DNA-Protein Cooperative Binding through Variable-Range Elastic Coupling

Joseph Rudnick and Robijn Bruinsma
Department of Physics, University of California at Los Angeles, Los Angeles, California 90095-1547 USA

ABSTRACT Cooperativity plays an important role in the action of proteins bound to DNA. A simple mechanism for cooperativity, in the form of a tension-mediated interaction between proteins bound to DNA at two different locations, is proposed. These proteins are *not* in direct physical contact. DNA segments intercalating bound proteins are modeled as a worm-like chain, which is free to deform in two dimensions. The tension-controlled protein-protein interaction is the consequence of two effects produced by the protein binding. The first is the introduction of a bend in the host DNA and the second is the modification of the bending modulus of the DNA in the immediate vicinity of the bound protein. The interaction between two bound proteins may be either attractive or repulsive, depending on their relative orientation on the DNA. Applied tension controls both the strength and the range of protein-protein interactions in this model. Properties of the cooperative interaction are discussed, along with experimental implications.

INTRODUCTION

The cooperative binding of proteins to DNA plays a significant role in the regulation of gene expression (Owen-Hughes and Workman, 1994) because it allows a sensitive response to small changes in protein concentration. In particular, it is well known that transcription factor proteins (Lodish et al., 1995) exhibit a significant level of cooperativity (Sun et al., 1997). The structural basis of the cooperativity is not fully understood (Sun et al., 1997), but it is known that long-range cooperativity is possible through loops (Schlief, 1992), formed as the result of association between two DNA-binding proteins. Looping is also believed to play an important role in gene access control. Cooperativity at shorter distances may be related to specific protein-protein interactions or to a generic cooperativity resulting from structural distortions induced by the binding of a protein to DNA (Lilley, 1995; Nelson, 1995). For example, the binding of transcription regulation proteins such as the important TATA-box promoters (TPB) involves amino acid intercalation into the stack of basepairs (Kim et al., 1993; Werner et al., 1996). The result is that kinks are produced in the form of sharp local bending angles in the DNA strand. This deformation may permit a better fit for other DNA-associating proteins, such as the polymerases. Disruptions of the basepair stacking sequence have no effect beyond about half a turn of the double helix (Kim et al., 1993; Werner et al., 1996), so it is expected that this form of cooperativity is restricted to the immediate neighborhood of the primary binding protein. Protein-induced deformations of the DNA strand are not restricted to transcription factors. DNA may *wrap* itself once or more around a protein, as happens in the case of complexation of DNA with nucleosomes, the gyrase enzyme, or bacterial RNA. Interestingly, nucleosome-binding appears to be cooperative with transcription factors (Owen-Hughes and Workman, 1994). DNA-deforming proteins will be referred to below as "architectural" proteins.

The aim of this article is to demonstrate the possibility of a variable-range form of cooperative DNA binding of architectural proteins with a range and strength that is regulated by the *tension* along the DNA strand. The cooperative interaction between proteins that are *not* in physical contact is mediated by the deformation of the intervening DNA strand. In the absence of tension, the proposed mechanism is absent. Our demonstration of the possibility of tension-controlled cooperativity is based on an analysis of the "worm-like chain" (WLC) model of DNA elasticity (Hagerman, 1981; Kam et al., 1981), which has been used in studies of single protein binding to DNA (Marko and Siggia, 1997). The WLC model is characterized by a single parameter, a length-scale ξ_p, known as the *persistence length.* It is the distance over which a (tensionless) WLC maintains orientational order in the presence of thermal fluctuations. That is to say, the autocorrelation between orientational order at two different locations of the chain falls off with distance, ℓ, as $e^{-\ell/\xi_p}$. Fitting the results of micromechanical in vitro studies of protein-free DNA chains to the predictions of the WLC model yields good results for a persistence length of $\sim$50 nm (Bustamante et al., 1994; Bensimon et al., 1994), although longer persistence lengths have been reported by different methods (Bednar et al., 1995). Further studies of the elastic properties of DNA in in vitro conditions can be found in Strick et al., 1996; Cluzel et al., 1996; Larson et al., 1997.

We employ the WLC model only to evaluate binding cooperativity due to deformations produced by architectural proteins in sections of the DNA strand that are not in the immediate neighborhood of the binding proteins themselves. The WLC model will not apply reliably when the

Received for publication 7 July 1998 and in final form 8 January 1999.

Address reprint requests to Dr. Joseph Rudnick, Physics Department, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095-1547. Tel.: 310-825-8535; Fax: 310-206-5668; E-mail: jrudnick@physics.ucla.edu.

0006-3495/99/04/1725/09 $2.00

two proteins are so close together that details of basepair action (i.e., roll, slide, and twist) play an important role in the mediation of their interaction. The primary binding of the protein itself is characterized by two phenomenological parameters: the single protein, zero-tension specific binding free energy $E^{(1)}(s)$, and the DNA bending angle α; see Fig. 1. The specific binding energy $E^{(1)}(s)$ is a sequence-sensitive quantity that depends on the location of the protein along the chain; the index s refers to the position of the protein along the DNA strand. It is usually in the range of 10–30 k_BT, with k_B Boltzmann's constant and T the temperature in Kelvin. At room temperature $k_BT \simeq 0.6$ kcal/mol. By comparison, the individual ionic bonds between basepairs in DNA are in the range 2–3 kcal/mol, and typical covalent bounds are the order of 60–120 kcal/mol. The protein-DNA complex will be assumed to be rigid for the tension levels envisioned, so the bending angle α is independent of tension. The values of $E^{(1)}(s)$ and α must be obtained either experimentally or by detailed structural modeling of DNA-protein interactions.

We will focus on the results of analytical and numerical studies that utilize the WLC model to compute the deformation energy of a DNA chain under tension with two identical architectural proteins attached and separated by a distance l. The two proteins are assumed to induce a bend into the DNA strand without any twisting. We find that there is, in general, both an enthalpic and an entropic contribution to the binding cooperativity proposed here (Wang and Giaever, 1988). The enthalpic contribution to the energy is associated with the configuration of DNA with attached proteins that minimizes the total classical energy of the system, while the entropic contribution is generated by fluctuations about that minimum energy configuration. We find that the enthalpic energy of cooperativity results from the deformation, here the kinking, of the DNA due to the presence of an attached protein, while the entropic contribution to the cooperative energy arises from the modification of the DNA's bending modulus induced by the proteins.

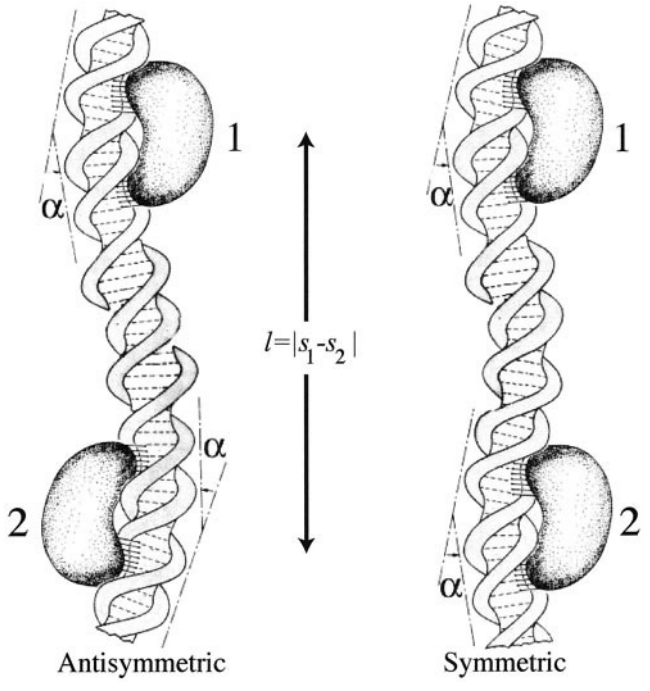

FIGURE 1 Strand of DNA containing two proteins. The proteins induce "kinking" on the DNA strand. Illustrated are the case of antisymmetric and symmetric configurations of the two bound proteins. The distance between proteins is expressed in terms of the locations, s_1 and s_2, of each of them on the strand.

The enthalpic cooperative correction to the binding energy $E^{(2)}$ between the two proteins (denoted by 1 and 2) that are separated by a distance $l = |s_1 - s_2|$ assumed large compared to the characteristic dimensions of the protein is given by

$$E^{(2)}_{\text{enth}} \simeq E^{(1)}_1(s_1) + E^{(1)}_2(s_2) - 2\alpha^2 k_B T\left(\frac{\xi_p}{\xi(F)}\right)(1 \mp e^{-|s_1-s_2|/\xi(F)}) \quad (1)$$

This formula, derived in Appendix A, holds when the bending angle α is small compared to $\pi/2$. The case of large bending angle is discussed later. The plus sign in Eq. 1 refers to the "symmetric" case with the two proteins bound on the same side of DNA while the minus sign refers to the "antisymmetric" case with the two proteins bound on opposite sides of the DNA (see Fig. 1). The chain-tension F in Eq. 1 is variable but assumed to be in the range of 10^{-2} to 10 pN. The tension-dependent length-scale $\xi(F)$ in Eq. 1 is of key importance; it sets the *range* of the binding cooperativity. This quantity is defined by:

$$\xi(F) = \sqrt{\frac{\xi_p k_B T}{F}} \quad (2)$$

For a 1-pN tension, $\xi(F)$ is ~7 nm, while for a 10^{-2} pN force it is ~70 nm. However, Eq. 1 is only valid as long as the "tension-length" $\xi(F)$ is less than the persistence length ξ_p. For larger tensions $\xi(F)$ is small compared to the persistence length $\xi_p \approx 50$ nm. Note that all parameters in Eq. 1 can, in principle, be determined experimentally.

According to Eq. 1, there is no bending-induced cooperativity between two proteins separated by a distance large compared to the tension length $\xi(F)$. In that limit, the only effect of tension is to reduce the single protein binding energy by an amount per protein $\Delta E_{\text{enth}} = \alpha^2\sqrt{k_B T \xi_p F}$ (using Eq. 2 to eliminate $\xi(F)$ in Eq. 1). This tension-induced reduction of the protein binding energy is discussed (Marko and Siggia, 1997) for the case of single-protein binding to DNA. For a 1-pN tension, the binding energy reduction is significant: $\sim 7\alpha^2 k_B T$. With increasing tension, the protein will be released from the DNA chain when ΔE_{enth} starts to approach the protein-binding free energy $E^{(1)}$. [If the DNA-protein interaction involves a number of turns of the DNA around the protein, then we must add the quantity ΔLF to ΔE_{enth}, with ΔL the excess DNA length wound around the protein. For low tensions F, this correc-

TABLE 1 Enthalpic energy of interaction

F (pN)	$\xi(F)$ (ξ_p = 500 Å)	$V_{enth}(l$ = 0 Å) (kcal/mol)	$V_{enth}(l$ = 50 Å) (kcal/mol)
0.1	221 Å	0.82	0.66
1.0	70 Å	2.6	1.27
10	22 Å	8.2	0.85

Table of the enthalpic interaction strength, as given by the last term on the right-hand side of Eq. 1, for physically reasonable values of the temperature, the persistence length ξ_p, and the tension, F. For the dependence of the binding cooperativity range, $\xi(F)$ on tension and other parameters, see Eq. 2.

tion is small compared to ΔE_{enth}, but for a 1-pN tension, it is comparable.]

A summary of expected values for the cooperative interaction between two bound proteins is displayed in Table 1. It is assumed that the bending angle that each enforces is 45°. Other quantities are appropriate to DNA at room temperature. These values should be compared to, for instance, the cooperative interaction energies between repressor molecules bound to adjacent sites of DNA, which are the order of 1.3 kcal/mol (Hyde and Spicer, 1995).

When the spacing between the two proteins is reduced to within a distance of order of the tension-length $\xi(F)$, then the binding energy increases exponentially for the antisymmetric arrangement while it decreases exponentially for the symmetric arrangement. We can interpret the last term in Eq. 1 as an *effective potential energy of interaction*, $V_{enth}(l)$, between two proteins given by:

$$V_{enth}(l) \cong \mp 2\alpha^2 k_B T\left(\frac{\xi_p}{\xi(F)}\right)e^{-l/\xi(F)} \tag{3}$$

with $l = |s_1 - s_2|$ the interprotein spacing. The minus sign in Eq. 3 is for the antisymmetric case. In Appendix A we compute $V_{enth}(l)$ for values of the bending angle α that range up to $\pi/2$. The result is shown in Fig. 2. Note that both the vertical and horizontal axes are dimensionless. The energy has been expressed in units of $2k_BT(\xi_p/\xi(F))$, the energy scale of the tension-induced binding energy reduction ΔE_{enth} (see Eqs. 1 and 3), while the distance is expressed in units of the tension length, $\xi(F)$. It follows from Fig. 2 that an increase in tension *reduces* the range of the interaction, as expected from Eq. 2, while it *increases* the strength of the cooperativity. If the spacing l between the proteins is small compared to the tension length $\xi(F)$, then the effective potential $V(0)$ cancels the ΔE_{enth} term in Eq. 1. The enthalpic energy gain obtained by bringing two proteins together along the chain from a large separation is equal to twice ΔE_{enth}.

For the symmetric case, the effective potential energy is repulsive. The corresponding energy plots are shown in Fig. 3. The enthalpic deformational energy now increases as the two proteins approach each other. In the limit of small bending angle α, two adjacent bending proteins with α in the symmetric conformation have the same tension-induced binding energy reduction as a single protein with a double bending angle of 2α. The energy scale for the cooperativity is thus in general set by ΔE_{enth}. The effective interaction potential is, then, always less than the single protein binding energy, since proteins are expected to unbind when ΔE_{enth} spacing approaches $E^{(1)}(s)$.

Although it would appear as if the enthalpic cooperativity depends on temperature within the WLC model (see Eqs. 1 and 3), this is not the case: ξ_P is inversely proportional to k_BT (see Eq. A.8) so neither ΔE_{enth} nor $\xi(F)$ depend on k_BT. There *is*, however, a purely entropic contribution to the cooperativity that is explicitly dependent on k_BT. In Appendix B, we obtain the following expression for the entropic correction to the cooperativity:

$$\Delta E^{(2)}_{ent} = k_BT\left\{\frac{d}{\xi(F)} - \ln\left(1 + \frac{d}{\xi(F)} + \frac{1}{4}\left(\frac{d}{\xi(F)}\right)^2\left[1 - e^{-(2l/\xi(F))}\right]\right)\right\} \tag{4}$$

for two proteins of length d separated by a distance $l = |s_1 - s_2|$. As noted previously, this energy results from the action of proteins in modifying the bending modulus of the

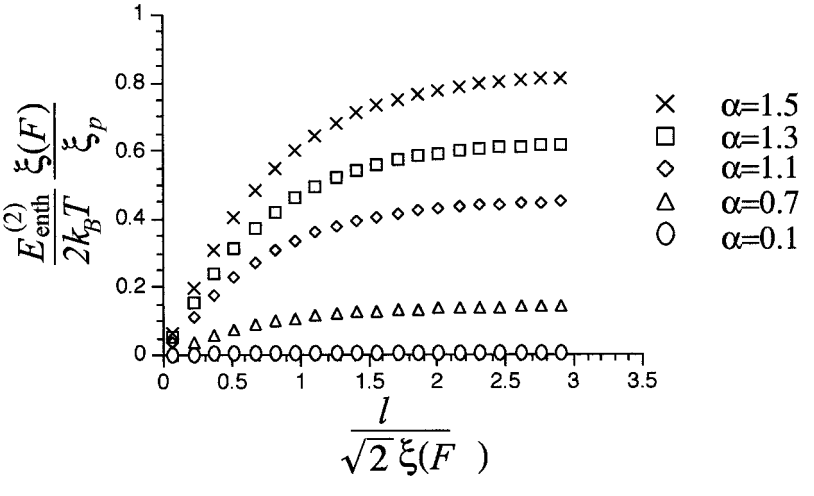

FIGURE 2 The attractive interaction for an antisymmetrically oriented pair of bound proteins. Plotted is the total energy of the configuration in units of k_BT, multiplied by the ratio $\xi(F)/\xi_p$, where $\xi(F)$ is the tension-dependent length scale defined in Eq. 2, and ξ_p is the persistence length of the DNA strand. Curves are displayed for various values of the "kink angle," α, as shown in Fig. 1.

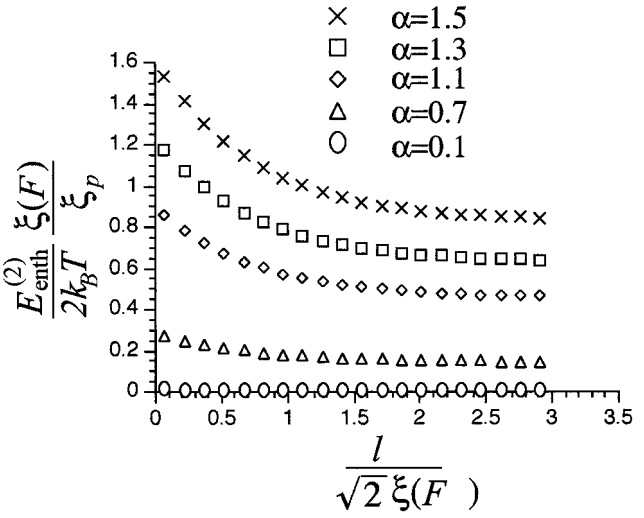

FIGURE 3 The repulsive interaction between two proteins bound in a symmetric configuration. Plotted is the total energy of the configuration in units of k_BT, multiplied by the ratio $\xi(F)/\xi_p$, where $\xi(F)$ is the tension-dependent length scale defined in Eq. 2, and ξ_p is the persistence length of the DNA strand. Curves are displayed for various values of the "kink angle," α, as shown in Fig. 1.

DNA strand to which they are attached. Equation 4 follows from the assumption that the section of DNA to which a protein is attached has become infinitely stiff. The entropic contribution does not depend on the bending angle α and is the same for the symmetric and antisymmetric configurations. In the limit of protein separations that are large compared to the tension length $\xi(F)$, the entropic contribution $\Delta E_{entr} = (1/2)\Delta E^{(2)}_{entr}(\infty)$ to the single-protein binding energy is:

$$\Delta E_{entr} = k_BT\left\{\frac{d}{2\xi(F)} - \ln\left(1 + \frac{d}{2\xi(F)}\right)\right\} \qquad (5)$$

This is, again, a negative quantity: the local constraints imposed on the DNA chain by the two binding proteins lower the entropy of the complex as compared to a free chain, and hence reduce the binding energy. The difference can, again, be interpreted as an effective entropic potential energy of interaction, which is now entropic. It is given by

$$V_{entr}(l) = k_BT\ln\left(1 - \frac{(d/\xi(F))^2 e^{-(2l/\xi(F))}}{4(1 + (d/2\xi(F))^2}\right) \qquad (6)$$

This effective entropic potential energy is always attractive. In Fig. 4 we show the entropic potential energy for two proteins of size d equal to 20 Å and for a 1-pN applied tension ($\xi(F = 1\text{ pN}) = 70$ Å).

For larger bending angles, the entropic interaction is both weaker and shorter in range than the enthalpic interaction. However, because this interaction does not depend on the magnitude of the bending angle, it dominates for zero bending angles, or bending angles that are very small. An interesting special case concerns the entropic interaction between two long strings of binding proteins. If we model a

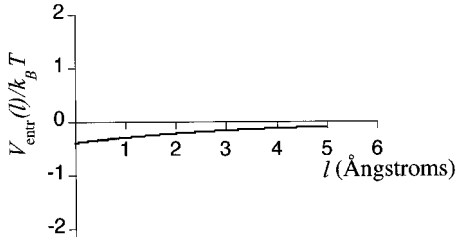

FIGURE 4 The entropic interaction potential between two identical bound proteins resulting from the effect of each of them on the bending modulus of the strand of DNA to which they are attached. The interaction is plotted in units of k_BT under the assumption that the region of affected DNA is 70 Å long, and that the applied tension is 1 pN.

polymerized string of proteins bound to DNA as a rigid section of size d, with d assumed large compared to the tension length, and with zero total bending angle, then two such strings are expected to have an effective entropic interaction potential given by:

$$\begin{array}{lcl} V_{entr}(l) & \cong & k_BT\ln(1 - e^{-(2l/\xi(F))}) \\ d/\xi(F) & \rightarrow & \infty \end{array} \qquad (7)$$

[Formally, there was a divergence in Eq. 7 for small spacings l, but Eq. 7 should, of course, not be expected to retain validity when l approaches a basepair spacing.] The free energy of two long rigid strings separated by a small gap is lowered by an amount of order k_BT if the intervening gap is filled in either by shifting one of the two strings to close the gap or by adding additional binding proteins inside the gap. This effect provide us with a curious *entropic* stabilization of mechanism of polymerization of proteins along DNA strands.

When we increase the bending angle beyond $\pi/2$, new physical effects appear. As shown in Figs. 5 and 6, there are in general two possible configurations for a symmetric two-protein/DNA complex. Up to now, it has been tacitly assumed that the "S" or "stretched" configuration was the appropriate one (as shown in Fig. 1) and indeed the S

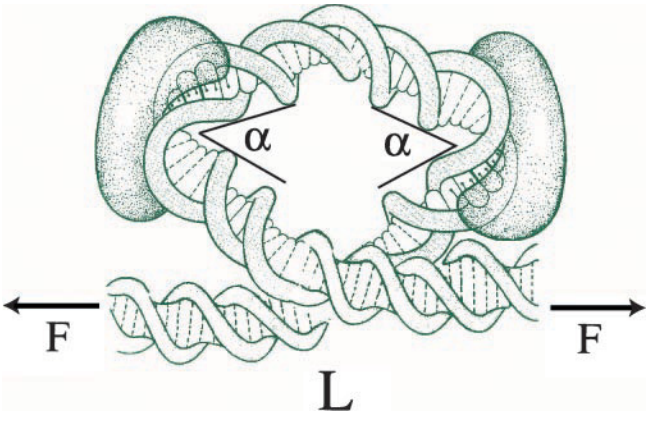

FIGURE 5 The "loop" or L configuration of two symmetrically bound proteins, when the bend angle enforced by a bound protein is $> \pi/2$. This is the preferred configuration when the applied tension is small.

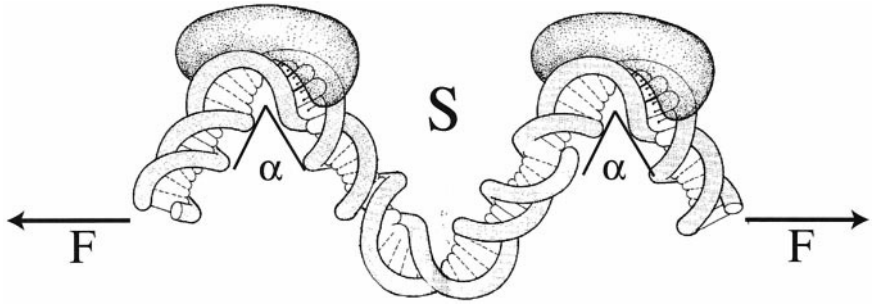

FIGURE 6 The stretched, or S, configuration of two symmetrically bound proteins when the bend angle exceeds $\pi/2$. This is the preferred configuration at high levels of applied tension.

configuration has the lower free energy for bending angles $\alpha < \pi/2$. However, when the bending angle exceeds $\pi/2$ this is no longer the case. For low tensions, the "L," or "looped" configuration has in fact the lower elastic free energy, while for higher tensions, the S configuration is more stable. The two regimes are separated by a mathematical singularity that has the character of a first-order phase transition. In Fig. 7 we show the extension, X, of the two-kink configuration for which the kink angle is greater than $\pi/2$, as a function of tension, F. There is a transition from the L to the S configuration with increasing tension visible as a discontinuity of the extension at the transition point. There is no transition in the antisymmetric case.

In summary, we have shown that, within the confines of the WLC model, tension can trigger cooperative binding for antisymmetrically arranged architectural proteins. The binding strength has both an enthalpic and an entropic contribution, and it has an appreciable magnitude for tensions of the order of 1-pN or higher. For larger bending angles, we find two competing configurations connected by a tension-induced phase-transition. Both the enthalpic and entropic contributions to cooperativity vanish in the limit of zero tension (see Eqs. 1 and 6). This last result is certainly not self-evident. The decrease in chain entropy imposed by two rigid sections can, for instance, be expected to depend on the spacing, even for zero tension. Interestingly, a study of the interaction between two stiff inclusions inside a *two dimensional* surface (such as a membrane) reports (Bruinsma et al., 1994) that in this case there is a long-range zero-tension interaction with both entropic and enthalpic contributions, both dropping off as the inverse fourth power of the spacing between the inclusions. The disappearance of tension-induced cooperativity at $F = 0$ thus must be related to the fact that we are dealing with a one-dimensional geometry.

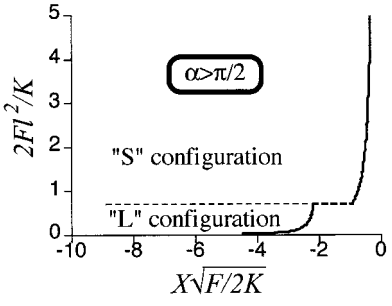

FIGURE 7 The overall extension of a very long strand of DNA containing two identical symmetrically bound proteins, the bend angle, α, of each of which is 2.1. The extension, X, is relative to the fully extended DNA strand, measured in units of $\sqrt{2K/F}$, where F is the applied tension and K is the bending modulus. The vertical axis is the applied tension, F, in units of $K/2l^2$, where l is the distance along the DNA backbone between proteins along the DNA strand. See Appendix A for a full discussion of the parameters utilized.

Experimental in vitro tests of the proposed mechanism can be performed by preparing a bundle of DNA strands, each strand containing bacterial gene operator sequences periodically spaced by a distance of l basepairs. The associated repressor protein (such as the lac repressor) binding specifically to the operator sites would induce local kinks at the operator sites. According to the model, the logarithm of the equilibrium repressor-operator binding constant K_{RO} contains a contribution that depends on the operator spacing l and the tension F of the DNA bundle according to Eq. 1.

A study of the kinetics of cooperative protein-DNA association can also be a testing ground. For instance, since the TATA box binding protein TBP is known to produce a large bending angle, the one-dimensional diffusion along the DNA of other proteins required for the RNA polymerase initiation complex, such as TFIIE, H, and J, that are nonspecifically bound to DNA, will be speeded up by bending-induced cooperativity. The reason is that the effective potential $\Delta V(l)$ turns the random one-dimensional diffusion into a directed process. It should be noted here that the weaker nonspecific binding of DNA associating proteins will not deform the DNA strand as much as specific binding. However, studies of the dependence of nonspecifically bound 434 repressors on the DNA flexibility indicate that nonspecific bonding also involves distortions of the local DNA structure, so there still ought to be a tension-controlled interaction between specific and nonspecifically bound proteins (Hogan and Austin, 1987). We thus predict that (modest) tension will actually increase the formation rate of the RNA polymerase initiation complex. At high tension levels, the formation rate will decrease with tension for reasons discussed earlier.

Another possible area where the present theory could be applied is histone-DNA interactions. According to our model calculations, a collection of nonspecifically bound proteins ought to adopt an antisymmetric zig-zag configuration under tension. The binding of histones to DNA is reported to produce a zig-zag nucleosome structure consis-

tent with our calculations (Thoma et al., 1979). Under tension, the zig-zag structure should thus be stabilized. It should be kept in mind, though, that the intervening linker histones may well affect the competition between different configurations (Thoma et al., 1979).

An important question for the relevance of the work presented here is whether DNA is under tension under in vivo conditions. DNA strands suspended in good solvent are in fact not under tension. We believe that this is an exceptional case. A DNA strand whose ends are fixed is, even for low extensions, typically under a tension in the range of 0.01 pN due to thermal fluctuations, as shown by the micromechanical studies (Bustamante et al., 1994; Bensimon et al., 1994). For extensions closer to one, the tension can be much larger. If a DNA strand is subject to the activity of force-transducing proteins—for instance during mitosis, RNA transcription, or homologous recombination—then much higher tensions can be generated. It is known that a single motor protein is able to generate a force of 10 pN or more (Yin et al., 1995). Finally, architectural proteins themselves generate tension if they attach to a DNA strand with fixed ends. When an architectural protein attaches, DNA material is required to accommodate the deformation of the DNA chain near the protein. A simple calculation shows that the self-induced tension of a DNA strand of length L whose ends are held fixed a distance X apart and which contains a line density ρ of bending proteins is given by:

$$F \cong \frac{4k_B T\langle\alpha^2\rangle\xi_p}{\left(1-\frac{X}{L}\right)^2}\rho^2 \qquad (8)$$

with $\langle\alpha^2\rangle$ the average of the square of the bending angle. If the line density is of the order of one protein/100 Å and if $\langle\alpha^2\rangle$ is of the order one, then this self-generated tension is of the order 1-pN (it should be possible to verify Eq. 8 in micromechanical measurements). It thus seems reasonable to assume that DNA is under tension for in vivo conditions.

Our results were derived with DNA-protein binding in mind, but the general aspects of our conclusions are related to work in other areas. We can consider the predicted aggregation of proteins under tension as a form of stress-induced decomposition. Stress-induced phase-separation of multicomponent systems is actually a classic phenomenon in solid-state materials for the case that different constituents have different elastic moduli (Cahn, 1961), just as envisioned in the present case.

We conclude by noting that a number of technical objections can be raised against the method used in our study. It is not really reasonable to assume that DNA twist plays no role, in view of the helical nature of DNA, and that we can be allowed to restrict ourselves to a purely two-dimensional arrangement. However, it is our current belief that the introduction of the twist degree of freedom introduces qualitative, but not quantitative, modifications to the results reported here. The assumption of internal structural rigidity of proteins also is questionable. For instance, it is well known that many enzymes can undergo stress-induced structural changes. Enzymes bound to DNA also may change their structure under stress. In addition, it is true that the thermodynamic ensemble in which the tension is kept fixed is not identical to the ensemble in which the length of the DNA segment is held constant, and the condition of fixed segment length is a more reasonable assumption in many in vivo situations. However, the calculations reported here were for two proteins attached to an asymptotically long segment, and in that limit, the two ensembles yield identical results. Finally, it must be kept in mind that the WLC model neglects the possible influences of intervening structures, such as other bound proteins, on the tension-induced interaction between two bound proteins, and that looping and other manifestations of DNA self-interaction are ignored. Despite all these caveats, we feel that the basic result—stress-induced aggregation of DNA-bound proteins—is robust and should remain present in more realistic models.

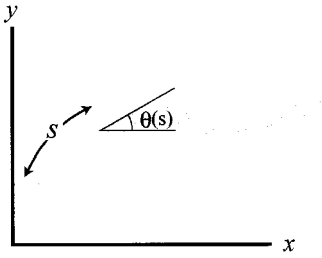

FIGURE 8 The parameters in Eq. A.1. The quantity $\theta(s)$ is the angle between the flexible strand and the horizontal axis, while s is arclength.

APPENDIX A

We model the DNA strand as a rod that is free to move in two dimensions. The shape of the rod is parametrized in terms of an angular variable θ, which will vary with distance, s, along the rod. This parametrization is pictured in Fig. 8. The energy of a configuration of the system is given by the following expression.

$$H[\theta(s)] = \int ds\left\{\frac{K}{2}\left(\frac{d\theta}{ds}\right)^2 - F\cos\theta\right\} \qquad (A.1)$$

The rigidity against bending is parametrized in terms of a stiffness parameter K, while F is the tension applied to the rod. The quantities referred to in the text are related to K and F by $\xi(F) = \sqrt{K/F}$ and $\xi_p = K/k_B T$. In the "classical" limit, which applies at very low temperatures, is determined by

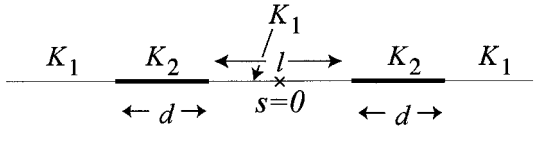

FIGURE 9 The regions of varying stiffness in a strand containing two bound proteins that alter the bending modulus where they attach to a strand of DNA. The values that the bending modulus, K, takes in the various regions are indicated. See Eq. B.1 and surrounding text in Appendix B.

the extremum equation

$$\frac{\delta H}{\delta\theta(s)} = -K\frac{d^2\theta}{ds^2} + F\sin\theta(s) = 0 \tag{A.2}$$

This equation is solved by quadratures by noting that it is identical to the first order differential equation

$$\frac{K}{2}\left(\frac{d\theta(s)}{ds}\right)^2 + F\cos\theta(s) = \kappa \tag{A.3}$$

with κ a constant. From Eq. A.3 we immediately obtain

$$\frac{d\theta}{\sqrt{\frac{2F}{K}(\Lambda - \cos\theta)}} = \pm ds \tag{A.4}$$

with Λ another constant. Integration of A.4 yields an implicit form for $\theta(s)$, in which s is represented as a function of θ. The specific function is a combination of elliptic integrals. There is no evident way to invert the expression thus derived to obtain θ as an explicit function of s.

However, when the bending of the DNA is small, so that θ is small, the cosine function in A.1 can be expanded. The zeroth order term is a constant, "background" energy. The energy of the system, as it depends on the configuration of the DNA, is given by

$$H[\theta(s)] = \int ds\left\{\frac{K}{2}\left(\frac{d\theta(s)}{ds}\right)^2 + \frac{F}{2}\theta(s)^2\right\} \tag{A.5}$$

The statistical mechanics of the system controlled by this energy can be analyzed in complete detail. Here, the extremum equation is

$$-K\frac{d^2\theta(s)}{ds^2} + F\theta(s) = 0 \tag{A.6}$$

the solution of which is

$$\theta(s) = Ae^{s/\xi(F)} + Be^{-s/\xi(F)} \tag{A.7}$$

where

$$\xi(F) = \sqrt{KF} \equiv \sqrt{\xi_p k_B T F}, \tag{A.8}$$

and A and B are constants.

As an example of the use of the small angle formulas A.5–A.7 we calculate the free energy cost of the presence of two proteins that kink the DNA to which they are attached. These insertions enforce a discontinuity in $d\theta/ds$ at the location of each kink. We assume that the magnitude of the discontinuity is α_- at the first kink and α_+ at the second. To simplify the analysis, we place the first kink at $s = -l/2$ and the second kink at $s = l/2$. The calculation of the classical solution subject to these constraints is relatively straightforward. One finds

$$\theta(s) = \begin{cases} \theta_< \exp\left(\sqrt{\frac{F}{K}}\left(s + \frac{l}{2}\right)\right) & s < -\frac{l}{2} \\ -\theta_-\dfrac{\sinh\sqrt{\frac{F}{K}}\left(s - \frac{l}{2}\right)}{\sinh\sqrt{\frac{F}{K}}l} + \theta_+\dfrac{\sinh\sqrt{\frac{F}{K}}\left(s + \frac{l}{2}\right)}{\sinh\sqrt{\frac{F}{K}}l} & -\frac{l}{2} < s < \frac{l}{2} \\ \theta_> \exp\left(-\sqrt{\frac{F}{K}}\left(s - \frac{l}{2}\right)\right) & \frac{l}{2} < s \end{cases} \tag{A.9}$$

According to A.9, the angle is equal to $\theta_<$ immediately to the right of $s = -l/2$ and θ_- immediately to the left of $s = -l/2$. By the same token $\theta(l/2 - \epsilon) = \theta_+$ and $\theta(l/2 + \epsilon) = \theta_>$. The kinks give rise to a discontinuity at $s = \pm l/2$, which means that $\theta_< \neq \theta_-$ and $\theta_> \neq \theta_+$. We write

$$\begin{aligned} \theta_< &= \theta_- - \alpha_- \\ \theta_> &= \theta_+ + \alpha_+ \end{aligned} \tag{A.10}$$

Then, the total energy, E, of the bent rod, as given by Eqs. A.5, A.9, and A.10, is given by

$$E = \frac{1}{2}\sqrt{FK}\left\{(\theta_- - \alpha_-)^2 + (\theta_+ + \alpha_+)^2 + \theta_+^2\coth\sqrt{\frac{F}{K}}l + \theta_-^2\coth\sqrt{\frac{F}{K}}l + \frac{2\theta_+\theta_-}{\sinh\sqrt{\frac{F}{K}}l}\right\} \tag{A.11}$$

If we replace θ_+ by $\Sigma + \Delta$ and θ_- by $\Sigma - \Delta$, then Eq. A.11 becomes

$$E(\Sigma, \Delta) = \sqrt{FK}\left\{\Sigma^2\frac{e^{\sqrt{\frac{F}{K}}l} + 1}{\sinh\sqrt{\frac{F}{K}}l} + \Delta^2\frac{e^{\sqrt{\frac{F}{K}}l} - 1}{\sinh\sqrt{\frac{F}{K}}l} + \Sigma(\alpha_+ - \alpha_-) + \Delta(\alpha_+ + \alpha_-) + \alpha_+^2 + \alpha_-^2\right\} \tag{A.12}$$

The partition function, Z, of the system is given by

$$Z = \int\int d\Sigma d\Delta\,\exp(-\beta E(\Sigma, \Delta)) \tag{A.13}$$

where $\beta = 1/k_B T$. This double Gaussian integral is evaluated by completing squares. Taking the log and multiplying by $k_B T$ to obtain the free energy, F, we find

$$F = \frac{\sqrt{FK}}{4}\left\{\alpha_+^2 + \alpha_-^2 + 2\alpha_+\alpha_- e^{-\sqrt{\frac{F}{K}}l}\right\} - \frac{k_B T}{2}\ln\left(1 - e^{-2\sqrt{\frac{F}{K}}l}\right) \tag{A.14}$$

The first term on the right-hand side of the expression for the free energy in Eq. A.14 is the mean field approximation to the free energy cost of two kinks in the rod. The second term is a partial contribution to the free energy due to fluctuations, in particular fluctuations in the angles θ_+ and θ_-. To complete the evaluation of the free energy associated with a pair of kinks, we must now average over all other fluctuations in the rod. To do this, we expand $\theta(s)$ in normal sinousoidal modes, subject to the condition that those modes do not alter the values of θ immediately to the right or the left of the locations of the kinks. The contribution to the free energy that depends on the distance, l, between the kinks arises from the fluctuations in that region. This fluctuation sum reduces to

$$k_B T\left\{\sum_{n=1}^{\infty}\ln\left[\left(\frac{n\pi}{l}\right)^2 + \frac{F}{K}\right] - \frac{1}{2}\ln\frac{F}{K}\right\} = \frac{k_B T}{2}\ln\left(1 - e^{-\sqrt{(F/K)}l}\right) \tag{A.15}$$

In arriving at the expression on the right-hand side of Eq. A.15, an infinite contribution to the sum over logarithms was subtracted. This "regularization" of the sum is consistent with standard field theoretical approaches to the evaluation of the free energy of systems such as the one under consideration here. The right-hand side of A.15 exactly cancels the fluctuation contribution to A.14. Thus, the free energy of the two-kink system consists entirely of the mean-field contribution.

APPENDIX B

In the limit that the kinks occupy an infinitesimal portion of the DNA, the free energy cost of a pair of them is, as noted in the appendix above, given entirely by mean-field theory. There is a mechanism leading to fluctuation-induced interaction energy, and that is the modification of the bending modulus by the proteins that cause the kinks. This modification can be modeled as follows. In the small-angle approximation, one replaces the energy expression in A.5 by

$$H[\theta(s)] = \int ds \left\{ \frac{K_{1,2}}{2} \left(\frac{d\theta(s)}{ds} \right)^2 + \frac{F}{2} \theta(s)^2 \right\} \qquad \text{(B.1)}$$

where the symbol $K_{1,2}$ stands for the two possible values, K_1 and K_2, that the bending modulus can now take on in the various regions. The subscript 2 applies in the regions that are in the immediate vicinity of the proteins, while the subscript 1 is appropriate everywhere else. This means that K_1 is the unsubscripted K in previous expressions. The effects of modifications of the bending modulus can be calculated separately from the consequences of the bending of the DNA by proteins. This means that we can ignore any discontinuities or alterations of the equilibrium value angle $\theta(s)$ that result from the presence of proteins.

The regions in which the bending modulus takes on its two possible values are indicated in Fig. 8. The bending modulus is equal to K_2 in the two heavily drawn regions, while it is equal to K_1 everywhere else. Matching conditions at the boundary between regions are that $\theta(s)$ and $K_i d\theta/ds$ are continuous.

In order to evaluate the contribution of fluctuations, it is necessary to determine the eigenfunctions of the Hamiltonian B.1. One searches for solutions to the equation

$$-K_{1,2} \frac{d^2\theta(s)}{ds^2} + F\theta(s) = \lambda\theta(s) \qquad \text{(B.2)}$$

There will be two sorts of solution: even and odd.

Even solutions

These solutions have the form

$$\theta(s) = \begin{cases} \cosh q_1 s & 0 < s < l/2 \\ A \cosh q_2(s - l/2) & \\ \quad + B \sinh q_2(s - l/2) & l/2 < s < l/2 + d \\ C \cosh q_1(s - l/2 - d) & \\ \quad + D \sinh q_1(s - l/2 - d) & l/2 + d < s \end{cases} \qquad \text{(B.3)}$$

where $\theta_e(-s) = \theta_e(s)$.

Here

$$q_1 = \sqrt{(F - \lambda)/K_1} \qquad \text{(B.4)}$$

and

$$q_2 = \sqrt{(F - \lambda)/K_2} = q_1 \sqrt{\frac{K_1}{K_2}} \qquad \text{(B.5)}$$

Making use of the boundary conditions, we find

$$A = \cosh q_1 l/2 \qquad \text{(B.6)}$$

$$B = \sqrt{\frac{K_1}{K_2}} \sinh q_1 l/2 \qquad \text{(B.7)}$$

$$C = \cosh q_1 l/2 \cosh q_2 d + \sqrt{\frac{K_1}{K_2}} \sinh q_1 l/2 \sinh q_2 d \qquad \text{(B.8)}$$

$$D = \sqrt{\frac{K_2}{K_1}} \cosh q_1 l/2 \sinh q_2 d + \sinh q_1 l/2 \cosh q_2 d \qquad \text{(B.9)}$$

Odd solutions

Here, the solutions are

$$\theta(s) = \begin{cases} \sinh q_1 s & 0 < s < l/2 \\ A \cosh q_2(s - l/2) & \\ \quad + B \sinh q_2(s - l/2) & l/2 < s < l/2 + d \\ C \cosh q_1(s - l/2 - d) & \\ \quad + D \sinh q_1(s - l/2 - d) & l/2 + d < s \end{cases} \qquad \text{(B.10)}$$

where $\theta_o(-s) = -\theta_o(s)$, and, in this case, the coefficients are given by

$$A = \sinh q_1 l/2 \qquad \text{(B.11)}$$

$$B = \sqrt{\frac{K_1}{K_2}} \cosh q_1 l/2 \qquad \text{(B.12)}$$

$$C = \sinh q_1 l/2 \cosh q_2 d + \sqrt{\frac{K_1}{K_2}} \cosh q_1 l/2 \sinh q_2 d \qquad \text{(B.13)}$$

$$D = \sqrt{\frac{K_2}{K_1}} \sinh q_1 l/2 \sinh q_2 d + \cosh q_1 l/2 \cosh q_2 d \qquad \text{(B.14)}$$

The ultimate goal of the calculation is the so-called Fredholm determinant of the differential operator on the right-hand side of Eq. B.2. This determinant is readily evaluated with the use of a by-now well-known trick (Coleman, 1985). The trick yields the following expression for the contribution to the free energy of fluctuations in $\theta(s)$.

$$A = \frac{k_B T}{2} \ln(\theta_e(L)\theta_o(L)) \qquad \text{(B.15)}$$

where the arguments of the logarithm are the even and odd solutions displayed in B.3 and B.10, where λ has been set equal to 0. There are, in addition, normalization factors. The factor that multiplies the even solution sets the magnitude of that solution equal to one at $s = 0$. The factor multiplying the odd solution sets the slope of that solution equal to unity at the origin. The argument L is the total length of the segment to which the proteins are attached. In the following development, the limit $L \to \infty$ is taken, and the contribution to the free energy relevant to the interaction between the two attached proteins is extracted.

In the limit of asymptotically large L the two factors in the argument of

the logarithm in Eq. B.15 are

$$e^{-q_1 d}\left\{\cosh q_2 d + \sinh q_2 d \cdot\left[\sqrt{\frac{K_1}{K_2}}\frac{1+e^{-q_1 l}}{2} + \sqrt{\frac{K_2}{K_1}}\frac{1-e^{-q_1 l}}{2}\right]\right\} \quad \text{(B.16)}$$

and

$$e^{-q_1 d}\left\{\cosh q_2 d + \sinh q_2 d \cdot\left[\sqrt{\frac{K_1}{K_2}}\frac{1-e^{-q_1 l}}{2} + \sqrt{\frac{K_2}{K_1}}\frac{1+e^{-q_1 l}}{2}\right]\right\} \quad \text{(B.17)}$$

After some algebra we find for the dependence of the free energy of the system on the parameters l, d, and the K_i values.

$$A = \frac{k_B T}{2}\ln(f) \quad \text{(B.18)}$$

where

$$f = e^{2(q_2-q_1)d} + \frac{e^{-2q_1 d}\sinh 2q_2 d}{2}\frac{(\sqrt{K_1}-\sqrt{K_2})^2}{\sqrt{K_1 K_2}} + \frac{e^{-2q_1 d}\sinh^2 q_2 d}{4}\frac{(K_1-K_2)^2}{K_1 K_2}[1-e^{-2q_1 l}] \quad \text{(B.19)}$$

Note that the dependence of the argument of the logarithm on the separation, l, between the two regions of differing stiffness constant is as $A - Be^{-2q_1 l}$. This implies an always attractive interaction.

If we now make the assumption that the bending modulus, K_2, in the immediate vicinity of the proteins is infinite, then making the connections summarized in Eq. A.8, we obtain the expression displayed in Eq. 4 for the entropic energy of interaction between attached proteins.

The authors thank C. Bustamante, D. Chatenay, and E. Siggia for helpful conversations. We are especially grateful to P. O'Lague for useful comments.

This work was supported in part by National Science Foundation Grant DMR-9708646 (to R. B.).

REFERENCES

Bednar, J., P. Furrer, V. Katrich, A. Z. Stasiak, J. Dubochet, and A. Stasiak. 1995. Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.* 254:579–594.

Bensimon, A. S., A. Chiffaudel, V. Croquette, F. Heslot, and D. Bensimon. 1994. Alignment and sensitive detection of DNA by a moving interface. *Science.* 265:2096–2098.

Bruinsma, R., M. Goulian, and P. Pincus. 1994. Self-assembly of membrane junctions. *Biophys. J.* 67:746–750.

Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. *Science.* 265:1599–1600.

Cahn, J. W. 1961. On spinodal decomposition. *Acta Metall.* 9:795–801.

Cluzel, P., A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. Caron. 1996. DNA: an extensible molecule. *Science.* 271:792–794.

Coleman, S. 1985. Aspects of Symmetry. 1985. Cambridge University Press, Cambridge, UK. 340–341.

Hagerman, P. J. 1981. Investigation of the flexibility of DNA using transient electric birefringence. *Biopolymers.* 20:1503–1535.

Hogan, M. E., and R. H. Austin. 1987. Importance of DNA stiffness in protein-DNA binding specificity. *Nature.* 329:263–266.

Hyre, D. E., and L. D. Spicer. 1995. Thermodynamic evaluation of binding interactions in the methionine receptor system of Escherichia coli using isothermal titration. *Biochemistry.* 34:3212–3221.

Kam, Z., N. Borochev, and H. Eisenberg. 1981. Dependence of laser light backscattering of DNA on NaCl concentration. *Biopolymers.* 20: 2671–2690.

Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature.* 365: 520–527.

Larson, R. G., T. T. Perkins, D. E. Smith, and S. Chu. 1997. Hydrodynamics of a DNA molecule in a force field. *Phys. Rev. E.* 55:1794–1797.

Lilley, D. M. J., editor. 1995. DNA-Protein Structural Interactions. IRL Press, Oxford.

Lodish, H., D. Baltimore, A. Berk, S. L. Zipursky, P. Matsudaira, and J. Darnell. 1995. Molecular Cell Biology, 3rd ed. W. H. Freeman, New York.

Marko, J. F., and E. D. Siggia. 1997. Driving proteins off DNA using applied tension. *Biophys. J.* 73:2173–2178.

Nelson, H. C. M. 1995. Structure and function of DNA-binding proteins. *Curr. Opin. Gen. Dev.* 5:180–189.

Owen-Hughes, T., and J. L. Workman. 1994. Experimental analysis of chromatin function in transcription control. *Crit. Rev. Euk. Gene Exp.* 4:403–441.

Schlief, R. 1992. DNA Looping. *Annu. Rev. Biochem.* 61:199–223.

Strick, T. R., J.-F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette. 1996. The elasticity of a single supercoiled DNA molecule. *Science.* 271:1835–1837.

Sun, L. J., B. R. Peterson, and G. L. Verdine. 1997. Dual role of the nuclear factor of activated T cells insert region in DNA recognition and cooperative contacts to activator protein 1. *PNAS.* 94:4919–4924.

Thoma, F., T. Koller, and A. Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* 83:403–427.

Wang, J. C., and G. N. Giaever. 1988. Action at a distance along DNA. *Science.* 240:300–304.

Werner, M. H., A. M. Gronenborn, and G. M. Close. 1996. Intercalation, DNA kinking and the control of transcription. *Science.* 271:778–784.

Yin, H., M. D. Wang, K. Svoboda, A. Landick, S. M. Block, and J. Gelles. 1995. Transcription against an applied force. *Science.* 270:1653–1657.