \hat{\mathbf{t}}_i) = \sum_{n_i=0}^{1} \exp[-\beta(E_i + F_{\tau,i})]. \tag{24}$$

This transfer matrix can again be written in terms of spherical harmonic components. The simple form (9) which is diagonal in the m's does not occur, due to the appearance of the term proportional to $\tau(\phi - \phi')$ in Eq. (23). Therefore, while still invariant under rotations, Eq. (24) no longer is invariant under azimuthal rotations of 2π or under sign reversal. Instead, we obtain, again dropping an overall constant,

$$\langle lm|T|l'm'\rangle = \frac{1}{m - m'} \int_{-1}^{1} dx P_{lm}(x) \int_{-1}^{1} dx' P_{l'm'}(x') \times \int_{-2\pi}^{2\pi} d(\phi - \phi') \times \sin\left[(m - m')\left(\pi - \frac{|\phi - \phi'|}{2} \right) \right] \times \exp\left[\frac{i(m + m')(\phi' - \phi)}{2} \right] T(\hat{\mathbf{t}}, \hat{\mathbf{t}}'). \tag{25}$$

Except for the case $\tau = 0$ where Eq. (25) reduces to Eq. (9), we now have complex-valued matrix elements for $m \neq m'$. The transfer matrix is Hermitian and, therefore, has real eigenvalues.

Extension and protein occupations are still computed from the maximum eigenvalue of T_τ according to Eqs. (10) and (11). Now the average linking number can be computed as well, as a derivative on torque τ. Referring to Eq. (18) we see that

$$\Delta\,\mathrm{Lk} = \Delta\,\mathrm{Tw} + \mathrm{Wr} = \frac{N}{2\pi}\frac{\partial \ln\lambda_1}{\partial\tau} \tag{26}$$

The intensive $\sigma = \Delta\,\mathrm{Lk}/\mathrm{Lk}_0$ is therefore

$$\sigma = \frac{1}{b\omega_0}\frac{\partial \ln\lambda_1}{\partial\tau}, \tag{27}$$

where $\omega_0 = 2\pi/3.5$ nm is the spatial rotation rate along the unstressed double helix and as before, b is the contour length per segment.

Effects of DNA-bending-and-untwisting proteins on the force-extension behavior at a fixed linking number are shown in Fig. 7. The parameters are as in Sec. II A ($\gamma = 0, b = 5$ nm, $a = 10, a' = 100$). The only new parameter demanded by the fixed-linking calculation is $\Theta = 1$, corresponding to untwisting of the double helix by 1 radian when a protein binds (the TATA binding protein generates about double this unwinding and a $\approx 90°$ bend over its < 10 bp binding size). A relatively low binding free energy μ

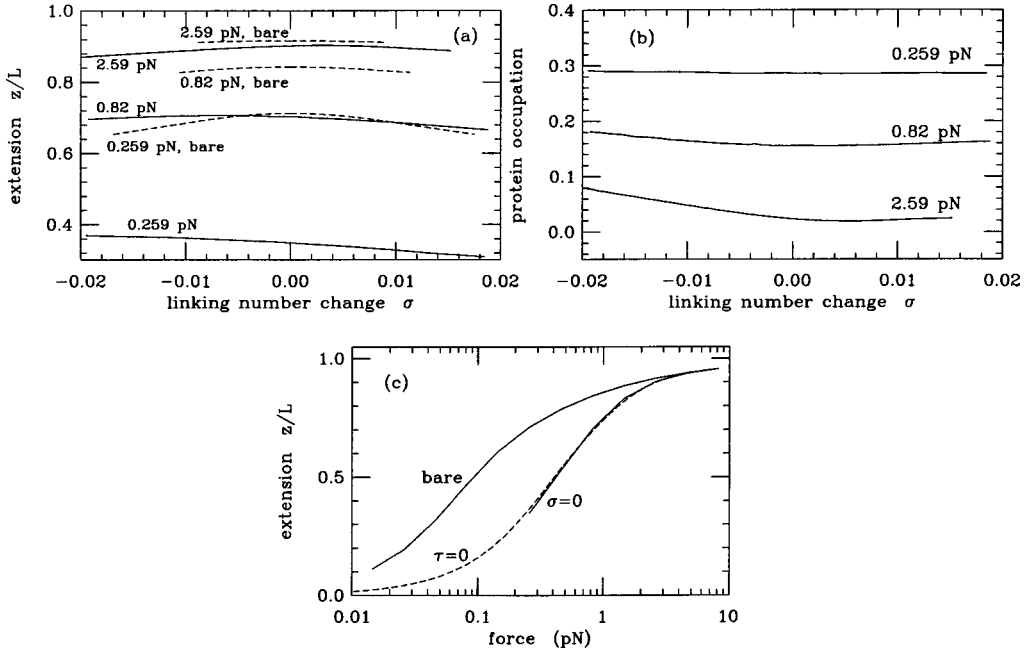

FIG. 7. Effects of DNA-bending-and-untwisting proteins on the force-extension behavior at a fixed linking number. DNA and protein parameters are as in Fig. 2, with also $\Theta=1$ corresponding to untwisting of the double helix by 1 radian when a protein binds, and a binding free energy $\mu=-1.61$. (a) Extension vs unwinding (fractional linking number change σ). Dashed curves show bare DNA ($\mu=-\infty$) and solid curves show protein effect ($\mu=-1.61$). Higher extensions correspond to a higher force. (b) Protein occupation vs unwinding corresponding to part (a), for $\mu=-1.61$. Unwinding the DNA enhances the protein binding, except at very large forces where the proteins are prevented from binding by the straightening of the DNA. Higher protein occupations correspond to lower forces. (c) Comparison of force-extension curves for bare DNA (left, solid), nicked DNA in presence of DNA-bending protein ($\tau=0$, dashed, result of Fig. 2 for $\mu=-1.61$), and twist-blocked DNA at $\sigma=0$ in the presence of DNA-bending protein (right, solid). The twist-blocked case is shifted to slightly higher forces, although the effect is weak since the protein occupation is relatively small.

$=-1.61$ is used, which would generate binding of about one-fifth of the segments by protein in absence of DNA tension or torque.

This computation differs from that of the previous sections, in that it is accurate only for sufficiently large forces such that the tangent vector is oriented appreciably along the force direction. At the same time, we have not incorporated structural transitions known to occur for the double helix at high forces and torques [26,38] into the present model. Therefore, the present calculation is applicable to experiment only for a sufficiently large force that $\langle \hat{\mathbf{t}}\cdot\hat{\mathbf{z}}\rangle > 0.4$ (the range where the corresponding Gaussian approximation is valid for the simple semiflexible polymer [25]), but not for such large forces and torques that one would expect the double helix to denature. The results discussed below are in this range of applicability.

Figure 7(a) shows extension versus unwinding (fractional linking number change σ). The dashed curves show bare DNA ($\mu=-\infty$), while solid curves are in presence of protein ($\mu=-1.61$). Higher extensions correspond to progressively higher forces; from bottom to top, Fig. 7(a) shows results for forces of 0.259, 0.82 and 2.59 pN. The bare DNA curves have their peaks at $\sigma=0$, with extension reduced by either overwinding or underwinding. In the presence of protein, the symmetry between overwinding and underwinding is broken, and the peaks of the extension-linkage curves are shifted away from $\sigma=0$. Corresponding protein-binding curves are shown in Fig. 7(b); binding is suppressed by DNA tension (as in Sec. II A) and is stimulated by DNA unwinding. For sufficient overwinding there is stimulation of protein binding due to the formation of chiral bends correlated with bound protein.

An interesting feature of the curves in Fig. 7(a) is that the shift of the extension peak to $\sigma<0$ at lower forces, but to $\sigma>0$ at high forces. At low forces, the chiral fluctuations induced by the linking number constraint strongly compact bare DNA; binding of protein relaxes the DNA torsional stress and causes the DNA to stretch out. The result is a peak in extension for $\sigma<0$. Conversely, at high forces, the compaction effect of the chiral fluctuations is supressed by applied tension; binding of protein therefore reduces DNA extension. Since protein binding is suppressed for $\sigma>0$ the peak extension is observed in that regime at high force.

Figure 7(c) shows force-extension curves in the presence of this DNA-bending-unwinding protein. Curves are shown for bare DNA, nicked DNA ($\tau=0$) in presence of DNA-bending protein (result of Fig. 2 for $\mu=-1.61$), and twist-blocked DNA at $\sigma=0$ in the presence of DNA-bending protein. The twist-blocked and nicked curves are almost identical for this case.

V. CONCLUSION

This paper has presented discrete semiflexible chain models for proteins or other molecules which deform the double

helix with which they bind. The major focus of this paper has been on DNA-bending proteins and, most roughly, we find that increasing tension favors DNA extension and, consequently, unbinding of protein in a manner qualitatively similar to that discussed previously [16].

We have discussed how proteins which bend the double helix generate an increase in entropic elastic stiffness which is well described as a reduction in the effective persistence length. The experimental signature that can be observed in current force-extension experiments is a gradual shift of log force vs extension curves to larger forces as protein concentration is gradually increased (Fig. 2). This effect has already been observed to some extent in experiments on IHF and HU protein by Amit and co-workers. We have also found that the stiffness of the protein-DNA complex impacts the sharpness of the high-force DNA extension and protein unbinding [Figs. 2(a) and 2(b) vs Figs. 2(c) and 2(d)].

Description of DNA elasticity in terms of an effective persistence length has been discussed in the case of Gaussian distributions of local permanent bends along DNA, in the limit of weak disorder that can be treated perturbatively [39,40]. It is somewhat surprising that the rather severe and isolated local bends of fixed bend angle studied in this paper, for either the cases where the bend locations are fluctuating (Fig. 2) or are fixed (Figs. 5 and 6) generate rather similar effects.

We have also discussed how proteins which bend and unwind the double helix (as HU and IHF are known to do) will impact DNA force-extension experiments with a fixed DNA linking number. At low force, unwinding stimulates protein binding, shifting the peak of the extension vs linking number curve in the direction of unwinding (Fig. 7). At high force, this shift can be reversed towards overwinding, since the reduction in DNA extension generated by the bends.

All our results are in the framework of equilibrium statistical mechanics. A basic requirement for application of our results for thermal fluctuation of protein occupation (Secs. II and IV) to experiments is that an extension measurement is made on a time scale long compared to the ''on'' and ''off'' times of the protein. The protein-binding degrees of freedom of DNA-distorting proteins may fluctuate slowly due to barriers to protein binding and unbinding. Applied tension may enhance these barriers, thus the measurement times required may vary with tension. For example, a DNA-bending protein might require thermal bending of DNA for its binding, which will be suppressed if the DNA is straightened out by applied tension.

In Sec. III, we consider an alternative ensemble of essentially permanent binding of protein, or the case where the off time is infinite. To apply this part of the paper to experiment, care must be taken to wait for the proteins to bind before carrying out extension measurements, and one requires independent knowledge that the off times are longer than the time of the experiment.

We speculate that tension-driven unbinding of proteins may be relevant biologically, since one expects tension to be generated on DNA *in vivo* during transcription, in any case where the RNA polymerase is in some way physically constrained. In *E. coli*, transcription and translation (assembly of amino acids into protein) of membrane proteins mechanically couple RNA polymerase to the cell membrane [41]. RNA polymerase is a linear motor capable of moving at tens of nm/sec and generating forces of up to 40 pN [42]. According to the estimates of the present paper, this level of force generation is sufficient to dissociate proteins bound to the double helix via tension conducted through the DNA (this is distinct from a contact mechanism where RNA polymerase directly pushes proteins off their binding site). If the dissociated proteins are transcription enhancers or repressors, or if their exit opens up binding sites for regulatory proteins, this amounts to a possible gene regulation mechanism, with DNA tension as part of the signal pathway.

Many extensions of the calculations of this paper are possible. First, many proteins bind to DNA with a relatively large binding site that excludes proteins over a range of more than a few nanometers (>10 bp). In these cases, the simple approach of setting the lattice parameter b to the binding-site size will start to generate artifacts that will degrade the model's description of dsDNA entropic elasticity (here, the <5 nm value of b used will modify the semiflexible polymer elasticity for forces $>k_BTA/b^2\approx 5$ pN at the top of the force range considered in this paper). A simple solution to this problem is to follow the approach of Sec. III to separate protein-binding vertices with always-bare vertices to build up whatever length of binding site is desired.

Cooperativity of protein binding has not been discussed in this paper, but is also easily incorporated. Coupling of adjacent protein-binding degrees of freedom can be included at the cost of deferring the initial summation over occupation numbers [Eq. (5)], and therefore doubling the rank of the transfer matrix. Most simply, binding cooperativity can be used to tune the width of the extension and protein-unbinding curves (e.g., Fig. 2). The bending flexibilities and spontaneous bend angles can be made functions of adjacent protein occupation numbers, allowing a description of the difference between the deformation of DNA generated by isolated proteins, and complexes of many adjacent proteins bound to double helix.

This paper has also not discussed the effect of orientation of bends with underlying double-helix structure; effectively, we have considered proteins which bend DNA with no attention to local helix orientation. For the 5-nm binding sites considered here and low occupations, this is likely a reasonable approximation, but for high protein occupation and especially in cases where there is strong cooperativity of binding, it would be useful to include phasing of bends with the double helix. This can be done by describing DNA base-pair orientations a triad of unit vectors [31], and then defining protein-generated bends in terms of base-pair orientation. The cost will again be an increase in transfer matrix dimension.

ACKNOWLEDGMENTS

This research was supported in part by the NSF through Grant Nos. DMR-9734178 and DMR-0203963, by a grant from the Johnson and Johnson Focused Giving Grant Program, and by a grant from Research Corporation.

[1] E.M. Blackwood and J.T. Kadonaga, Science **281**, 60 (1998).
[2] M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, and P. Lu, Science **271**, 1247 (1996).
[3] H. Xu and T.R. Hoover, Curr. Microbiol. **4**, 138 (2001).
[4] I.Y. Goryshin and W.S. Reznikoff, J. Biol. Chem. **273**, 7367 (1998).
[5] S.E. Milsom, S.E. Halford, M.L. Embleton, and M.D. Szczelkun, J. Mol. Biol. **311**, 515 (2001).
[6] N.J. Trun and J.F. Marko, Am. Soc. Microbio. News **64**, 276 (1998).
[7] P. Rice, S. Yang, K. Mizuuchi, and H.A. Nash, Cell **87**, 1295 (1996).
[8] S.W. White, K.S. Wilson, K. Appelt, and I. Tanaka, Acta Crystallogr., Sect. D: Biol. Crystallogr. **55**, 801 (1999).
[9] K. Wojtuszewski, M.E. Hawkins, J.L. Cole, and I. Mukerji, Biochemistry **40**, 2588 (2001).
[10] R. Amit, A.B. Oppenheim, and J. Stavans, Biophys. J. **84**, 2467 (2003).
[11] J.O. Thomas and A.A. Travers, Trends Biochem. Sci. **26**, 167 (2001).
[12] J.O. Thomas, Biochem. Soc. Trans. **29**, 395 (2001).
[13] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond, Nature (London) **389**, 251 (1997).
[14] S. Jones, P. van Heyningen, H.M. Berman, and J.M. Thornton, J. Mol. Biol. **287**, 877 (1999).
[15] B.M.J. Ali, R. Amit, I. Braslavsky, A.B. Oppenheim, O. Gileadi, and J. Stavans, Proc. Natl. Acad. Sci. U.S.A. **98**, 10 658 (2001).
[16] J.F. Marko and E.D. Siggia, Biophys. J. **73**, 2173 (1997).
[17] L. Finzi and J. Gelles, Science **267**, 378 (1995).
[18] J. Rudnick and R. Bruinsma, Biophys. J. **76**, 1725-33 (1999).
[19] H. Diamant and D. Andelman, Phys. Rev. E **61**, 6740 (2000).
[20] C. Storm and P. Nelson, e-print condmat physics/0206088.
[21] N.A. Davis, S.S. Majee, and J.D. Kahn, J. Mol. Biol. **291**, 249 (1999).
[22] X. Zhao and W. Herr, Cell **108**, 615 (2002).
[23] B. Teter, S.D. Goodman, and D.J. Galas, Plasmid **43**, 73 (2000).
[24] B. Schnurr (unpublished).
[25] J.F. Marko and E.D. Siggia, Macromolecules **28**, 8759 (1995).
[26] J.F. Leger, J. Robert, L. Bordieu, D. Chatenay, and J. Marko, Proc. Natl. Acad. Sci. U.S.A. **95**, 12295 (1998).
[27] M. Hegner, S.B. Smith, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **96**, 10109 (1999).
[28] P. Cluzel, A. Lebrun, Ch. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).
[29] S. Husale, W. Grange, and M. Hegner, Single Mol. **3**, 91 (2002).
[30] M.L. Bennink, O.D. Schärer, R. Kanaar, K. Sakata-Sogawa, J.M. Schins, J.S. Kanger, B.G. de Grooth, and J. Greve, Cytometry **36**, 200 (1999).
[31] J.F. Marko and E.D. Siggia, Macromolecules **27**, 981 (1994).
[32] J.F. Marko and E.D. Siggia, Science **265**, 1599 (1994); Phys. Rev. E **52**, 2912 (1995).
[33] T.R. Strick, J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette, Science **271**, 1835 (1996).
[34] T. Strick, V. Croquette, and D. Bensimon, Biophys. J. **74**, 2016 (1998).
[35] B. Fain, J. Rudnick, and S. Ostlund, Phys. Rev. E **55**, 7364 (1997).
[36] D. Moroz and P. Nelson, Proc. Natl. Acad. Sci. U.S.A. **94**, 14418 (1997); Macromolecules **31**, 6333 (1998).
[37] C. Bouchiat and M. Mezard, Phys. Rev. Lett. **80**, 1556 (1998).
[38] A. Sarkar, J.-F. Leger, D. Chatenay, and J.F. Marko, Phys. Rev. E **63**, 051903 (2001).
[39] D. Bensimon, D. Dohmi, and M. Mezard, Europhys. Lett. **42**, 97 (1998).
[40] P. Nelson, Phys. Rev. Lett. **80**, 5810 (2000).
[41] L.F. Liu and J.C. Wang, Proc. Natl. Acad. Sci. U.S.A. **84**, 7024 (1987).
[42] M.D. Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S.M. Block, Science **282**, 902 (1998).