Statistical-Mechanical Theory of DNA Looping

Yongli Zhang,* Abbye E. McEwen,† Donald M. Crothers,* and Stephen D. Levene†
*Departments of Molecular Biophysics and Biochemistry and Chemistry, Yale University, New Haven, Connecticut; and †Institute of Biomedical Sciences and Technology and Department of Molecular and Cell Biology, University of Texas at Dallas, Richardson, Texas

ABSTRACT The lack of a rigorous analytical theory for DNA looping has caused many DNA-loop-mediated phenomena to be interpreted using theories describing the related process of DNA cyclization. However, distinctions in the mechanics of DNA looping versus cyclization can have profound quantitative effects on the thermodynamics of loop closure. We have extended a statistical mechanical theory recently developed for DNA cyclization to model DNA looping, taking into account protein flexibility. Notwithstanding the underlying theoretical similarity, we find that the topological constraint of loop closure leads to the coexistence of multiple classes of loops mediated by the same protein structure. These loop topologies are characterized by dramatic differences in twist and writhe; because of the strong coupling of twist and writhe within a loop, DNA looping can exhibit a complex overall helical dependence in terms of amplitude, phase, and deviations from uniform helical periodicity. Moreover, the DNA-length dependence of optimal looping efficiency depends on protein elasticity, protein geometry, and the presence of intrinsic DNA bends. We derive a rigorous theory of loop formation that connects global mechanical and geometric properties of both DNA and protein and demonstrates the importance of protein flexibility in loop-mediated protein-DNA interactions.

INTRODUCTION

The formation of DNA loops mediated by proteins bound at distant sites along a single molecule is an essential mechanistic aspect of many biological processes including gene regulation, DNA replication, and recombination (for reviews, see Schleif (1) and Matthews (2)). In *Escherichia coli*, DNA looping represses gene expression at the *ara*, *gal*, *lac*, and *deo* operons (3–6) and activates transcription from the *glnALG* operon (7). The size of DNA loops formed in these systems varies between ~100 and 600 basepairs. In eukaryotes, a variety of transcription factors bind to enhancers that are hundreds to several thousands of basepairs away from their promoters and interact with RNA polymerases directly or through mediators to achieve combinatorial gene regulation (8). DNA looping is required to juxtapose two recombination sites in intramolecular site-specific recombination (9–11) and is also employed by a number of restriction endonucleases such as *Sfi*I and *Ngo*MIV, which recognize and cut two copies of well separated cognate sites simultaneously (12–14). The biological importance of DNA loop formation is underscored by the abundance of architectural proteins in the cell, such as HU, IHF, and HMG, which facilitate looping by bending the intervening DNA between protein-recognition sites (15). Moreover, DNA looping has been shown to be subject to regulation through the binding of effector molecules that alter protein conformation or protein-DNA interactions (16).

Two characteristics of DNA looping have been demonstrated by in vitro and in vivo experiments. One is cooperative binding of a protein to its two cognate sites, which can be demonstrated by footprinting methods (17). DNA looping can increase the occupancies of both binding sites; in particular, it can significantly enhance protein association to the lower-affinity site because of the tethering effect of DNA looping. This is a general mechanism by which many transcription factors recruit RNA polymerases in gene regulation. Another hallmark is the helical dependence of loop formation (1,3), which arises because of DNA's limited torsional flexibility and the requirement for correct torsional alignment of the two protein-binding sites. Although many methods have been developed to directly observe DNA looping in vitro, such as scanning-probe (7) and electron microscopy (18), and single-molecule techniques (19), assays based on helical dependence have been the only way to identify DNA looping in vivo. In these experiments, the DNA length between two protein binding sites is varied and the yield of DNA loop formation is monitored, for example, by the repression or activation of a reporter gene (20). Using this helical-twist assay, DNA looping in the *ara* operon was first discovered (3).

Our knowledge about the roles of DNA bending, twist, and their respective energetics in DNA looping has come largely from analyses of DNA cyclization (1,21,22). Shore et al. first showed that circularization efficiencies of DNA fragments, which are quantitatively described by J factors, oscillate with DNA length and therefore torsional phase (23,24). The J factor is defined as the free DNA end concentration whose bimolecular ligation efficiency equals that of the two ends of a cyclizing DNA molecule (25). For short DNA fragments J factors are limited by the significant bending and twisting energies required to form closed

Submitted July 14, 2005, and accepted for publication November 28, 2005.

Address reprint requests to Stephen Levene, Dept. of Molecular and Cell Biology, University of Texas at Dallas, PO Box 830688, Richardson, TX 75083-0688. Tel.: 972-883-2503; Fax: 972-883-2409; E-mail: sdlevene@utdallas.edu.

Yongli Zhang's present address is Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

0006-3495/06/03/1903/10 $2.00

doi: 10.1529/biophysj.105.070490

circles, whereas for long DNA, the chain entropy loss during circularization exceeds the elastic-energy decrease and reduces the *J* factor. Because of this competition between bending and twisting energetics and entropy, there is an optimal DNA length for cyclization (26). Analogous behavior has been expected for DNA looping, especially with respect to the helical dependence discussed above.

Quantitative analyses of DNA looping and cyclization are challenging problems in statistical mechanics and have been largely limited to Monte Carlo or Brownian dynamics simulations (27–31). Analytical solutions are available only for some ideal and special cases. An important contribution in this area is the theory of Shimada and Yamakawa (32), which is based on a homogeneous and continuous elastic rod model of DNA. This theory has been applied extensively to DNA cyclization (23,33) and also to DNA looping (21,22,34). The Shimada-Yamakawa theory makes use of a perturbation approach, in which small configurational fluctuations of a DNA chain around the most probable configuration are accounted for in the evaluation of the partition function.

The elastic-equilibrium conformation is obvious for the homogeneous DNA circle studied by Shimada and Yamakawa. However, the search for the elastic-energy minimum of homogeneous DNA molecules with complex geometry, such as in DNA looping, supercoiling, and the case of inhomogeneous DNA sequences containing curvature and nonuniform DNA flexibility, is not trivial (4,35,36). We have developed a statistical-mechanical theory for sequence-dependent DNA circles and applied it to the problem of DNA cyclization, combining computation of the equilibrium conformation with subsequent evaluation of thermodynamic quantities using a harmonic approximation (26). In this model DNA configuration is described by parameters defined at dinucleotide steps, i.e., tilt, roll, and twist (18), which allows straightforward incorporation of intrinsic or protein-induced DNA curvature at the basepair level. As in Shimada and Yamakawa's method, the theory takes advantage of small fluctuations around one stable mechanical configuration in small DNA circles (e.g., $<\sim$1000 bp). Once the mechanical equilibrium configuration under certain constraints is found with an iterative algorithm, fluctuations around the equilibrium conformation can be taken into account with the harmonic approximation. The new method is much more computationally efficient than Monte Carlo simulation, has comparable accuracy, and has been applied successfully to analyze experimental results from DNA cyclization (26).

Here we extend this theory to DNA looping. The basis of the extension is to treat the protein subunits as connected rigid bodies and to allow for a limited number of degrees of freedom between the subunits. Motions of the subunits are assumed to be governed by harmonic potentials and an associated set of force constants, neglecting the anharmonic terms often required for proteins undergoing large conformational fluctuations among their modular domains. Indeed, the use of a harmonic approximation is supported by the success of continuum elastic models that are based only on shape- and mass-distribution information in descriptions of protein motion (37). Similar to the description used for individual DNA basepairs in the model, protein geometry and dynamics are described by three rigid-body rotation angles (tilt, roll, and twist). Therefore, DNA looping can be viewed as a generalization of DNA cyclization in which the protein component is characterized by a particular set of local geometric constraints and elastic constants. This treatment not only unifies the theoretical descriptions of DNA cyclization and looping, but also allows consideration of flexibilities at protein-DNA and protein-protein interfaces and application of the concepts of linking number and writhe. In previous work, proteins were considered rigid, and their effects on DNA configuration were represented by a set of constraints applied to DNA ends (1,38,39). With the approach described here, programs developed for analyzing DNA cyclization can be used to analyze DNA looping with only minor modifications.

Our method is most applicable to the problem of short DNA loops, in which the free energy of a wormlike chain is dominated by bending and torsional elasticity. As in the previous theory of cyclization, possible modes of DNA self-contact and contacts between protein and DNA at positions other than the binding sites are not considered (26). For large loops, contributions to the free energy from chain entropy and DNA-DNA contacts can become highly significant. Several alternative treatments of DNA looping have appeared recently. One of these addresses the excluded-volume contribution to DNA looping within large open-circular molecules (40), whereas two others consider the effect on looping of traction at the ends of a DNA chain (41,42). None of these treatments includes helical phasing effects on DNA looping. In contrast, a method based on the Kirchhoff elastic-rod model, which includes the helical-phase dependence, has been presented (39,43). However, this approach does not include thermal fluctuations per se and therefore is not directly applicable to calculations of the *J* factor. The comprehensive treatment of small DNA loops described here is thus far unique to the extent that it accounts for sequence- and protein-dependent conformational and flexibility parameters, thermal fluctuations, and helical phasing effects.

THEORETICAL METHODS

DNA-loop model

The complete description of the model and theory for DNA cyclization was presented recently (26). Here we focus only on modifications necessary to treat DNA looping. In this work the protein subunits that mediate loop formation are modeled exclusively as two identical and connected rigid bodies, as shown in Fig. 1. There are three additional sets of rigid-body rotation angles that are defined in addition to those for dinucleotide steps: two sets for the interfaces between protein and the last (DP) and first (PD) basepairs of the DNA and one set for the interface between the two protein domains (PP), where the symbols in parentheses are used to indicate the corresponding angles through subscripts. The local Cartesian-coordinate

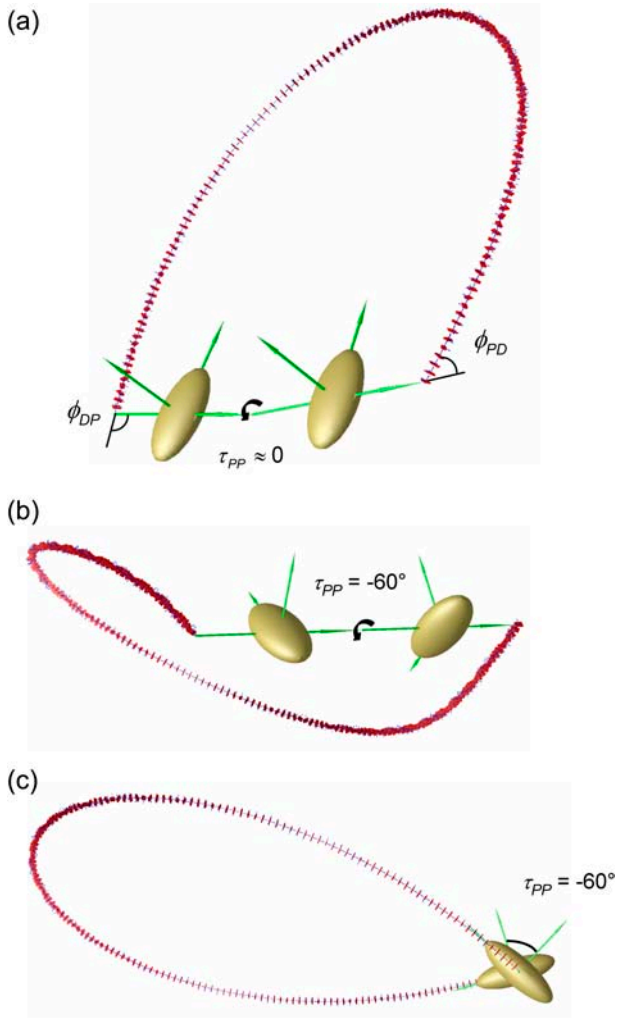

FIGURE 1 Rigid-body models for studies of protein-mediated DNA looping. (*a*) A prototype 137-bp DNA loop generated by interactions with a pair of rigid, DNA-binding protein subunits is shown. DNA basepairs are represented by rectangular slabs (*red*) with axes (*blue*) that indicate the orientation of the local Cartesian-coordinate frame whose origin lies at the center of each basepair. Two sets of coordinate axes (*green*) represent the local coordinate frames embedded in the protein subunits (*gold ellipsoids*) that mediate DNA looping. The coupling of protein and DNA geometry is characterized by tilt, roll, and twist values for the DNA-protein, protein-protein, and protein-DNA interfaces. Three of these variables are shown here: the DNA-protein roll angle, ϕ_{DP}; the protein-protein twist angle, τ_{PP}; and the protein-DNA roll angle, ϕ_{PD}. (*b*) Prototype 179-bp loop with protein-protein twist angle, τ_{PP}, equal to $-60°$. The view is from the base of the loop toward the DNA apex. (*c*) Loop conformation shown in *b*, viewed from the side, perpendicular to the loop-dyad axis.

frames for protein subunits are defined such that their origins coincide with vertices of a circular chain and their z axes point toward the next vertex in succession. Thus protein dimensions can be modeled in terms of a noncanonical value for the helix rise corresponding to particular segments within a circular polymer chain.

Angles are expressed in degrees, and length in units of the DNA helical rise, ℓ_{bp} = 3.4 Å, throughout. Most of our analysis focuses on the basic phenomenon of DNA looping apart from any effects of DNA-sequence-dependent structure or flexibility. Therefore, all calculations used canonical mechanical parameters for duplex DNA: helical twist $\tau_0 = 34.45°$, a sequence-independent twist-angle standard deviation, or twisting flexibility, $\sigma_\tau = 4.388°$, and standard deviations, or bending flexibilites, for all tilt and roll angles, σ_θ and σ_ϕ, respectively, of 4.678° (equivalent to a persistence length of 150 bp). Except for specific cases where intrinsic DNA bending is considered, the average values of tilt and roll are taken to be zero. Computations were carried out on a Dell laptop with 1 GHz Pentium III CPU and 256 Mb memory. The program source code is written in Fortran 90 and is available upon request.

Simplified protein geometries and flexibility parameters

For DNA loops with either zero or nonzero end-to-end distances, constraints are directly applied to the DNA ends, as in the case of DNA cyclization. We modeled DNA loops formed during site synapsis using protein-dependent parameters roll = $\phi_{DP} = \phi_{PD} = 90°$ and twist = $\tau_{DP} = \tau_{PD} = 34.45°$. The angle τ_{PP} was considered an adjustable parameter that we denote the axial angle and, unless specified, all other protein-related angular parameters were set equal to 0°. In these cases, the DNA ends (the centers of two protein-binding sites on DNA) are separated by twice the protein-arm length ℓ_p and displaced from one another along the $+x$ direction, or toward the major groove of DNA. Projected along the x axis, the axial angle is the included angle between the tangents to the DNA at the two protein-binding sites and is altered by varying the twist between protein subunits (Fig. 1, *b* and *c*). An axial angle equal to 0 corresponds to antiparallel axes at the ends as shown in Fig. 1 *a*. The case of a rigid protein assembly is modeled by setting the standard deviations of the DP, PP, and PD sets of rigid-body rotation angles to 1×10^{-8}°.

RESULTS AND DISCUSSION

DNA loops having zero end-to-end distance and antiparallel helical axes

DNA loops containing N basepairs in which the two ends meet in an antiparallel orientation can be empirically described by the following formula:

$$\begin{cases} \text{Tilt}: & \theta_i = -A_i \cos(180 + \delta) \\ \text{Roll}: & \phi_i = A_i \sin(180 + \delta) \\ \text{Twist}: & \tau_i = \tau^0 \end{cases}, \quad (1)$$

where τ^0 is the intrinsic DNA twist and δ an arbitrary angle related to the unconstrained torsional degree of freedom of DNA. The coefficients A_i are given by

$$A_i = \frac{1}{N} f\left(\frac{i}{N-1}\right), \; i = 0, \ldots, N-1, \quad (2)$$

with

$$f(x) = \begin{cases} g(x), & 0 \le x \le 0.5 \\ g(0.5 - x), & 0.5 < x \le 1 \end{cases}, \quad (3)$$

where

$$g(x) = \sum_{i=1}^{5} a_i x^i, \quad 0 \le x \le 0.5. \quad (4)$$

The coefficients in Eq. 4 were obtained by fitting the space curve corresponding to the DNA helical axis that gives the minimum elastic energy conformation of DNA loops of different sizes as follows: $a_0 = -335.0142$, $a_1 = 2318.881$,

$a_2 = -1299.164$, $a_3 = -4483.366$, $a_4 = 38169.74$, and $a_5 = -54753.5$. The error for end-to-end distances computed using Eq. 1 is $<2\%$ of the DNA length from 50 bp to 100 bp, and $<0.5\%$ of that from 100 bp to 500 bp. The torsional phase angle between two ends is $\xi = -(N-2)\tau - 2\delta$. The entire loop lies in a plane, and the angle between the normal vector of the plane and the x axis of the external coordinate frame can be shown to be $\psi = 180 + \tau - \delta$. The expressions for ξ and ψ suggest that δ is related to DNA bending isotropy. Loop configurations with different δ-values are related to each other by globally twisting DNA molecules. Since the orientation of the first basepair is fixed, this global twist is equivalent to rotation of the loop plane, which corresponds to the rotational symmetry met in DNA cyclization of homogeneous DNA with bending isotropy (26) . Therefore, J factors for configurations with different δ-values are identical.

If DNA looping needs to be torsionally in-phase, only two degenerate loop configurations are available, breaking the rotational symmetry. These loop geometries can be expressed by Eq. 1 with two different δ-values, i.e., $\delta_1 = -(N-2)\tau/2$ and $\delta_2 = 180 - (N-2)\tau/2$, which satisfy the torsional phase requirement $\xi = 360 \times n$, $n = 0, \pm 1, \pm 2, \ldots$. In contrast to DNA cyclization, no twist change is involved in forming these ideal DNA loops for any DNA length and thus the helical dependence vanishes in this case. From the expression given above for ψ it is clear that the helical axes of the two loops are coincident and their directions are reversed. Fig. 2 shows the bending profile of the loop configuration corresponding to δ_1 for a 150-bp DNA. Surprisingly, the maximal J factor occurs at approximately the same DNA length, or 460 bp (data not shown), as in DNA cyclization (26). This can be partly explained by the fact that the total bending magnitude of the loop is 290°, close to a full circle, instead of 180°.

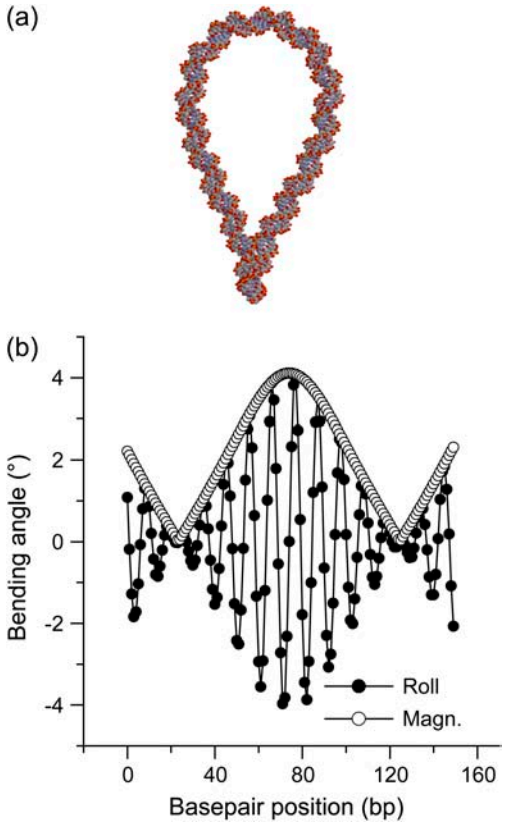

FIGURE 2 Conformation of an antiparallel, 150-bp DNA loop with zero end-to-end distance. (*a*) Computed space-filling model of the loop generated with 3DNA (49). The ends of the DNA juxtapose exactly with antiparallel helical axes and exact torsional phasing. (*b*) Equilibrium roll and magnitude of the loop shown in *a*. The bending magnitude of each dinucleotide step is defined as $\sqrt{\theta_i^2 + \phi_i^2}$, where θ_i and ϕ_i are the tilt and roll, respectively, of the ith dinucleotide step.

DNA looping with finite end-to-end distance, antiparallel helical axes, and in-phase torsional constraint

Separation of the DNA ends breaks the rotational symmetry, restoring the dependence on helical twist. Fig. 3 *a* shows the J factor as a function of DNA length for end-to-end distances of 10 bp and 30 bp. The helical dependence increases with end-to-end separation. Starting from the two loop configurations (corresponding to δ_1 and δ_2) with zero end-to-end distance and in-phase torsional alignment as initial configurations, two mechanical equilibrium configurations are obtained by using the iterative algorithm described previously (26). The J factor in Fig. 3 *a* is the sum of separate J factors calculated for the two configurations. Note that in all cases involving configurations that differ in linking number, equilibration between the two forms requires breakage of at least one of the protein-DNA interfaces. The contributions from each of these configurations are shown in detail for the case where the ends are separated by 10 bp. Interestingly, the length dependence of J computed from the individual configurations are out of phase and have a periodicity of two helical turns, which results from the half-twist dependence of the phase angles δ_1 and δ_2. However, their sum displays a periodicity of one helical turn. Fig. 3, *b* and *c*, shows two such configurations for DNA molecules that are torsionally in-phase ($N = 210$ bp) or out-of-phase ($N = 215$ bp).

In the case of cyclization, the helical-phase dependence of the J factor persists at DNA lengths well beyond that corresponding to the maximum value of J, which lies near 500 bp. This is clearly not the case for DNA looping. In Fig. 3 *a*, the periodic dependence of J on DNA length for 10-bp end-to-end separation decays nearly to zero well before the maximum J value is reached. Although the periodicity of J is not attenuated quite as strongly for 30-bp end separation, there is a <4-fold variation in the value of J near 300 bp, as opposed to the >10-fold variation in cyclization J factors expected in this length range. The differences between looping and cyclization are largely due to substantial differences in the relative contributions of DNA writhe in the two processes, discussed at length below.

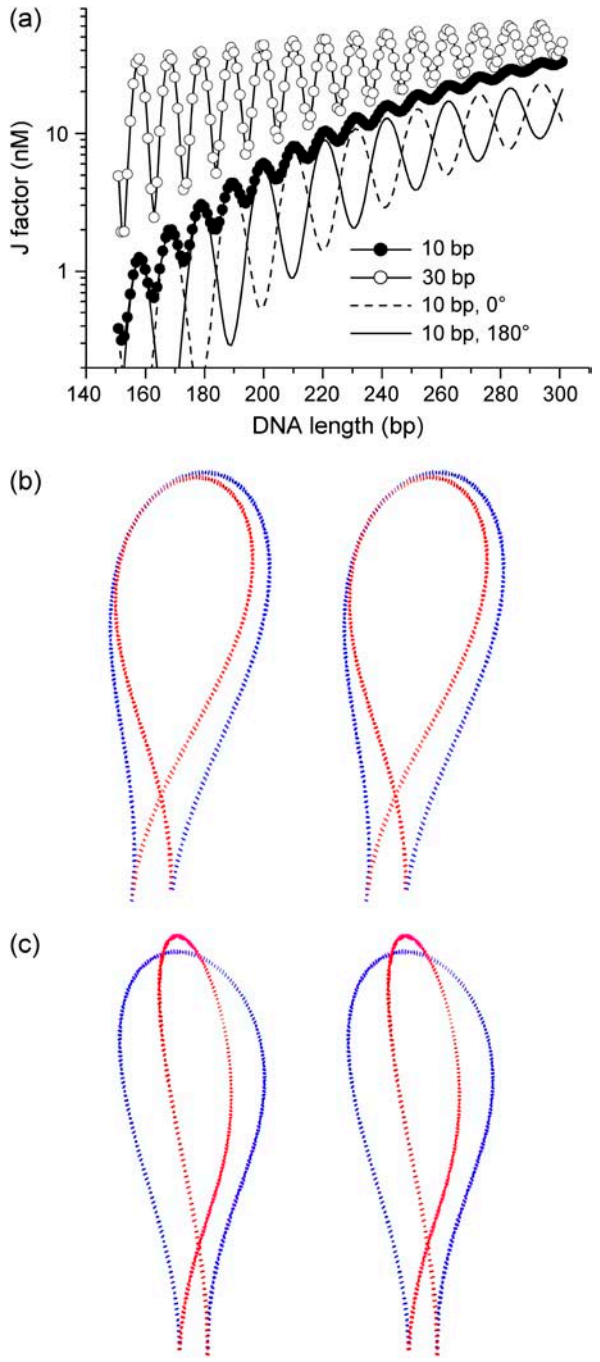

FIGURE 3 DNA-length-dependent J factor and loop configuration as a function of end-to-end separation. (*a*) The helical dependence of DNA looping is shown for values of the end-to-end separation equal to 10 bp and 30 bp. The two configurations for the 10-bp separation are obtained from corresponding configurations with zero end-to-end separation by using an iterative algorithm. Therefore, the two configurations are designated by the initial configurations with phase angles $\delta = -(N-2)\tau/2+0$ (0°, *dashed line*) and $\delta = -(N-2)\tau/2+180$ (180°, *solid line*), as described in the text. (*b* and *c*) Stereo models of the two equilibrium configurations for 210-bp (*b*) and 215-bp (*c*) antiparallel DNA loops with end-to-end separation equal to 10 bp. The 210- and 215-bp DNA correspond to an adjacent peak and valley, respectively, of the curve in *a*. Conformations shown in blue correspond to $\delta = 0$; those shown in red are for $\delta = 180°$. Note that for N-bp DNA, the chain contour length is equal to $(N-1)\ell_{bp}$.

DNA looping in synapsis

Intramolecular reactions of most site-specific recombination systems (9–11) and a number of DNA restriction endonucleases, such as *Sfi*I and *Ngo*MIV (12), proceed through protein-mediated intermediate structures in which a pair of DNA sites are brought together in space and the intervening DNA is looped out. The intermediate nucleoprotein complex involved in site pairing and strand cleavage (and also exchange, in the case of recombinases) is termed the synaptic complex. In these systems, two characteristic geometric parameters are of interest: the average through-space distance between the sites and the average crossing angle between the two ends of the loop, which we denote the axial angle. The latter quantity can be described in terms of the twist angle between the protein domains, τ_{PP} (Fig. 1 *b*), and we shall use these terms interchangeably. Here we focus on the effect of protein geometry on DNA looping, leaving consideration of protein flexibility to the following section.

Fig. 4 shows the helical dependence of looping (Fig. 4 *a*) and the elastic-minimum configuration of DNA loops (Fig. 4 *b*) for different values of the axial angle. The most prominent feature of these results is that the phase of the helical dependence is shifted as a function of the axial angle,

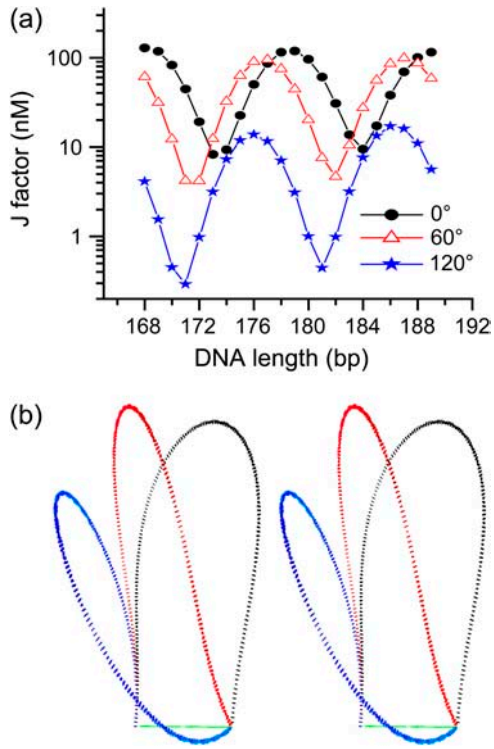

FIGURE 4 Dependence of the J factor on axial angle. (*a*) DNA-length dependence of J for axial angles of 0°, 60°, and 120° with the end-to-end separation set equal to 40 bp. Note that the positions of the extrema shift to the left with increasing values of the axial angle. (*b*) Stereo models of minimum elastic-energy conformations of 179-bp loops color coded in accord with the corresponding axial-angle values in *a*.

characterized by a relative global shift of the curve along the x axis. This implies that DNA looping does not always occur most efficiently when two sites are separated by an integral number of helical turns, as has been suggested for some simple DNA looping systems studied previously. The axial angle also globally modulates J factors, which is apparent from the vertical shift in the J versus length curve and effects on the amplitude of the helical dependence. The torsion-angle-independent value of J, averaged over a full helical turn, decreases with increasing axial angle, whereas the amplitude of the helical dependence increases. The above observations can be qualitatively explained by analogous results from DNA cyclization. As in cyclization, DNA forms loops most efficiently when the number of helical turns in the loop is close to an integer value. It is therefore appropriate to consider this issue in terms of the linking number for the looped conformation, Lk, which involves contributions from the geometries of both the protein and DNA.

We define the loop helical turn ($H_{t,loop}$) as the sum of the DNA twist and the twist introduced by the protein subunits, divided by 360. Therefore, changing the twist angle τ_{PP}, the axial angle, will shift the phase of the helical dependence relative to that of the DNA alone. For a loop with $N = 179$ bp and $\tau_{PP} = 0$, the total twist is simply equal to that for the DNA loop. Because this loop has 17.0 helical turns, only one loop topoisomer contributes to the J factor. The value of J is a local maximum at $\tau_{PP} = 0$ and, as shown in Fig. 5 *a*, decreases monotonically for both $\tau_{PP} > 0$ and $\tau_{PP} < 0$. Contributions to J from other topoisomers of the 179-bp loop are $<5\%$ over the range $-135° > \tau_{PP} > +120°$. The twist for the planar equilibrium conformation of a 173-bp loop is 16.5 helical turns; thus, there are two alternative loops that can be efficiently formed (Fig. 5 *a*): either a loop with $H_{t,loop} = 17.0$ and $\tau_{PP} > 0$ or a loop with $H_{t,loop} = 16.0$ and $\tau_{PP} < 0$. The J value at $\tau_{PP} = 0$ is a local minimum and there is a bimodal dependence on axial angle for loops in which the DNA twist is half-integral. We investigated the phase shift of the J factor and found that this quantity is a nonlinear function of the axial angle. From Fig. 4 *a*, the calculated phase shifts for 60° and 120° axial angles relative to 0° are ~52° and 103°, respectively. Moreover, the local maxima for the total J curve for $N = 173$, shown in Fig. 5 *a*, are located at −58.5° and 63°, positions that are not in agreement with predicted angle values based solely on $H_{t,loop}$ −166° and 194°, respectively).

These deviations can be explained by the fact that writhe makes an important contribution to the overall Lk for the loop. This aspect of DNA looping is dramatically different from that in the cyclization of small DNA molecules. The conformations of small DNA circles are close to planar and the writhe contribution is small relative to DNA twist (26,30,44,45). In the case of protein-mediated looping, nonzero values of the axial angle impose an intrinsically nonplanar conformation on the DNA. The relative contributions of loop writhe and twist for the $Lk = 16$ topoisomer of a 173-bp loop are shown as a function of axial angle in Fig. 5 *b*.

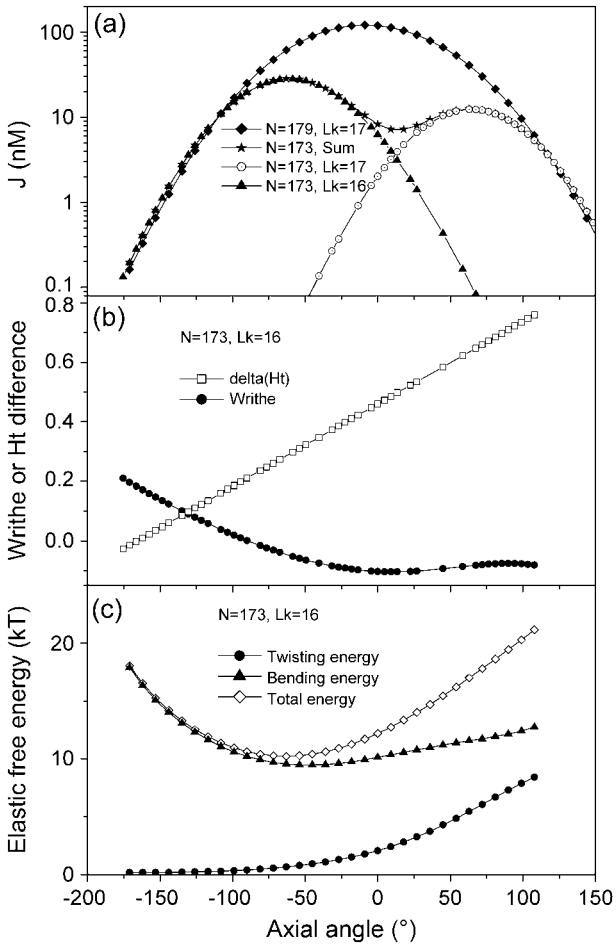

FIGURE 5 J factor, loop-geometry parameters, and elastic free energies as functions of axial angle. (*a*) J factor values for loop topoisomers corresponding to 179-bp and 173-bp loops in Fig. 4. The principal contribution to J for $N = 179$ bp comes from a single loop topoisomer with $Lk = 17$. For $N = 173$ bp, the overall J factor is the sum of contributions from two loop topoisomers with Lk values of 16 and 17, generating a bimodal dependence of J on axial angle as described in the text. (*b*) Excess helical twist, ΔH_t, and writhe of the loop formed by the $Lk = 16$ topoisomer for $N = 173$ bp as a function of axial angle. Excess twist is computed from the expression $H_{t,\,loop} - 16$, where $H_{t,\,loop}$ is the loop helical turn value described in the text, and depends linearly on the axial angle. The writhing number of the loop was calculated using the method of Vologodskii (28,50). (*c*) Elastic free energies of the $Lk = 16$ loop topoisomer for $N = 173$ bp calculated according to Eq. 38 of Zhang and Crothers (26). The individual contributions of bending and twisting energies are shown along with their sum.

In Fig. 5 *c*, we plot the axial-angle-dependent values of the bending and twisting free energies for the $Lk = 16$ topoisomer and their sum, which is the total elastic free energy of the loop. The minimum value of the total elastic energy occurs at $\tau_{PP} = -58.5°$, coincident with the position of the J factor maximum for this topoisomer (Fig. 5 *a*). This mechanical state can be achieved with very little twist deformation of the loop, but at the expense of significant bending energy. Further reduction of the axial angle requires even less twisting energy; however, the bending energy

increases monotonically. In contrast, for $\tau_{PP} > -58.5°$, somewhat less bending energy is required, but the twisting energy begins to increase significantly with increasing axial angle. Since the sense of the bending deformation for $\tau_{PP} > 0$ opposes the needed reduction in loop linking number, the elastic energy cannot be decreased by increasing the axial angle. The only way that the loop geometry can compensate for this is through twist deformation. This asymmetry arises because we are considering the contribution of only one loop topoisomer to the elastic free energy.

Effects of binding-site symmetry

Twofold symmetry in the DNA-binding domain of a protein or the sequence of its cognate site may allow formation of two alternative looping geometries that conserve protein-DNA contacts. This is shown in Fig. 6, which depicts looping mediated by the type-II restriction endonuclease *Sfi*I as an example. *Sfi*I binds two copies of its recognition sequence and, in the presence of Mg^{2+}, catalyzes the concerted cleavage of all four DNA strands. These symmetric recognition sequences can be juxtaposed via two alternative loop geometries whose axial angles differ by 180°. If the reaction steps subsequent to synapsis are independent of DNA orientation, in particular that of the spacer sequence shown in Fig. 6, then the overall efficiency of the reaction should be related to the total J factor of the two geometries. Fig. 7 shows the configuration corresponding to these geometries and the helical dependence of the individual and total J factors for 179-bp DNA loops. From Fig. 7, *a* and *c*, it can be seen that the J dependences for the two geometries are out of phase, and may shift relative to each other along the vertical axis.

If the reaction is governed by the thermodynamics of loop formation, then the experimentally observed product yield should again be proportional to the sum of J values for the two alternative geometries. When the axial angle equals 90° (Fig. 7 *a*), two out-of-phase loop topoisomers with similar torsion-angle-independent J factors contribute nearly equally to the overall value of J. The net helical dependence is

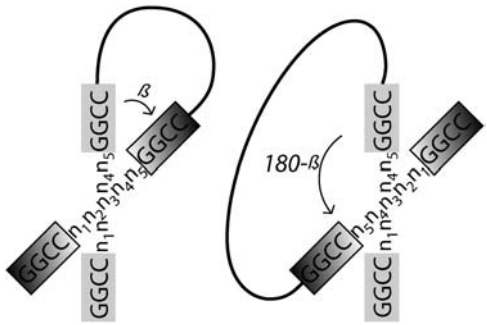

FIGURE 6 Alternative looping geometries in systems with binding-site symmetry. Two loop geometries can be formed if a protein's DNA-binding domain or its cognate binding site has twofold symmetry. The example shown here is the *Sfi*I recognition sequence, GGCCnnnn$^{\downarrow}$nGGCC, where n is any base and the arrow indicates the location of the cleavage site. Each subunit of the tetrameric *Sfi*I binds to one GGCC sequence and two dimers bound at two distant copies of the recognition site associate to form a synaptic complex by looping the intervening DNA. Four DNA backbones are then cut in concert to release the looped DNA. The two geometries are related by reversing the intrinsic DNA direction at the binding site if the protein is rigid, forming a negative (*left*) or positive (*right*) crossing according to the right-hand rule. Due to the twofold symmetry of the protein dimer, this reversal does not affect protein-DNA interactions. Given the angle shown in the figure, the protein twist τ_{PP} used to model the geometry is $-\beta$ and $180 - \beta$ for the left and right configurations, respectively.

strongly diminished (<1.7-fold variation) and the overall helical periodicity is one-half the helical repeat of the DNA. The insensitivity of J to helical phasing occurs despite the strong helical dependence of the individual loop geometries. However, once the axial angle deviates from a right angle, the J factor contributed by the geometry with the smaller absolute value of axial angle dominates the total J factor (Fig. 7 *c*), which restores the helical dependence.

Two alternate looping geometries and their out-of-phase contributions to loop formation were observed for *Sfi*I by native polyacrylamide gel electrophoresis of the synaptic complexes (12). Our calculations not only explain the observation, but also predict that the *Sfi*I tetramer binds its two cognate sites to form a crossed DNA structure, because the

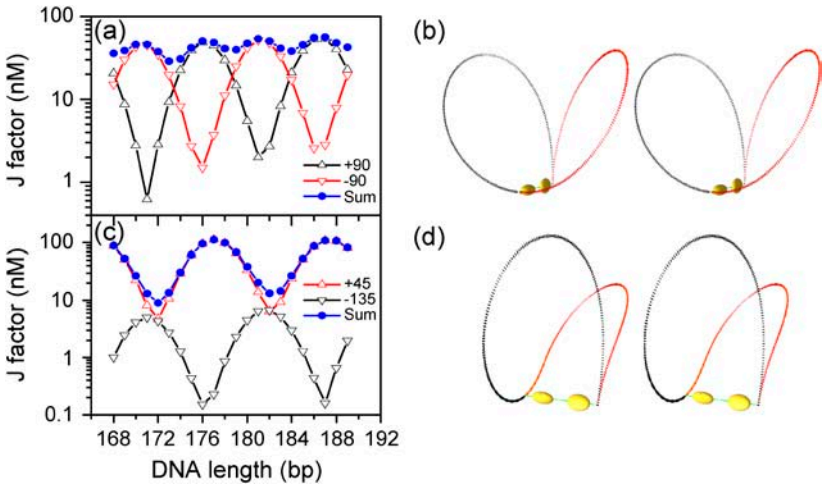

FIGURE 7 Effects of binding-site symmetry on the J factor and loop geometry. (*a*) Individual J factors computed for the two alternative looping geometries of a synaptic complex with twofold symmetry and their sum. The τ_{PP} values for these conformations are +90° and −90°. (*b*) Stereo models of the equilibrium loop geometries color coded to correspond with the J factor values in *a*. (*c*) Individual J factor contributions for loops with $\tau_{PP} = 45°$ and −135°. (*d*) Stereo models corresponding to the equilibrium loop geometries in *c*. DNA length was equal to 179 bp in all cases and the end-to-end separation was equal to 40 bp.

two DNA looping geometries have similar contributions in the helical-twist assay. Interestingly, a cocrystal structure of a similar endonuclease, *Ngo*MIV, with two copies of its recognition site (14) exhibits an axial angle of −60° (or equivalently 120°), instead of ~90° (or −90°). The difference may be due to either slightly different architectures for protein-DNA and protein-protein association of the two endonucleases, or flexibility at the protein-protein interfaces that allows preferences for axial angles to be perturbed by crystal-packing forces. It would be informative to investigate DNA loop geometries for the *Ngo*MIV system in solution by a helical-twist assay similar to that done with *Sfi*I.

Effects of protein flexibility

DNA loops as small as 40 to 70 bp have been widely observed in experiments, for example, in LacR- and AraC-mediated DNA looping (20,46). For the case of AraC, no lower limit on site spacing was found. Because it is energetically unlikely for DNA to form such small loops with rigid end constraints, plasticity in the protein assembly has been proposed to explain these observations. Protein plasticity can manifest itself through two mechanisms: multiple stable protein states with different configurations, as demonstrated by LacR (31), and protein conformational flexibility around specific states. To investigate the potential effects of protein elasticity, we considered DNA loop configurations mediated by the extended conformation of the LacR tetramer (31). The minimum-elastic-energy conformation of this protein structure corresponds to $\phi_{DP} = \phi_{PD} = 90.0°$ with all other DP, PD, and PP parameters set equal to 0°.

Fig. 8, *a* and *b*, shows loop configurations for 137-bp and 53-bp DNAs in which roll and twist fluctuations were permitted at the protein-protein and both protein-DNA interface steps. Both DNA and protein configurations change with protein flexibility. By increasing protein flexibility, the system alleviates the need for strong DNA bending required to close the loop. Remarkably, DNA loops as small as 30 bp can easily be formed, as evidenced in the *J* dependence shown in Fig. 8 *b*. Higher protein flexibilities enable looping more dramatically for short DNAs because the bending energy required by looping is sharply decreased, but less dramatically for longer DNA segments because of the greater loss of chain entropy. Consequently, the peak corresponding to optimal DNA loop length is shifted to smaller values when protein flexibility is increased. This comparison not only explains the experimental observations of looping for short DNAs, but also provides a convenient method to assess the effective degree of protein flexibility from the optimum DNA length for loop formation.

Effects of intrinsic or protein-induced DNA curvature

Static DNA curvature can significantly alter the thermodynamics of DNA looping. In a dramatic demonstration of this

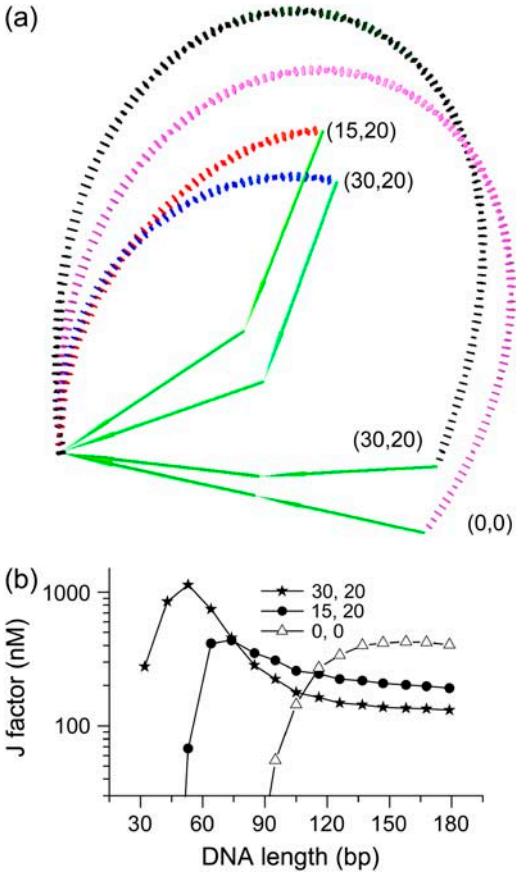

FIGURE 8 Effects of protein flexibility on loop configuration and *J* factor. (*a*) Equilibrium conformations of 53-bp (*blue*, *red*) and 137-bp (*violet*, *black*) loops mediated by LacR in its extended conformation, as described in the text. Protein flexibilities are specified as pairs of values ($\sigma_\phi^{PP} = \sigma_\tau^{PP}$, $\sigma_\phi^{DP} = \sigma_\phi^{PD} = \sigma_\tau^{PD} = \sigma_\tau^{DP}$), where the former number gives the bending and twisting fluctuations in degrees for the protein-protein interface and the latter corresponds to the values for the protein-DNA interfaces. (*b*) Variation of the *J* factor with DNA length for flexible protein assemblies. Only DNAs with intrinsic torsional in-phase ends are shown for clarity. Protein-flexibility parameters are those given (*a*).

effect, Kahn and co-workers were able to design hyperstable LacR-DNA loops by incorporating A-tracts within intervening DNA (31,47). To examine the general effects of DNA curvature on the helical dependence, we carried out calculations for a single kink of 36° or 72° introduced at different positions in loops mediated by the extended LacR conformation. The calculated dependence of *J* factor on DNA length is shown in Fig. 9. Depending on kink position, a kink can shift the phase and change the amplitude of the helical dependence. If the kink position is held constant, an increase in kink magnitude increases the amplitude of the helical dependence. The strong dependence of DNA looping on kink position and magnitude creates an opportunity for flexible genetic control by architectural proteins through DNA bending. The helical dependence of DNA looping modulated by DNA bending differs significantly from that in DNA cyclization, in which the relative position of a single

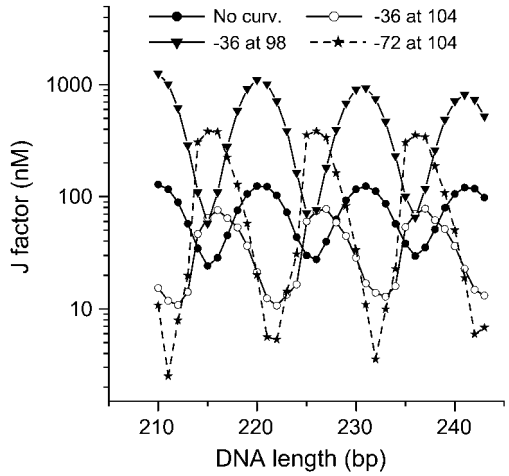

FIGURE 9 Dependence of *J* factor on intrinsic DNA bending. Effects of a single kink on the length dependence of *J* was modeled in terms of a roll angle of −36° or −72° placed at specified base-pair positions. The protein-protein and protein-DNA parameters were those of the extended LacR conformation used in the analysis shown in Fig. 8, and flexibility parameters were fixed at 30° for the protein-protein interface and 20° for protein-DNA interfaces. The kink was placed at either the 98th or 104th dinucleotide step.

kink does not alter the *J* factor and amplitude of the helical dependence. This comparison suggests that it is intrinsically inaccurate to determine DNA torsional rigidity by DNA looping if intrinsic bending is present. It also implies that a rigorous examination of effects of DNA bending on looping requires an additional phasing assay, in which the position of induced DNA curvature is systematically varied to determine the optimal location of the bending locus (48).

CONCLUSIONS

We have developed a statistical mechanical theory for DNA looping and used the theory to investigate the helical dependence of DNA-loop formation. Our results suggest that the helical dependence of DNA looping is affected by many factors and lead to the conclusion that whereas a positive helical-twist assay can often confirm DNA looping, a negative result cannot exclude DNA looping. Since it is difficult to explore the architecture of DNA loops with current experimental techniques, this theory will be useful for more reliably analyzing DNA looping with limited experimental data. Our theory has advantages over previous approaches based exclusively on DNA mechanics, particularly when protein flexibility is taken into account. In these cases, entropy effects become important and are responsible for the observed decay of looping efficiency with DNA length. Effects that involve chain entropy cannot be accounted for using models based on minimum-elastic-energy conformations (26).

This work was supported by the Joint Division of Mathematical Sciences/National Institute of General Medical Sciences' Mathematical Biology Initiative (NIH GM 67242 to S.D.L. and I.K. Darcy), by National Institutes of Health grant GM 21966 to D.M.C., and by the National Foundation for Cancer Research through the Yale-NFCR Center for Protein and Nucleic Acid Chemistry.

REFERENCES

1. Schleif, R. 1992. DNA looping. *Annu. Rev. Biochem.* 61:199–223.
2. Matthews, K. S. 1992. DNA looping. *Microbiol. Rev.* 56:123–136.
3. Dunn, T. M., S. Hahn, S. Ogden, and R. F. Schleif. 1984. An operator at −280 base-pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. USA.* 81: 5017–5020.
4. Geanacopoulos, M., G. Vasmatzis, V. B. Zhurkin, and S. Adhya. 2001. Gal repressosome contains an antiparallel DNA loop. *Nat. Struct. Biol.* 8:432–436.
5. Mossing, M. C., and M. T. Record. 1986. Upstream operators enhance repression of the *lac* promoter. *Science.* 233:889–892.
6. Valentin-Hansen, P., B. Albrechtsen, and J. E. Love Larsen. 1986. DNA-protein recognition: demonstration of three genetically separated operator elements that are required for repression of the *Escherichia coli deoCABD* promoters by the DeoR repressor. *EMBO J.* 5:2015–2021.
7. Rippe, K., M. Guthold, P. H. von Hippel, and C. Bustamante. 1997. Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of *E. coli* RNA polymerase · σ(54) holoenzyme by scanning force microscopy. *J. Mol. Biol.* 270:125–138.
8. Ogata, K., K. Sato, and T. Tahirov. 2003. Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr. Opin. Struct. Biol.* 13:40–48.
9. Van Duyne, G. D. 2001. A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* 30:87–104.
10. Benjamin, H., and N. Cozzarelli. 1986. DNA-directed synapsis in recombination: slithering and random collision of sites. *Proc. Robert A. Welch Found. Conf. Chem. Res.* 29:107–126.
11. Tsen, H., and S. D. Levene. 1997. Supercoiling-dependent flexibility of adenosine-tract-containing DNA detected by a topological method. *Proc. Natl. Acad. Sci. USA.* 94:2817–2822.
12. Watson, M. A., D. M. Gowers, and S. E. Halford. 2000. Alternative geometries of DNA looping: An analysis using the *Sfi*I endonuclease. *J. Mol. Biol.* 298:461–475.
13. Embleton, M. L., S. A. Williams, M. A. Watson, and S. E. Halford. 1999. Specificity from the synapsis of DNA elements by the *Sfi*I endonuclease. *J. Mol. Biol.* 289:785–797.
14. Deibert, M., S. Grazulis, G. Sasnauskas, V. Siksnys, and R. Huber. 2000. Structure of the tetrameric restriction endonuclease *NgoM*IV in complex with cleaved DNA. *Nat. Struct. Biol.* 7:792–799.
15. Ross, E. D., P. R. Hardwidge, and L. J. Maher. 2001. HMG proteins and DNA flexibility in transcription activation. *Mol. Cell. Biol.* 21: 6598–6605.
16. Schleif, R. 2000. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* 16:559–565.
17. Hochschild, A. 1991. Detecting cooperative protein-DNA interactions and DNA loop formation by footprinting. *Methods Enzymol.* 208: 343–361.
18. Bloomfield, V. A., D. M. Crothers, and I. Tinoco. 1999. Nucleic Acids: Structures, Properties, and Functions. University Science Books, Sausalito, CA.
19. Finzi, L., and J. Gelles. 1995. Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science.* 267:378–380.
20. Müller, J., S. Oehler, and B. Müller-Hill. 1996. Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J. Mol. Biol.* 257:21–29.

21. Rippe, K., P. H. von Hippel, and J. Langowski. 1995. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* 20:500–506.

22. Rippe, K. 2001. Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* 26:733–740.

23. Shore, D., J. Langowski, and R. L. Baldwin. 1981. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA.* 78:4833–4837.

24. Shore, D., and R. L. Baldwin. 1983. Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J. Mol. Biol.* 170: 957–981.

25. Crothers, D. M., J. Drak, J. D. Kahn, and S. D. Levene. 1992. DNA bending, flexibility, and helical repeat by cyclization kinetics. *Methods Enzymol.* 212:3–29.

26. Zhang, Y. L., and D. M. Crothers. 2003. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.* 84:136–153.

27. Klenin, K., H. Merlitz, and J. Langowski. 1998. A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* 74:780–788.

28. Vologodskii, A. V., V. V. Anshelevich, A. V. Lukashin, and M. D. Frank-Kamenetskii. 1979. Statistical-mechanics of supercoils and the torsional stiffness of the DNA double helix. *Nature.* 280:294–298.

29. Hagerman, P. J. 1985. Analysis of the ring-closure probabilities of isotropic wormlike chains: application to duplex DNA. *Biopolymers.* 24: 1881–1897.

30. Levene, S. D., and D. M. Crothers. 1986. Ring-closure probabilities for DNA fragments by Monte-Carlo simulation. *J. Mol. Biol.* 189:61–72.

31. Edelman, L. M., R. Cheong, and J. D. Kahn. 2003. Fluorescence resonance energy transfer over ~130 basepairs in hyperstable *lac* repressor-DNA loops. *Biophys. J.* 84:1131–1145.

32. Shimada, J., and H. Yamakawa. 1984. Ring-closure probabilities for twisted wormlike chains: application to DNA. *Macromolecules.* 17: 689–698.

33. Bacolla, A., R. Gellibolian, M. Shimizu, S. Amirhaeri, S. Kang, K. Ohshima, J. E. Larson, S. C. Harvey, B. D. Stollar, and R. D. Wells. 1997. Flexible DNA: genetically unstable CTG·CAG and CGG·CCG from human hereditary neuromuscular disease genes. *J. Biol. Chem.* 272:16783–16792.

34. Law, S. M., G. R. Bellomy, P. J. Schlax, and M. T. Record. 1993. *In vivo* thermodynamic analysis of repression with and without looping in *lac* constructs: estimates of free and local *lac* repressor concentrations and of physical properties of a region of supercoiled plasmid DNA *in vivo*. *J. Mol. Biol.* 230:161–173.

35. Hao, M. H., and W. K. Olson. 1989. Global equilibrium-configurations of supercoiled DNA. *Macromolecules.* 22:3292–3303.

36. Coleman, B. D., W. K. Olson, and D. Swigon. 2003. Theory of sequence-dependent DNA elasticity. *J. Chem. Phys.* 118:7127–7140.

37. Ming, D., Y. Kong, M. A. Lambert, Z. Huang, and J. Ma. 2002. How to describe protein motion without amino acid sequence and atomic coordinates. *Proc. Natl. Acad. Sci. USA.* 99:8620–8625.

38. Tobias, I., B. D. Coleman, and W. K. Olson. 1994. The dependence of DNA tertiary structure on end conditions: theory and implications for topological transitions. *J. Chem. Phys.* 101:10990–10996.

39. Balaeff, A., L. Mahadevan, and K. Schulten. 1999. Elastic rod model of a DNA loop in the *lac* operon. *Phys. Rev. Lett.* 83:4900–4903.

40. Hanke, A., and R. Metzler. 2003. Entropy loss in long-distance DNA looping. *Biophys. J.* 85:167–173.

41. Yan, J., R. Kawamura, and J. F. Marko. 2005. Statistics of loop formation along double helix DNAs. *Phys. Rev. E Stat.* 71:061905.

42. Blumberg, S., A. V. Tkachenko, and J. C. Meiners. 2005. Disruption of protein-mediated DNA looping by tension in the substrate DNA. *Biophys. J.* 88:1692–1701.

43. Balaeff, A., C. R. Koudella, L. Mahadevan, and K. Schulten. 2004. Modelling DNA loops using continuum and statistical mechanics. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 362:1355–1371.

44. Shore, D., and R. L. Baldwin. 1983. Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol.* 170:983–1007.

45. Horowitz, D. S., and J. C. Wang. 1984. Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol.* 173:75–91.

46. Lee, D. H., and R. F. Schleif. 1989. *In vivo* DNA loops in *araCBAD*: size limits and helical repeat. *Proc. Natl. Acad. Sci. USA.* 86:476–480.

47. Mehta, R. A., and J. D. Kahn. 1999. Designed hyperstable *lac* repressor · DNA loop topologies suggest alternative loop geometries. *J. Mol. Biol.* 294:67–77.

48. Ross, E. D., A. M. Keating, and L. J. Maher. 2000. DNA constraints on transcription activation *in vitro*. *J. Mol. Biol.* 297:321–334.

49. Lu, X. J., and W. K. Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108–5121.

50. Klenin, K., and J. Langowski. 2000. Computation of writhe in modeling of supercoiled DNA. *Biopolymers.* 54:307–317.