First-principles calculation of DNA looping in tethered particle experiments

Kevin B Towles[1]**, John F Beausang**[1]**, Hernan G Garcia**[2]**, Rob Phillips**[3]**, and Philip C Nelson**[1]

[1]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia PA 19104

[2]Department of Physics, California Institute of Technology, Pasadena CA 91125

[3]Division of Engineering and Applied Science, California Institute of Technology, Pasadena CA 91125

E-mail: `nelson@physics.upenn.edu`

Abstract. We calculate the probability of DNA loop formation mediated by regulatory proteins such as Lac repressor (LacI), using a mathematical model of DNA elasticity. Our model is adapted to calculating quantities directly observable in Tethered Particle Motion (TPM) experiments, and it accounts for all the entropic forces present in such experiments. Our model has *no free parameters*; it characterizes DNA elasticity using information obtained in other kinds of experiments. It assumes a harmonic elastic energy function (or wormlike chain type elasticity), but our Monte Carlo calculation scheme is flexible enough to accommodate arbitrary elastic energy functions. We show how to compute both the "looping J factor" (or equivalently, the looping free energy) for various DNA construct geometries and LacI concentrations, as well as the detailed probability density function of bead excursions. We also show how to extract the same quantities from recent experimental data on tethered particle motion, and then compare to our model's predictions. In particular, we present a new method to correct observed data for finite camera shutter time and other experimental effects.

Although the currently available experimental data give large uncertainties, our first-principles predictions for the looping free energy change are confirmed to within about $1\,k_{\mathrm{B}}T$, for loops of length around 300 basepairs. More significantly, our model successfully reproduces the detailed distributions of bead excursion, including their surprising three-peak

structure, without any fit parameters and without invoking any alternative conformation of the LacI tetramer. Indeed, the model qualitatively reproduces the observed dependence of these distributions on tether length (e.g., phasing) and on LacI concentration (titration). However, for short DNA loops (around 95 basepairs) the experiments show more looping than is predicted by the harmonic-elasticity model, echoing other recent experimental results. Because the experiments we study are done *in vitro,* this anomalously high looping cannot be rationalized as resulting from the presence of DNA-bending proteins or other cellular machinery. We also show that it is unlikely to be the result of a hypothetical "open" conformation of the LacI tetramer.

PACS numbers: 87.14.gk,87.80.Nj, 82.37.Rs, 82.35.Pq, 36.20.Ey

1. Introduction and summary

1.1. Background

Living cells must orchestrate a multitude of biochemical processes. Bacteria, for example, must rigorously suppress any unnecessary activities to maximize their growth rate, while maintaining the potential to carry out those activities should conditions change. For example, in a glucose-rich medium *E. coli* turn off the deployment of the machinery needed to metabolize lactose; when starved of glucose, but supplied with lactose, they switch this machinery on. This switch mechanism—the "*lac* operon"—was historically the first genetic regulatory system to be discovered. Physically, the mechanism involves the binding of a regulatory protein, called LacI, to a specific sequence of DNA (the "operator") situated near the beginning of the set of genes coding for the lactose metabolism enzymes. Some recent

reviews of the *lac* system include Refs. [1–4]; see also Ref. [5] for looping in the *lambda* system.

Long after the discovery of genetic switching, it was found that some regulatory proteins, including LacI, exist in multimeric forms with two binding heads for DNA, and that their normal operation involves binding both sites to distant operators, forming a loop [6–11]. The looping mechanism seems to confer advantages in terms of function [12]. From the biophysical perspective, it is remarkable that in some cases loop formation, and its associated gene repression, proceed *in vivo* even when the distance between operators is much less than a persistence length of DNA [13]. For this and other reasons, a number of experimental methods have been brought to bear on reproducing DNA looping *in vitro*, to minimize the effects of unknown factors and focus on the one process of interest. Reconstituting DNA looping behavior in this way is an important step in clarifying the mechanism of gene regulation.

Tethered particle motion (TPM) is an attractive technique for this purpose [14]. In this method, a long DNA construct is prepared with two (or more) operator sequences at a desired spacing near the middle. One end is anchored to a wall, and the other to an otherwise free, optically visible bead. The bead motion is passively monitored, typically by tracking microscopy, and used as an indirect reporter of conformational changes in the DNA, including loop formation and breakdown (Fig. 1).

1.2. Goals of this paper

The recent surge of interest in DNA looping motivated us to ask: Can we understand TPM data quantitatively, starting from simple models of DNA elasticity? What is the simplest model that captures the main trends? How well can we predict data from TPM experiments, using *no fitting parameters?*

To answer such questions, we had to combine and improve a number of existing calculation tools. This paper explains how to obtain a simple elastic-rod model for DNA,
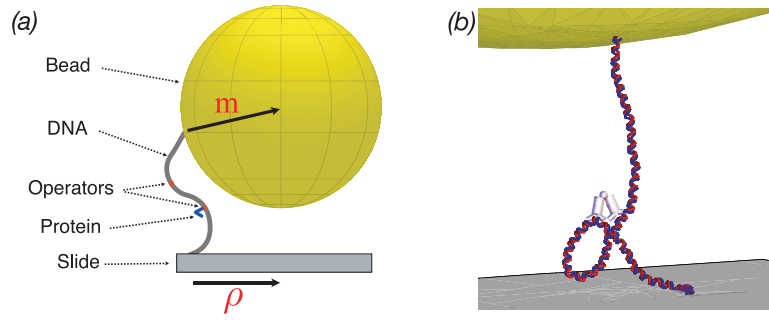
Figure 1. (*a*) Cartoon of a DNA molecule flexibly linking a bead to a surface via freely pivoting attachments (not to scale). The motion of the bead's center is observed and tracked, for example as described in Ref. [15]. In each video frame, the position vector, usually projected to the xy plane, is found. After drift subtraction, the mean of this position vector defines the anchoring point. The projected distance from this anchoring point to the instantaneous bead center is the bead excursion ρ. A regulatory protein, for example a LacI tetramer, is shown bound to a specific "operator" site on the DNA. (*b*) The conformational change of interest to us is loop formation: A loop forms when the repressor also binds to a second operator. The figure shows an actual representative looped configuration from the simulations described in this paper, drawn to scale. Figs. 2 and 6 explain the graphical representations of DNA and LacI used here.

and a geometric characterization of the repressor–DNA complex, from existing (non-TPM) experiments. From this starting point, with no additional fitting parameters, we show how to calculate experimentally observable quantities of TPM experiments (such as the fraction of time spent in various looped states and the distribution of bead excursions), as functions of experimentally controlled parameters (operator separation and repressor concentration), and compare to recent experiments.

Although our main interest is TPM experiments, our method is more generally applicable. Thus as a secondary project, we also compute looping J factors for a DNA construct with no bead or wall ("pure looping"). This situation is closer to the one that prevails *in vivo*; although in that case many other uncertainties enter, it is nevertheless interesting to

compare our results to the experimental data.

1.3. Assumptions, methods, and results of this paper

Supplementary Information Sect. S1 gives a summary of the notation used in this paper. Some readers may wish to skip to Sect. 1.3.3, where we summarize our results. Sect. 1.3.4 gives an outline of the main text and the supplement; in addition, the other subsections of this introduction give forward references showing where certain key material can be found.

1.3.1. Outline of assumptions First we summarize key assumptions and simplifications made in our analysis. Some will be justified in the main text, whereas others are taken in the spirit of seeking the model that is "as simple as possible, but not more so."

All our results are obtained using equilibrium statistical mechanics; we make no attempt to obtain rate constants, although these are experimentally available from TPM data [14, 16–18]. Our model treats DNA as a homogeneous, helical, elastic body, described by a 3×3 elastic compliance matrix (discussed in Sect. 3). Thus we neglect, for now, the effect of DNA sequence information [19], so our results may be compared only to experiments done with random-sequence DNA constructs. Despite this reduction, our model is more realistic than ones that have previously been used for TPM theory; for example, we include the substantial bend anisotropy, and twist–bend coupling, of DNA elasticity. We also neglect long-range electrostatic interactions (as is appropriate at the high salt conditions in the experiments we study), assuming that electrostatic effects can be summarized in effective values of the elastic compliances.

The presence of a large reporter bead at one end of the DNA construct, and a wall at the other end, significantly perturb looping in TPM experiments. We treat the bead as a sphere, the wall as a plane, and the steric exclusion between them as a hard-wall interaction. We neglect nonspecific DNA–protein interactions ("wrapping" [20]).

1.3.2. Outline of methods   Our method builds on prior work [15, 21]. Sect. 7 discusses other theoretical approaches in the literature.

Our calculations must include the effects of chain entropy on loop formation, because we consider loop lengths as large as 510 basepairs. We must also account for entropic-force effects created by the large bead at one end of the DNA and the wall at the other end, in addition to the specific orientation constraints imposed on the two operators by the repressor protein complex. To our knowledge such a complete, first-principles approach to calculating DNA looping for tethered particle motion has not previously been attempted. In part because of these complications, we chose to calculate using a Monte Carlo method called "Gaussian sampling" (discussed in Sect. 4 and Sects. S6–S7). Gaussian sampling is distinguished from Markov-chain methods (e.g., Metropolis Monte Carlo) in that successive sampled chains are independent of their predecessors.

We must also address a number of points before we can compare our results to experiments. For example, DNA simulations report a quantity called the "looping J factor." But TPM experiments instead report the time spent in looped versus unlooped states, which depends on both J and a binding constant K_d. We present a method to extract both J and K_d separately from TPM data (discussed in Sect. S5). We also describe two new data-analysis tools: *(1)* A correction to our theoretical results on bead excursion, needed to account for the effect of finite camera shutter time on the experimental results (discussed in Sect. S2), and *(2)* another correction needed to make contact with a widely used statistic, the finite-sample RMS bead excursion (discussed in Sect. S3). (To be precise, the latter two corrections do both involve phenomenological parameters, but we obtain these from TPM data that are different from the ones we are seeking to explain. Each correction could in any case be avoided by taking the experimental data differently, as described in the Supplementary Information.)

1.3.3. Outline of results   Some of our results were first outlined in Refs. [22, 23]. The assumptions sketched above amount to a highly reductionist approach to looping. Moreover,

we have given ourselves no freedom to tweak the model with adjustable parameters, other than the few obtained from non-TPM experiments (four elastic constants and the geometry of the repressor tetramer); all other parameters we used had known values (e.g., bead size and details of the DNA construct). So it is not surprising that some of our results are only in qualitative agreement with experiment. Nevertheless, we find that:

- Our physical model quantitatively predicts basic aspects of the TPM experiments, such as the effects of varying tether length and bead size (see Fig. 3).

- The model can roughly explain the overall value of the looping J factor obtained in experiments for a range of loop lengths near 300 basepairs (discussed in Sect. 5).

- Perhaps most surprising, the same simple model predicts rather well the observed, detailed structure of the *distribution* of bead excursions, including its dependence on loop lengths near 300 basepairs (see Fig. 12). The distinctive three-peaks structure of this distribution [24–26] has sometimes been taken as *prima facie* evidence for a hypothetical alternate "open" conformation of the repressor protein. But we show that it can also arise without that hypothesis, as a consequence of the contributions of loops with different topologies.

- Notwithstanding those successes, our simple model does *not* successfully extrapolate to predict the magnitude of the J factor for loop lengths near 100 basepairs, at least according to the limited, preliminary experimental data now available. Instead, there it underestimates J, pointing to a breakdown of some of its hypotheses in this high-strain situation. Perhaps the needed modification is a nonlinear elastic theory of DNA [27, 28], significant flexibility in the tetramer, additional nonspecific binding of DNA to the repressor protein, or some combination of these.

- However, our model does give a reasonable account of the structure of the bead excursion distribution even for loop lengths near $100\,\mathrm{bp}$ (see Fig. 13).

- Because previous authors have proposed the specific hypothesis that one of the excursion-distribution peaks reflects an "open" conformation of LacI, we simulated that situation as well. We argue that this hypothesis cannot by itself explain the high degree of looping observed experimentally for short DNA constructs (discussed in Sect. 5.4.3).

Our calculations also quantify the importance of the orientation constraint for binding to the tetramer, via a concept we call the "differential J factor" (discussed in Sect. 5.2). Finally,

our simple model of blur correction quantitatively predicts the observed dependence of apparent bead motion on camera shutter time, and we expect it will be useful for future TPM experiments (discussed in Sect. S2.2).

1.3.4. Organization of this paper Sect. 2 gives an overview of various single-molecule experiments used recently to study looping, emphasizing the particular capabilities of TPM. Sect. 3 derives the elastic model of DNA to be used in this paper. Sect. 4 introduces our Monte Carlo method, and gives a crucial check that theory and experiment are both working properly, by showing to what extent we can accurately predict the excursion of the tethered bead in the absence of looping. Sect. 5 shows how to extend the simulation to study looping, defines the looping J factor, and gives results on J as a function of loop length, both with and without the effect of the tethered bead and surface, and for both the closed (V-shaped) and hypothesized open conformation of the lac repressor tetramer. Sect. 6 gives a more refined measure of bead motion, the probability distribution of the bead excursions. Sect. 7 discusses the relation between our work and earlier theoretical papers, and finally Sect. 8 gives general discussion.

The Supplementary Information has its own table of contents; it contains information more directly related to the experimental data, details of our Monte Carlo algorithm, and some additional calculations in our model. For example, we checked our work by calculating cyclization J factors and comparing to the classic Shimada–Yamakawa result.

2. Survey of experiments on looping

Experimental measurements of DNA loop formation have fallen into four main classes. Readers familiar with the experiments may wish to skip directly to Sect. 3.

Cyclization In these *in vitro* experiments, many identical, linear DNA constructs are prepared with overhanging, complementary ends. Ligase enzyme captures transient states

in which either two ends of the same DNA join, forming a ring, or else ends of two different DNAs join, forming a dimer. Under suitable conditions the ratio of rings to dimers after the reaction runs to completion gives information about the equilibrium populations of those paired states, and hence about loop formation (e.g., [29–34]). Unfortunately, the interpretation of these experiments is complicated by the role of the large, complex ligase enzyme, the need to be in a very specific kinetic regime, and so on [35]. Moreover, the process of interest to gene regulation is looping, which is geometrically quite different from cyclization.

In vivo repression Other experiments measured the output of an operon as its controlling promoter was switched by a repressor (e.g., [13,36–39]); theory then connects those results to looping J factor values (or looping free energy changes) [40–43]. Although the experiments showed that short loops form surprisingly easily, their quantitative interpretation is obscured by uncertainties due to the complex world inside a living cell, for example, supercoiling and the many other DNA-binding proteins (such as HU, H-NS and IHF) present in cells.

Magnetic tweezer To introduce supercoiling in an *in vitro* preparation, some experiments manipulate the DNA using a magnetic bead in a trap. Some earlier implementations unavoidably also introduced extensional stress on the DNA [44]; however, recent work has overcome this limitation [25].

Tethered particle In the present work we study TPM experiments [45], which can report directly on looping state under controlled, *in vitro* conditions. Recent work on looping via TPM includes Refs. [16–18, 26, 46–48]. TPM experiments do require significant analysis to determine looping state from bead motion, but techniques such as dead-time correction [16] and Hidden Markov modeling [17, 18] now exist to handle this. Like cyclization, the TPM experiments we studied have the biologically unrealistic (but theoretically convenient) feature that the supercoiling stress applied externally to the loop is zero. (For a theoretical approach

to looping with supercoiling see e.g., Ref. [49].)

Additional advantages of TPM include the fact that it does not involve fluorescence, and so is not subject to bleaching; thus an experiment can generate an unlimited data sample simply by tracking a bead for a long time. Moreover, the DNA is in solution, and minimally affected by the distant bead. Some implementations of TPM do not track individual trajectories, instead observing the blurred average image of each bead [14, 24]; this article will focus on particle-tracking implementations (see, e.g., Refs. [15, 50]). Other experimental aspects, including the attachment of the DNA of interest to the mobile bead at one end and the immobile surface at the other, are discussed in the original articles cited above.

TPM experiments also offer the ability to separate the overall probability of looping, at least partially, into the contributions of individual loop types (see Sect. 6). This additional degree of resolution allows more detailed comparison with experiment than is possible when we observe only the level of gene repression. Finally, TPM and other *in vitro* methods also present the opportunity to dissect the experimentally observed looping probability into separate numerical values for the looping J factor and the binding constant, via a titration curve (discussed in Sect. S5). In contrast, some *in vivo* methods must obtain a value for the binding constant from a single data point (repression with auxiliary operator deleted), and moreover must rely on the accuracy of an estimate for the effective repressor concentration in the cell [42].

3. Elasticity theory used in this paper

This section derives the elastic model of DNA to be used in this paper. Sect. 3.2 first obtains the elasticity matrix up to an overall constant from structural information; then Sect. 3.3 fixes the constant by requiring a particular value for the persistence length. Our simulation method involves matrix exponentiation, and may be simpler than other methods sometimes used in the literature.

3.1. General framework

The physical model of DNA as a uniform, isotropic, slender, linearly-elastic rod [51] has proven to give an adequate description of DNA mechanics for some purposes, notably for computing the force–extension relation of long DNA [52, 53]. However, this simple model is not obviously appropriate for describing the formation of structures involving DNA loops of length comparable to a helical repeat ($\ell_{\mathrm{helix}} = 3.5\,\mathrm{nm}$). For example, in this paper we are interested in loops as short as 9 times the helical repeat length. On length scales comparable to ℓ_{helix}, the bend stiffness anisotropy of the molecule certainly becomes significant, as well as elastic cross-coupling between bend and twist [54, 55]. Sect. 3.2 below spells out the details of the elasticity theory we will use. (Sect. S8.1 explores the importance of including the anisotropy by studying an alternative model.)

In other respects, our elastic model will be standard. We assume that the unstressed state of DNA may be regarded as a stack of plates ("segments"), each with thickness ℓ_0 and each with a chosen reference point and an inscribed coordinate frame at that point (Fig. 2). Each plate is shifted a distance ℓ_0 along its $\hat{\mathbf{E}}_3$-axis relative to its predecessor, and also rotated by $2\pi\ell_0/(\ell_{\mathrm{helix}})$ about the same axis. Next we need to quantify the elastic energy cost for a deviation from this unstressed state.

We restrict attention to a harmonic elasticity model, that is, we assume that the elastic energy at each junction is a quadratic function of bend and excess twist, neglecting the possibility of elastic breakdown at high strain [27, 28, 32]. We do this because ultimately we are interested in *testing* the harmonic model, by confronting its predictions with experiment, and also because there is not yet a unique candidate for the detailed, three-dimensional form of an effective nonlinear elastic function.

We neglect stretch elasticity of the segments because there is no externally applied stretching force in TPM experiments (and any entropic stretching force is insignificant in this context [21]). Thus the displacement of each segment is always ℓ_0; the "pose"
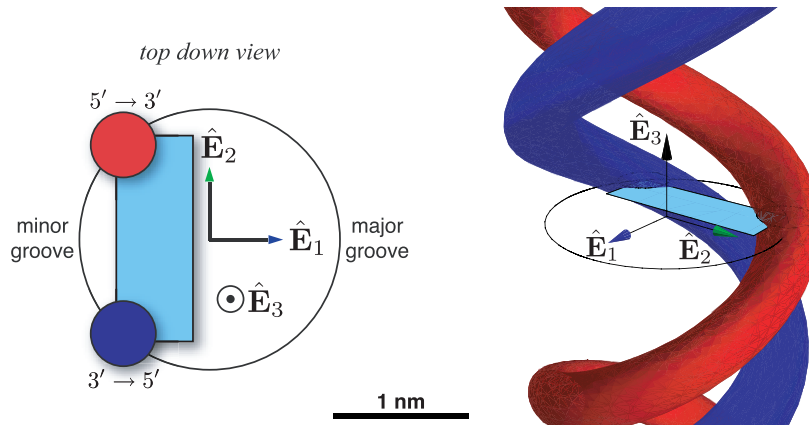
Figure 2. Basepair geometry [54]. The rectangle represents a DNA basepair. The red and blue dots are the phosphate backbones. The circle is the outer envelope of the double helix, $2\,\mathrm{nm}$ in diameter. We set up an orthonormal frame (*left*) where $\hat{\mathbf{E}}_3$ is out of the page, $\hat{\mathbf{E}}_1$ points to the major groove, and $\hat{\mathbf{E}}_2$ completes the triad and points towards the $5' \to 3'$ strand (red) as defined by the positive $\hat{\mathbf{E}}_3$ direction. "Positive roll" is then defined as a positive rotation about $\hat{\mathbf{E}}_2$ as we pass from this basepair to the one on top of it (="bend into the major groove"). Similarly "tilt" is rotation about $\hat{\mathbf{E}}_1$, and "twist" is excess rotation about $\hat{\mathbf{E}}_3$ (in addition to the natural helical twist). For our purposes, a DNA chain conformation is a sequence of such frames. Graphically we represent it in Figs. 1 and 6 as a chain of double-helical segments, as shown on the *right*.

(position and orientation) of each segment relative to its predecessor is completely specified by the angular orientation. For simplicity, we also neglect the sequence-dependence of DNA elasticity, so our results will apply only to random-sequence DNA constructs; all our comparisons to experiments will involve DNA of this type. Because we are making a finite-element approximation to a continuum elasticity model, we have some freedom in choosing the contour length ℓ_0 of each segment, as long as it is much shorter than the persistence length, about 150 basepairs. To speed up calculations, we have chosen a segment length corresponding to one fifth of a helical repeat (about 2 basepairs). Making our segments commensurate with the helical repeat also has the advantage of showing clearly any helical

phasing effects, i.e., modulation of looping with period equal to ℓ_{helix}.

Let $\Delta\theta_i$ be the excess rotation angles (beyond the natural twist) from one segment to the next and let $\Omega_i = \Delta\theta_i/\ell_0$ denote the corresponding strain rates per unit contour length, where $i = 1, 2, 3$ correspond to tilt, roll, and twist (see Fig. 2). We will define the elastic deformation free energy per unit contour length as

$$\mathcal{E} \equiv (\tfrac{1}{2}k_{\text{B}}T)\Omega^t\mathsf{Q}\Omega, \tag{1}$$

so the stiffness matrix Q has units of length and is independent of the choice of segment length ℓ_0. (The compliance matrix is then Q^{-1}.) In the traditional wormlike chain model Q is diagonal, with the bend and twist persistence lengths on the diagonal. We next propose a more realistic choice for this matrix.

3.2. Relative elastic constants

To get values for the elements of Q, we first note that (neglecting sequence-dependence) B-form DNA has a symmetry under $180°$ rotation about any line perpendicular to its long axis and passing through its major groove. (Such a line is labeled $\hat{\mathsf{E}}_1$ in Fig. 2.) This symmetry forbids any harmonic-elasticity coupling between twist and tilt (that is, between small rotations about $\hat{\mathsf{E}}_3$ and $\hat{\mathsf{E}}_1$ in the figure), and also between tilt and roll [54]. Thus the symmetric 3×3 matrix Q has only four independent nonzero entries [54, 56].

Next, we adapt a strategy used by W. Olson and coworkers [57], who examined crystal structures of many DNA oligomers and of DNA–protein complexes. They then supposed that each basepair is subjected to random external forces (e.g., crystal forces), the same for every type of basepair junction, analogous to the random forces in thermal equilibrium but of an unknown overall magnitude. The observed deformations of basepairs in this imagined random external force tell us about the elastic compliances for deformation of each basepair type, and in particular the covariances of deformations give the off-diagonal terms. Finally, we adjust the overall scale of the resulting elastic-energy matrix to obtain the desired persistence

length of DNA in the buffer conditions appropriate to the TPM experiments of interest.

The method outlined above, although rough, nevertheless captures the basic structure of DNA elasticity while preserving the required overall persistence length. To carry it out, we took the published covariance matrices for the $\Delta\theta_i$ of various basepair steps [57] and averaged them to obtain an elastic compliance matrix. We inverted this matrix and observed that indeed the (12), (13), (21), (31) entries of Q were much smaller than the others; we subsequently set them to be exactly zero. These steps yielded the entries of Q, up to an overall scale factor, as

$$\mathsf{Q} = \gamma \times \begin{bmatrix} 0.084 & 0 & 0 \\ 0 & 0.046 & 0.016 \\ 0 & 0.016 & 0.047 \end{bmatrix}. \tag{2}$$

The overall constant γ has units of length; it will be specified in Sect. 3.3.

The expected anisotropy is evident in the form of the matrix: The tilt eigenvalue (0.084) is much larger than the smaller of the two remaining eigenvalues (0.030). Note that the near-degeneracy of the last two diagonal elements means that the eigenvectors are strongly mixed: The smaller eigenvalue corresponds to a mixed deformation, with positive roll and negative twist. Thus, bending the DNA tends to untwist it [57]. Note, too, that the numerical values of the diagonal entries are not a good guide to the relative actual bend stiffnesses, because the eigenvalues of the 2×2 submatrix may be quite different from its diagonal entries.

3.3. Specification of overall scale factor

The persistence length ξ of a polymer is defined by the falloff in correlation between the long-axis directions of nearby elements when the polymer is free (no external forces). Thus $\langle \hat{\mathbf{E}}_3(s) \cdot \hat{\mathbf{E}}_3(s+t) \rangle \to \mathrm{e}^{-|t|/\xi}$ at large t, where s, t are contour lengths [52]. We now discuss how to compute ξ for an elastic matrix of the form Eqn. (2), as a function of the unknown parameter γ that sets the strength of Q; demanding a particular value of ξ will then fix the value of γ. (A similar discussion recently appeared in [58].)

To compute ξ given a choice of γ, we first generate a string of random rotation matrices, each representing relative rotations of one segment relative to its predecessor. These matrices are drawn from a distribution that is centered on the identity matrix and weighted by the Boltzmann factor $\mathrm{e}^{-\mathcal{E}\ell_0/k_\mathrm{B}T}$. More explicitly, we choose a value of ℓ_0, then diagonalize the matrix Q/ℓ_0, writing it as $\mathsf{T}^t\mathsf{D}\mathsf{T}$ for an orthogonal matrix T. We then use the diagonal entries of D as inverse variances for three Gaussian random variables $\{\Psi_i\}$, and let $\Delta\theta = \mathsf{T}^t\Psi$, obtaining three random variables $\Delta\theta_i$ with the desired statistical properties. We convert these random angles into a rotation matrix by computing the matrix exponential $\exp(\sum_{i=1}^{3}\Delta\theta_i\mathsf{J}_i)$, where J_i are the rotation generator matrices. (For example, $\mathsf{J}_3 = \left[\begin{smallmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix}\right]$.) Finally, we multiply the resulting rotation on the left by the natural, unstressed DNA rotation $\exp((2\pi\ell_0/\ell_\mathrm{helix})\mathsf{J}_3)$, obtaining $\mathsf{R}(1)$, then repeat all these steps to make a long string of matrices $\mathsf{R}(1),\ \mathsf{R}(2),\cdots$.

Next, we step through the matrix string, cumulatively applying each rotation $\mathsf{R}(k)$ in turn to an initial orientation to obtain the orientations of successive basepairs from a standard orientation for the first one. That is, let the frame vector at arclength position s be $\hat{\mathbf{E}}_a(s)$. We express it in components using the fixed lab frame as $[\hat{\mathbf{E}}_a(s)]_i = \mathsf{h}_{ia}(k)$, $i = 1, 2, 3$, where $s = k\ell_0$, $\mathsf{h}_{ia}(0) = \delta_{ia}$, and $\mathsf{h}(k+1) = \mathsf{h}(k)\mathsf{R}(k)$. Finally we average the quantity $\langle \hat{\mathbf{E}}_3(s) \cdot \hat{\mathbf{E}}_3(s+t) \rangle$ over the generated chains, average over s for various fixed t, confirm the exponential decay in t, and extract the decay length ξ.

In solvent conditions used for TPM by Han et al. [26,48], the persistence length has been previously measured by other means to be around $44\,\mathrm{nm}$ [59,60]; see also Sect. 4.2, where we show that this value is consistent with TPM calibration data. Applying the above procedure to Eqn. (2) and requiring $\xi = 44\,\mathrm{nm}$ fixes γ: We then have

$$\mathsf{Q} = \begin{bmatrix} 67\,\mathrm{nm} & 0 & 0 \\ 0 & 37\,\mathrm{nm} & 13.0\,\mathrm{nm} \\ 0 & 13.0\,\mathrm{nm} & 37\,\mathrm{nm} \end{bmatrix}. \tag{3}$$

Eqn. (3) is the form suitable for angles $\Delta\theta$ expressed in radians; for angles in degrees the

matrix should be multiplied by $(\pi/180)^2$.

4. Calculation of TPM distributions without looping

This section introduces our Monte Carlo method, and gives a crucial check that theory and experiment are both working properly, by showing to what extent we can accurately predict the excursion of the tethered bead in the absence of looping. Some details relevant to experimental data (blur correction and finite-sample effects) are relegated to the Supplement.

We begin our analysis by predicting the motion of a tethered particle in terms of the tether length and bead size, both of which were systematically varied in the experiments of Ref. [48]. Besides being a basic polymer science question, such *a priori* knowledge of, say, the root-mean-square bead excursion for simple tethers sets the stage for our calculations involving looped tethers in Sect. 6. More generally, in other kinds of experiments the tether length may be changing in time, in a way that we would like to measure, as a processive enzyme walks along DNA or RNA [61], or as proteins bind to the DNA, etc. Finally, by comparing theory to experiment, we gain confidence both that the experiment is working as desired and that our underlying assumptions about the polymer mechanics, bead–wall interactions, and so on, are adequate.

Although the end-end distribution of a semiflexible polymer such as DNA is a classical problem in polymer physics, the present problem differs from that one in several respects. For example, the DNA is not isolated, but instead is attached to a planar surface, and hence experiences an effective entropic stretching force due to the steric exclusion from half of space; a similar effective repulsion exists between the DNA and the large bead. More important than these effects, however, is the steric exclusion of the bead from the wall. Ref. [21] argued that the effect of this exclusion would be to create an entropic stretching force on the DNA.

Additional subtleties of the problem include the fact that the polymer itself has two

additional length scales in addition to the bead radius, namely its persistence length ξ and total length L, and the fact that we do not observe the polymer endpoint, but rather the center of the attached bead. Some of these effects have been studied analytically for the case with applied stretching force (e.g., [62]), but for *zero* applied stretching force the steric constraints, not fully treatable in that formalism, become important. For this reason, Refs. [15, 21] developed a Monte Carlo calculation method.‡ A similar method was independently used for a study of DNA cyclization by Czapla et al. [65], who call it "Gaussian sampling." Here we generalize that method to use the elasticity theory described in Sect. 3. We also extend our earlier work by computing the dependence of the RMS bead excursion on both tether length and bead size, and comparing to experimental data in which both were systematically varied.

4.1. Gaussian sampling

The Gaussian sampling approach is not a Markov-chain algorithm; each chain is generated independently of all the others, in the Boltzmann distribution associated with the elastic energy function. What makes this approach feasible is that the elastic energy functions of each junction between links are all independent (because we assume that there is no cooperativity between basepairs separated by more than our segment length ℓ_0). Thus, the random bends between links are also independent; we generate a chain by creating a string of rotation matrices each generated as described in Sect. 3.3. To implement the steric constraints, we next suppose additional energy terms of hard-wall type (i.e. either zero or infinity). Although it is an approximation to real mesoscopic force functions, the hard-wall approximation is reasonable in the high salt conditions studied in typical experiments. Together with the approximate representation of a real microscope slide as a perfect plane (a "wall"), it has proven successful in our earlier work [15].

The constraint energy terms set the probability of the sterically-forbidden chains to zero.

‡ Refs. [63, 64] studied related spatial-constraint effects in an analytical formalism; the present paper gives a numerical approach.

In practice, then, we generate many chains, find each chain segment's spatial position (and that of the bead) by following the $\hat{\mathbf{E}}_3$-axis of each orientation triad, and discard the chain if any steric constraint is violated. All our thermodynamic averages are then taken over the remaining ("allowed") chains. For short tethers, many chains will be discarded, but as long as the fraction of "allowed" chains is not too small the procedure is tractable.

We treat the biotin and digoxigenin linkages attaching the DNA to bead and wall as freely flexible pivots, and so the orientation of the first chain segment, and that of the bead relative to the last segment, are taken to be uniformly distributed in the half-spaces allowed by the respective surfaces. This approach has previously been successful in explaining experimental results [15,21,62,66]. That is, the initial chain segment's orientation is a uniformly distributed random rotation subject to the half-space constraint; subsequent segments are then determined by successive matrix multiplication by the rotations distributed as in Sect. 3.3; the final vector $\mathbf{m}$ describing the bead orientation relative to its attachment point (black arrow in Fig. 1*a*) is again taken to be uniformly distributed in the half-space defined by the final chain segment.

The steric constraints we implemented were *(i)* chain–wall, *(ii)* chain–bead, and *(iii)* bead–wall exclusion. For the short DNA tethers we consider, chain–chain excluded volume is not expected to be a significant effect (although it would be important if supercoiling stress were applied to the bead [56]).

We can see the trends in the data more clearly if we reduce the distribution of bead position to the root-mean-square excursion $\rho_{\mathrm{RMS}} \equiv \sqrt{\langle \rho^2 \rangle}$, a quantity often used in experiments to characterize tethered particle motion. A closely related quantity is the finite-sample RMS excursion, for example $\rho_{\mathrm{RMS,\,4\,s}} \equiv \sqrt{\langle \rho^2 \rangle_{4\,\mathrm{s}}}$. Here the expectation value is limited to a sample consisting of $(4\,\mathrm{s})/(0.03\,\mathrm{s})$ consecutive video frames at a frame rate of $1/(0.03\,\mathrm{s})$. Note that whereas ρ_{RMS} is a single number for each bead-tether combination, in contrast $\rho_{\mathrm{RMS,\,4\,s}}$ has a probability distribution. One of our goals in the remainder of this paper is to predict ρ_{RMS} (in this section), or the distribution of $\rho_{\mathrm{RMS,\,4\,s}}$ (in Sect. 6), as functions of
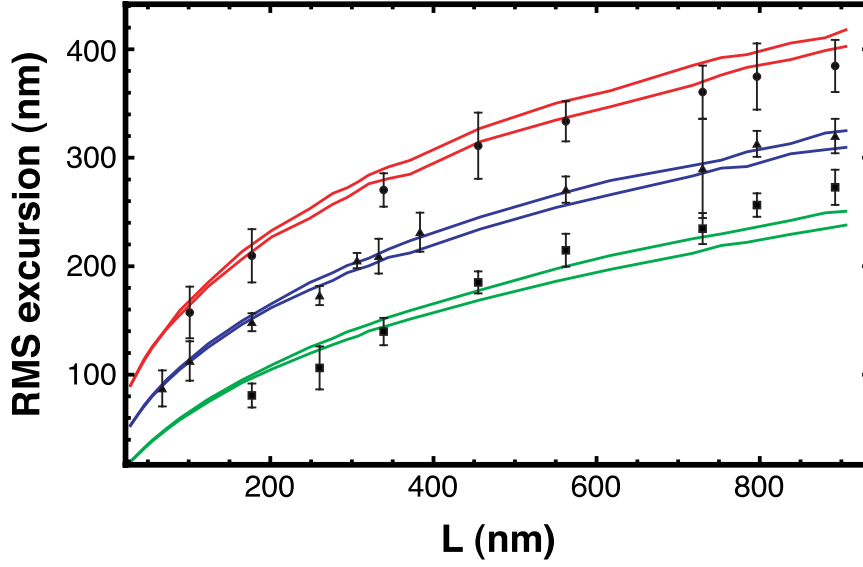
Figure 3. Theoretical prediction of equilibrium bead excursion. *Dots:* Experimental values for RMS excursion of bead center, $\rho_{\mathrm{RMS},t}$, for random-sequence DNA and three different bead sizes: Top to bottom, $R_{\mathrm{bead}} = 485$, 245, and $100\,\mathrm{nm}$. (Data from Ref. [48].) The sampling times were $t = 20$, 10, and $5\,\mathrm{s}$ respectively. For these rather long times the finite-sample correction is negligible; nevertheless, we included this correction (via a method given in Sect. S3). Each dot represents 20–200 different observed beads with the given tether length. Dots and their error bars were computed by the method described in Fig. S1. *Curves:* Theoretically predicted RMS motion, corrected for the blurring effect of finite shutter time. For each of the three bead sizes studied, two curves are shown. From top to bottom, each pair of curves assumes persistence length values $\xi = 47$ and $39\,\mathrm{nm}$, respectively. There are *no fit parameters;* the theoretical model uses values for bead diameter given by the manufacturer's specification.

bead size, tether length, and tether looping state.

4.2. Calibration curve results

Sect. 4.1 explained how, given values of L, R_{bead}, and ξ, we generate many chain/bead configurations. From these configurations, we can in principle compute quantities like

ρ_{RMS}. (An additional correction, to account for finite camera shutter speed, is explained in Sect. S2.2.) We compute $\rho_{\mathrm{RMS},t}$ in this way and compare it to the experiments of Han et al. [48]. We took L to be $0.34\,\mathrm{nm}$ times the number of basepairs in each construct, and accepted the manufacturer's specifications of R_{bead} for beads of three different sizes, leaving us with just one remaining parameter, the persistence length ξ. The finite sampling times used in the experiment had an insignificant effect (data not shown), but nevertheless we included this aspect of the experiment (see Sect. S3) for consistency with our later study of the probability density function of bead excursion in Sect. 6. In that context, the finite sampling time is important.

DNA stretching experiments using high-salt buffer similar to that used in the TPM experiments we study obtained a persistence length of $\xi = 45\,\mathrm{nm}$ [59], or $43\,\mathrm{nm}$ [60]. When we turn to TPM, Fig. 3 shows that indeed taking ξ in the range 39–$47\,\mathrm{nm}$ reproduces the trends of the data fairly well with *no fitting*, even though this is a very different class of experiment from stretching. (Previous work came to a similar conclusion [15], although it considered only a single bead size.) The curve with bead size $245\,\mathrm{nm}$ is particularly well predicted; all TPM data appearing in the rest of this paper were taken with this value of R_{bead}. Throughout the rest of this paper we will use the value $\xi = 44\,\mathrm{nm}$.

5. DNA looping

This section attempts to distill loop formation into a mathematical problem, the calculation of a quantity called the "looping J factor" (Sects. 5.1–5.2; some geometrical details about the looping synapse are deferred to Sect. S4). Sect. S5 in the Supplement explains how we extracted J from experimental data. Next, Sect. 5.3 describes the calculation of J (more details are in Sects. S6–S7) and Sect. 5.4 compares to experiment. For loops of length near $300\,\mathrm{bp}$, our absolute prediction for J agrees with the preliminary experimental data now available to within about a factor of 3; equivalently the corresponding looping free energies

agree to within about $1\,k_{\mathrm{B}}T$. However, the hypotheses embodied in our model cannot explain the observed J factor for short loops, near $95\,\mathrm{bp}$ between operators. We will argue that the hypothesis of an alternate "open" LacI conformation is not sufficient to resolve this discrepancy.

5.1. Geometric structure of the loop complex

5.1.1. DNA construct The experiments of Han et al. [26] studied DNA looping for random-sequence DNA in two classes, forming "long" and "short" loops. (They also studied special sequences [67], which we do not discuss in the present paper.) Both "long" and "short" loop DNA constructs had the general form

$$\texttt{wall-}(N_1\,\mathrm{bp})\texttt{-}(N_2\,\mathrm{bp})\texttt{-}(N_3\,\mathrm{bp})\texttt{-}(N_4\,\mathrm{bp})\texttt{-}(N_5\,\mathrm{bp})\texttt{-bead} \ . \tag{4}$$

The "short" constructs had $N_2 = 20\,\mathrm{bp}$ (the $\mathrm{O_{id}}$ operator), $N_4 = 21\,\mathrm{bp}$ (the $\mathrm{O_1}$ operator), and $N_1 = 144\,\mathrm{bp}$, $N_3 = 89 + I\,\mathrm{bp}$, $N_5 = 171\,\mathrm{bp}$, where I is an integer equal to 0, 5, or 11. The "long" constructs had $N_2 = 21\,\mathrm{bp}$ ($\mathrm{O_1}$), $N_4 = 20\,\mathrm{bp}$ ($\mathrm{O_{id}}$), and $N_1 = 427\,\mathrm{bp}$, $N_3 = 300 + I\,\mathrm{bp}$, $N_5 = 132 - I\,\mathrm{bp}$, where I is an integer between 0 and 10.§ For the purpose of labeling loop topologies, we choose a conventional direction along the DNA that runs from $\mathrm{O_{id}}$ to $\mathrm{O_1}$. Thus for the "short" constructs this direction runs from the wall to the bead, whereas for the "long" constructs it runs from bead to wall.

The artificial sequence $\mathrm{O_{id}}$ ("ideal operator") binds DNA more strongly than the wild type $\mathrm{O_1}$. In fact, in the range of $[\mathrm{LacI}]$ values we study, $\mathrm{O_{id}}$ is essentially always bound [26], and the looping transition consists of binding/unbinding of the already-bound LacI to $\mathrm{O_1}$.

§ Some of the actual constructs used in the experiment differed from the simple formula above by 1–2 basepairs [26].
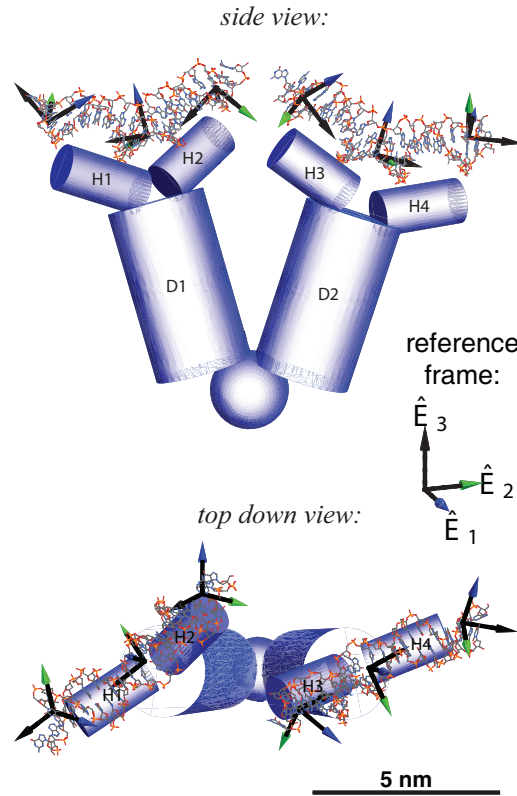
side view:

top down view:

5 nm

Figure 4. Cartoon of the LacI tetramer (solid shapes) bound to two operator DNA segments (shown as wireframes). The tetramer consists of dimers D1 and D2, with binding heads H1–H4. The wireframes show in detail the dispositions of the operators relative to each other, as given in Protein Data Bank entry 1LBG.pdb. In the present work we summarize the entire structure by the six orthonormal frames shown, which represent the *entry/exit* and *center* frames discussed in the main text and Sect. S4. The axes with blue, green, and black arrowheads represent $\hat{\mathbf{E}}_1$, $\hat{\mathbf{E}}_2$, and $\hat{\mathbf{E}}_3$ respectively. These six frames were determined from the PDB file by the method described in Sect. S4.

5.1.2. DNA binding and its degeneracy The LacI protein is a tetramer consisting of two identical dimers (D1, D2), each with two heads (H1–H4) that bind the DNA.‖ Fig. 4 shows a cartoon, drawn to scale, based on the RCSB Protein Data Bank entry 1LBG.pdb [70] (see also [71–73]). Two segments of bound DNA (operators of type O_{id}) appear as well.

‖ Lac repressor essentially always exists as tetramers under the conditions of the experiments studied here [68, 69].

The cartoon is meant to portray the level of detail with which we treat the tetramer in our calculations: We regard the protein as a clamp holding the two bound operators rigid relative to each other. Thus, as soon as we specify the pose (position and orientation) of the DNA bound to head H1 (say), we have also specified its exit from H2 as well as its entry and exit at H3 and H4. Fig. 4 shows six particular poses, represented by orthonormal triads, associated to the *entry*/*exit* and *center* basepairs. These are described in greater detail below and in Sect. S4. The axes are color-coded; the blue, green, and black arrows correspond to the axis vectors $\hat{\mathbf{E}}_1$, $\hat{\mathbf{E}}_2$, and $\hat{\mathbf{E}}_3$ in Fig. 2.

Actually, each binding site has *two* energetically equivalent binding orientations, due to a two-fold symmetry of the LacI dimer [4], so Fig. 4 shows only one of four possibilities. (The DNA sequence of the operator need not be a palindrome to have this degeneracy.) The symmetry operation on the DNA that relates these orientations is the same one described in Sect. 3.2: $180°$ rotation about the frame vector $\hat{\mathbf{E}}_1$ passing through the operator center and pointing to the major groove.

Referring to Eqn. (4), we will speak of the DNA as "starting" at the wall or bead, "entering" a binding site at one end of O_{id}, "exiting" that binding site to the interoperator segment, and (if looped) "then entering" the other site at O_1 and "finally exiting" to "arrive" at the bead or wall. Sect. S4 describes our mathematical characterization of the geometry of LacI for the purposes of our simulation. Here we only note that because of an approximate twofold symmetry in the tetramer, it is immaterial which dimer binds to O_{id}. However, we do need to distinguish the two binding orientations at each site, because they have inequivalent effects on the rest of the DNA. We will distinguish them at O_{id} by the label $\beta = 1, 2$. Similarly, we introduce a label $\alpha = 1, 2$ denoting the binding orientation at O_1. Fig. 5 defines our conventions for these labels, which amount to specifying four topologically distinct classes of loops.¶

¶ Each of these classes in turn can be further subdivided into distinct topoisomers. For example, we can take any of the loops shown in Fig. 5, detach the DNA from one binding head, twirl it about its axis by one full
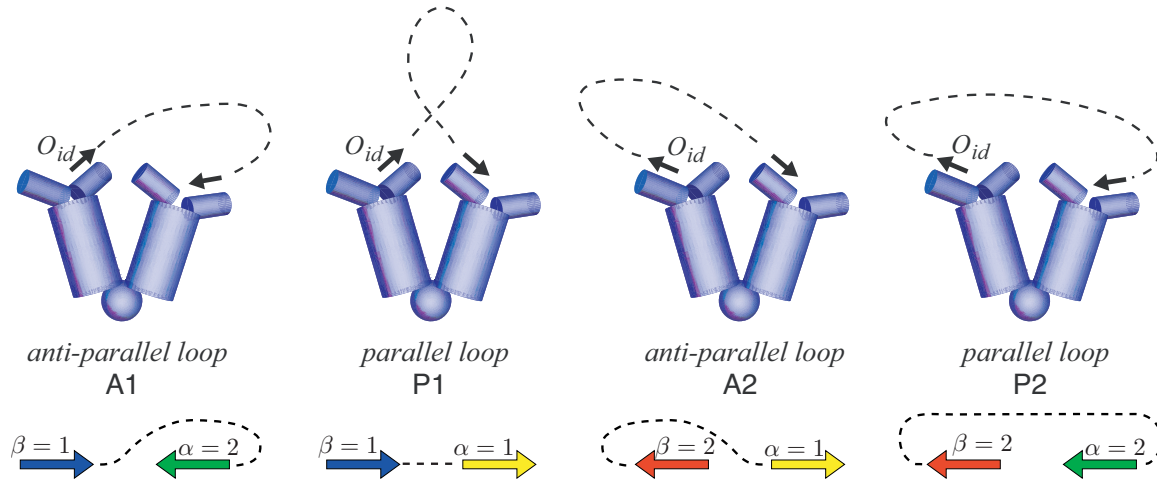
Figure 5. Four possible orientations of simulated looped chain (*dashed lines*). Our convention is that the arrows run from O_{id} to O_1. Two binary variables describe the binding orientation at the two operators as shown. If the chain exits O_{id} at an inner headgroup (H2 or H3 in Fig. 4), we say $\beta = 1$. If the chain enters O_1 at an inner headgroup, we say $\alpha = 1$. There are two "parallel" loop configurations (P1, P2), for which the entry and exit trajectories of the chain have nearly parallel $\hat{\mathbf{E}}_3$ axes; likewise, there are two "antiparallel" loop configurations (A1, A2), for which the entry and exit trajectories of the chain are nearly opposite. Configurations A1, A2 look equivalent under the symmetry that reverses DNA direction and exchanges the two LacI dimers. However, this apparent degeneracy is broken when we add the bead to one end, and the wall to the other.

Fig. 5 also identifies each looping topology using names consistent with previous LacI looping studies [74, 75]. These topologies are grouped into two general categories characterized by the relative orientation of the two bound operator sequences: parallel (P1, P2) or anti-parallel (A1, A2).

The dashed lines in Fig. 5 represent the DNA loops and are added as visual aids; they are not results of our calculations.

revolution, then reattach it, resulting in a topologically distinct loop with the same values of α and β. Because in experiments the topoisomer class of a loop can neither be observed nor controlled, however, we will not make any use of this subdivision in this paper.

5.2. The looping J factor

TPM (and some of the other experiments described in Sect. 2) provides information about the fraction $P(\text{looped})$ of time that a DNA tether spends in one of its looped conformations. Suppose that a repressor tetramer is already bound to operator O_{id}. Then we can regard the looping transition as a combination of two subprocesses, namely *(i)* the occasional spontaneous bending of the DNA to bring LacI and the other operator (O_1) into proximity, and *(ii)* binding of O_1 to LacI. The first of these processes will be characterized by a quantity called the "looping J factor" below, whereas the second is characterized by a chemical binding constant K_{d}. The looping J factor is the quantity of interest to us in this paper, as it is the one that we will subsequently attempt to predict theoretically. It is a generalization of the classical J factor from DNA cyclization [76–78], which can roughly be regarded as the concentration of one operator in the vicinity of the other. In this section we define J mathematically; Sect. S5 describes how to obtain it from TPM data. (Sect. S5 will explain the relation between J and the "looping free energy change" ΔG_{loop} discussed by other authors.) Sect. 5.3 describes how we compute J from our theoretical model, and Sect. 5.4.2 makes comparisons to available experimental results.

The overall dependence of looping on the length of the intervening DNA between the operators can be qualitatively understood as reflecting two competing phenomena. First, a short tether confines the second operator into a small region about the first one, increasing the effective concentration. But if the required loop is too short, then forming it will entail a large bending elastic energy cost, depressing the probability by a Boltzmann factor. For these reasons, the cyclization J factor exhibits a peak at DNA length about $460\,\text{bp}$ [79]. Later work extended Shimada and Yamakawa's calculation in many ways, using a variety of mathematical techniques [43, 65, 80–94]; Sect. 7 will comment on some of this work.

We now state the definition of the J factor to be used in this paper, and introduce the closely related "differential J factor," which we call $\tilde{J}$. As outlined above, we consider
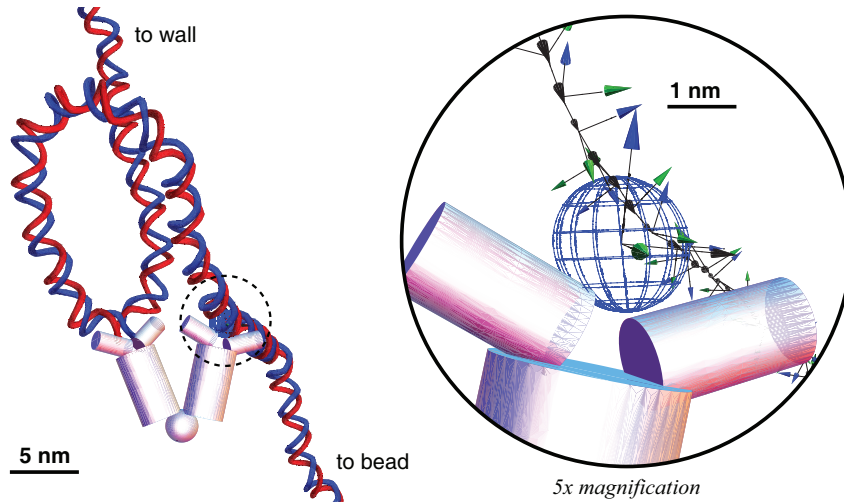
Figure 6. Illustration of the notion of target pose with a representative looped chain from our simulations. The chain shown is considered to be "looped" in the sense of Sect. 5.3.2 because the center of its O_1 operator matches its target within a certain tolerance (shown not to scale by the *blue-caged sphere*), and the orientation of the operator (*small arrows* in the inset) aligns with the target orientation (*large arrows* in the inset). DNA elasticity may favor thermal fluctuations that generate encounters with O_1 correctly oriented for binding (enhancing looping), or on the contrary, it may favor encounters incorrectly oriented for binding, depending in part on the number of basepairs between the two operators. Figs. 7–8 show this phenomenon in our numerical results.

fluctuations of the DNA chain conformation only, and ask how often operator O_1's position and orientation fluctuate to coincide with a "target" representing the available binding site on a LacI tetramer already bound to O_{id} (see Fig. 6). (A precise characterization of the target is given in Sect. S4.) A chain conformation is regarded as "looped" if the pose (position and orientation) of O_1 matches the target to within certain tolerances. We express the spatial tolerance as a small volume δv in space (with dimensions $(length)^3$), and the orientation tolerance as a small volume $\delta \omega$ in the group of rotations, normalized so that the full group has volume $8\pi^2$ [95]. The total group volume may be regarded as solid angle 4π for the director $\hat{\mathbf{E}}_3$, times angular range 2π for the rotations of the frame about $\hat{\mathbf{E}}_3$. Thus $\delta\omega$ is dimensionless.

Our Gaussian sampling Monte Carlo code generates many DNA chain conformations in a Boltzmann distribution. If we suppose that a LacI tetramer is bound to O_{id} with binding orientation β, then a certain fraction of these chains are looped in the above sense with O_1 binding orientation $\alpha = 1$; a different fraction are looped with $\alpha = 2$. Clearly both of these fractions go to zero if we take the tolerances δv or $\delta \omega$ to be small, so we define "differential J factors" as

$$\widetilde{J}_\alpha^{(\beta)} = \lim_{\delta v, \delta \omega \to 0} \left(\text{fraction in looped conformation } \alpha, \text{ given } \beta\right) / (\delta v \delta \omega). \qquad (5)$$

It is convenient to introduce the abbreviations

$$\widetilde{J}_{\text{tot}} \equiv \tfrac{1}{4} \sum_{\alpha, \beta} \widetilde{J}_\alpha^{(\beta)} \ \text{ and } \ J = 8\pi^2 \widetilde{J}_{\text{tot}}. \qquad (6)$$

Note that $\widetilde{J}$ and J naturally carry the dimensions of concentration. Our justification for the conventions in Eqn. (6) is that J defined in this way is a generalization of the familiar cyclization J factor [76–78]. To see this, suppose that we consider a very long loop. Then whenever O_1 wanders into its target volume, its orientation will be isotropically distributed, and in particular all four of the $\widetilde{J}_\alpha^{(\beta)}$ are equal. If a LacI tetramer is bound to O_{id}, then the effective concentration J of O_1 in the neighborhood of its other binding site (regardless of orientation) is related to the probabilities defined in Eqn. (5) by (say) $J = 8\pi^2 \widetilde{J}_1^{(1)}$. For arbitrary loop length (not necessarily long), we replace the last factor by its average, obtaining Eqn. (6).[+]

$\widetilde{J}_\alpha^{(\beta)}$ depends on the position and orientation of the target; Sect. 5.4 will take these to be defined by the crystallographic structure of the repressor tetramer. But more generally, we can regard $\widetilde{J}_\alpha^{(\beta)}$ as a function of *arbitrary* target pose, which we will compute and display in Sect. 5.3.1.

Although in principle TPM experiments can obtain the absolute magnitude of J, in practice the available experimental data are still sparse. Fortunately, the *ratio* of J factors for

[+] For the case of cyclization there are no labels α, β and no average; we then have $J = 8\pi^2 \widetilde{J}$, which with Eqn. (5) agrees with the definition in Ref. [65].

two different situations is more readily obtainable than the absolute magnitude (see Sect. S5).

For this reason, we will sometimes report experimental values normalized to a mean value $\bar{J}$,

which we define as

$$\bar{J}(\text{long}) = \text{mean of measured } J \text{ values over the range } 300 \leq L_{\text{loop}} \leq 310 \, \text{bp} . \quad (7)$$

5.3. Calculation of looping J factor

Sects. S6–S7 describe how we generalized the Gaussian sampling Monte Carlo algorithm of

Sect. 4.1 to handle looping. Sect. S9 describes how we checked our code, and our definitions

such as Eqns. (5–6), by calculating the *cyclization J* factor and comparing to the classic result

of Shimada and Yamakawa.

5.3.1. Orientation distribution of looped states Each binding orientation of O_{id}, with

$\beta = 1, 2$, yields a characteristic distribution $\mathcal{C}_\beta$ of allowed chains, each with a particular

pose for the center basepair of O_1. Of these, a small subset $\mathcal{C}_\beta^*$ will be "hits," i.e. will have that

center basepair inside its target volume for binding of the other site on LacI (see Fig. 6 inset).

We are ultimately interested in a smaller subset still, namely those chains $\mathcal{C}_{\beta\alpha}^*$ for which O_1

is also in one of its two target orientations. First, however, it is instructive to examine the

distribution of orientations for O_1 in $\mathcal{C}_\beta^*$. (The importance of this distribution was discussed

long ago by Flory and coauthors [77].)

For each "hit" configuration, we stored the orientation of the O_1 center segment relative

to the exit segment of O_{id}. Figs. 7 and 8 show the distribution of the tangential ($\hat{\mathbf{E}}_3$) and

normal vectors ($\hat{\mathbf{E}}_1$), respectively, for the "short" loop construct with loop length equal to

89 bp ($I = 0$ in the notation of Eqn. (4)). In these graphs we have taken the unit sphere and

divided into 20 finite-solid-angle bins. The coloring shown on each face of the icosahedron

represents the population of the corresponding angular bin.

Figs. 7 and 8 show that the orientation of hits is quite anisotropic, and not in general

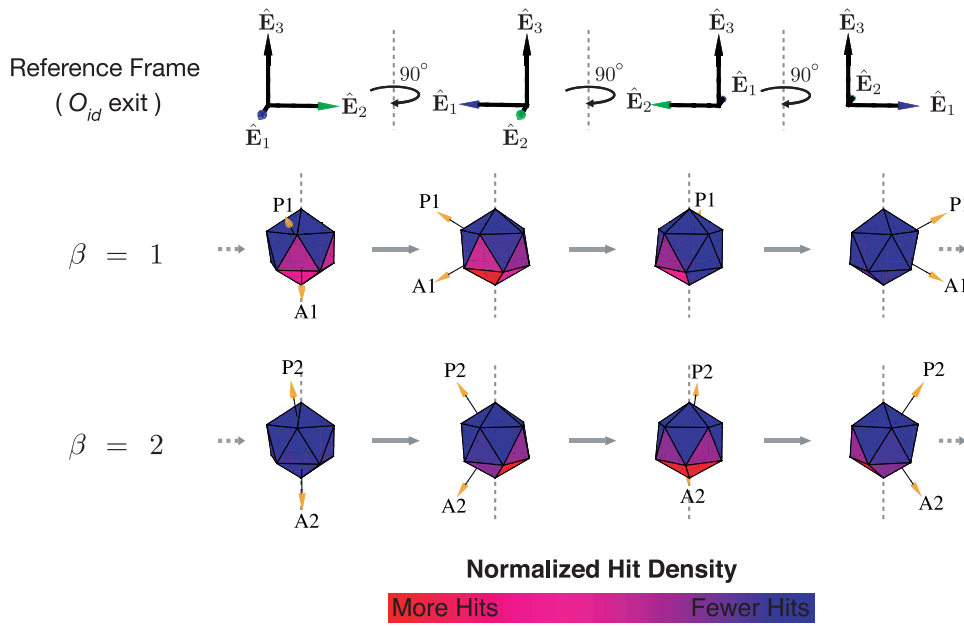peaked in the target orientation for forming any type of loop. These trends are characteristic

Figure 7. Distribution of the chain tangent vector for generated chains ending in the target volume ("hits," see Sect. 5.3.2) for the short construct tether. The possible directions for $\hat{\mathbf{E}}_3$ at the center of O_1 have been divided into twenty bins and the observed probabilities to land in each bin are assigned colors. Each row of the figure shows an icosahedron painted with the corresponding colors, from various viewpoints. The red faces correspond to the most populated bins; bluer faces correspond to lower hit densities. The four views represent clockwise rotations of the viewpoint by $90°$ about $\hat{\mathbf{E}}_3$ for the two binding orientations at O_{id}. The reference coordinate frames at top represent the orientation of the *exit* frame of O_{id}. Directions labeled P1, etc., refer to the target pose for the corresponding loop type, which does not in general agree with the most-populated bin. A total of about $7.5 \cdot 10^{10}$ chains were generated, resulting in 1673 hits with $\beta = 1$ and $12\,540$ hits with $\beta = 2$.

of all loop lengths; however, the same plots of the "long" loop construct (not shown) reveal a broader, though still peaked, distribution. The broadening of the distribution as the loop length increases is to be expected and is a natural consequence of the lability of long DNA loops. As the loop length is increased one segment at a time, the distribution of the tangential vector evolves slowly, but the peak of normal vector distribution rotates with each added segment
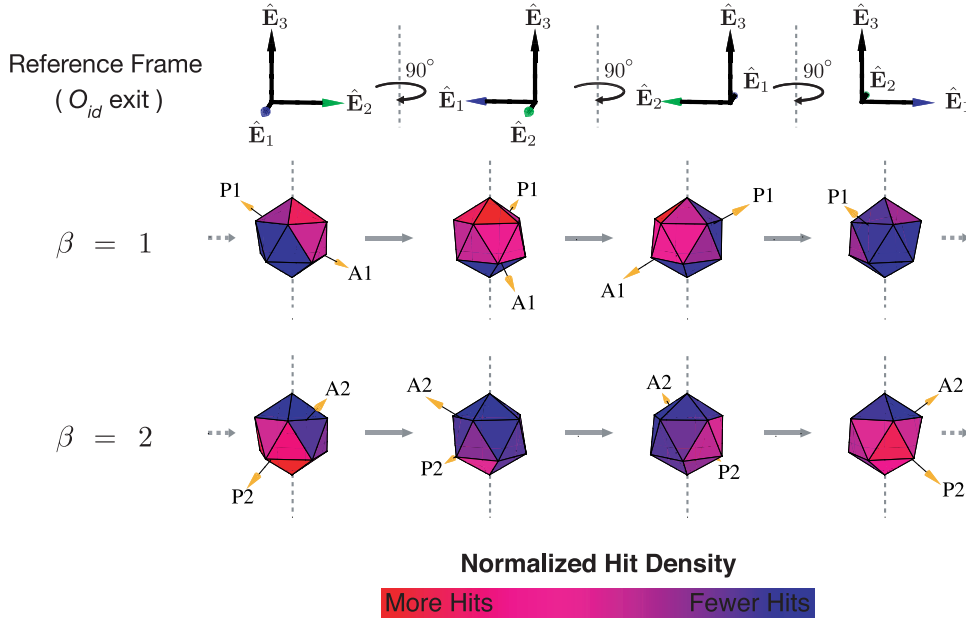
Figure 8. Distribution of the normal vector $\hat{\mathbf{E}}_1$ for the short construct tether. Other conventions are similar to Fig. 7, except the directions labeled P1, etc., correspond to the target normal vector $\hat{\mathbf{E}}_1$ for the corresponding loop type.

by about $2\pi\ell_0/\ell_{\text{helix}} \approx 2\pi/5$ radians (data not shown). This rotation of the normal vector distribution with changes in loop length corresponds to the helical nature of DNA; as the peak rotates about the fixed target orientation, we get an approximately periodic modulation in the J factor called "helical phasing" [96]. A more quantitative treatment of this behavior follows in Sect. 5.4.

5.3.2. Looping criteria and tolerance choices Chains generated with the target segment located within the target volume δv ("hits") pass the first constraint, the spatial tolerance check, as mentioned above (see also Fig. 6). All results correspond to a spatial tolerance of $\delta v = (4\pi/3)(2\,\text{nm})^3$. Classification of chains as looped or not is further dependent on an orientational constraint defined by $\delta\omega$. We required that the tangent vector to the chain, $\hat{\mathbf{E}}_3$, at the center of O_1 lie within a cone of angular radius $\pi/4$ radians of the target direction. We also required that the major-groove direction at the center of O_1, $\hat{\mathbf{E}}_1$, projected to the 1-2

plane of the target orientation, must match the corresponding target frame vector to within $2\pi/5$ radians. In other words, we checked whether the orientation of the major groove of the generated chain's central operator segment matches its target orientation. If both of these conditions are met, the "hit" conformation is considered to be "looped." The group volume corresponding to these angular tolerances is thus $\delta\omega = 2\pi(1 - \cos\frac{\pi}{4})(2 \times \frac{2\pi}{5}) \approx 4.63$, which is much smaller than the full group volume $8\pi^2$. After a chain is classified as looped or not, we proceed as described in Sect. S7.

According to Eqn. (5), we are interested in a limit as the tolerances $\delta v, \ \delta\omega$ approach zero. In practice we must of course keep these quantities finite, but we checked that we were reasonably close to the limiting behavior by checking two other choices of these tolerances: We cut the spatial tolerance in half, leaving the orientational tolerances the same, and we cut the orientational tolerances in half, leaving the spatial tolerance at $(4\pi/3)(2\,\mathrm{nm})^3$. We found that, although the magnitude of the phasing oscillations increased slightly for each reduction of the tolerances, nevertheless in each case the qualitative effect on the J factor calculations (and also on the RMS probability distributions, Sect. 6) was minimal (data not shown).

We have chosen to report results of the larger tolerance for two reasons: First, the number of hits is proportional to the tolerance, so we obtain better statistics with larger tolerance; second, larger tolerances may actually do a better job of representing the real experimental situation, specifically flexibility in the head regions of the Lac repressor, which we do not otherwise include. Recent all-atom simulations suggest that this flexibility is substantial [97].

5.4. J factor results

Before presenting results for looping in TPM experiments, we briefly describe a simpler warmup calculation. Then Sect. 5.4.2 describes a calculation that can be compared to TPM data, with moderately good agreement; Sect. 6.2 shows a much more striking agreement of theory with another kind of TPM data.
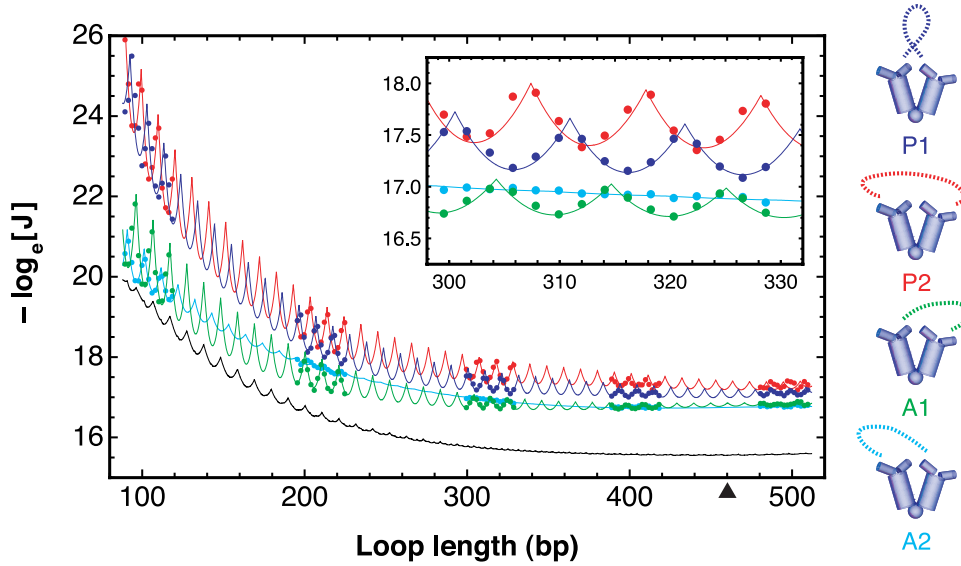
Figure 9. J factor for pure looping, as a function of loop length L_{loop} in basepairs. The vertical axis shows minus the natural logarithm of J (measured in molar). (Some authors call this quantity $\Delta G_{\mathrm{loop}}/k_{\mathrm{B}}T$; see Sect. S5.5.) Thus, higher points on the curves indicate more difficult looping; the curve rises at the left because of the high elastic energy cost of a short loop. The triangle at 460 bp roughly corresponds to the minimum of the overall looping J factor. *Dots:* Our Monte Carlo results. Blue, red, green, and cyan represent the quantities $2\pi^2 \widetilde{J}_\alpha^{(\beta)}$ corresponding to P1, P2, A1, and A2 loops, respectively. *Curves:* Each set has been summarized by an interpolating function described in Sect. 5.4.1. *Black curve:* The sum of the colored curves, that is, the overall looping J factor assuming that each looping topology is equally weighted (see Eqn. (6)). *Inset:* An enlarged portion of the graph for loop lengths of ≈ 300–330 bp.

5.4.1. Pure looping   One can imagine an experiment involving a DNA construct with only the two operators and the basepairs between them, that is, no flanking segments joining the loop to a wall and a bead. Here we present results on this form of the looping J factor ("pure looping"). We will also plot our results alongside corresponding experimental numbers for *in vivo* looping, even though the latter correspond to rather different physical conditions.

The J factor for this situation, defined via Eqns. (5–6), can be calculated by a simplified

version of our Monte Carlo algorithm that generates only the interoperator DNA segments and hence omits the steric-constraint checking. Fig. 9 shows our calculation of this quantity as a function of loop length. Three sets of Monte Carlo data are reported, each spanning three helical repeats. The data for each topology are summarized by a global interpolating function equal to the minimum of a collection of parabolas, centered on L_{loop} values separated by ℓ_{helix}. The interpolating functions are specified by the overall phasing (horizontal shift), a scaling function which determines the widths of the parabolas as a function of loop length (physically representing effective twist stiffness), and an envelope function describing the heights of the successive minima (physically representing competing effects of bend stiffness and entropy). The figure shows that indeed interpolating functions of this form globally summarize our simulation data over a wide range of L_{loop} values. At shorter loop lengths, the contributions of a single topology seem to dominate at any particular loop length, resulting in a noticeable modulation of the overall looping J factor; however, at longer loop lengths (e.g., 300 bp), the contributions of each topology are all similar and tend to cancel out each others' modulations. The anti-parallel loop topologies are predicted to be the preferred state, accounting for 90% or more of the looped chains for loop lengths of about 89 to 120 bp.

Fig. 10 shows the free energy of looping for an *in vivo* repression study [13], as interpreted by Saiz et al. [42], along with our Monte Carlo results for the total pure looping J factor. The cellular environment is far from ideal in terms of understanding DNA looping behavior: For example, superhelical stress, other DNA binding proteins, and molecular crowding all complicate the interpretation. Moreover, some analyses assume that LacI is free in solution at a known concentration [42], whereas much of it is instead likely to be nonspecifically bound to DNA [40, 41] or otherwise unavailable.

Despite these reservations, the comparison to our predictions is interesting: Our calculation seems able to predict the rough magnitude of the *in vivo* looping J factor, to within about a factor of two, at long loop lengths. At shorter loop lengths, however, *in vivo*
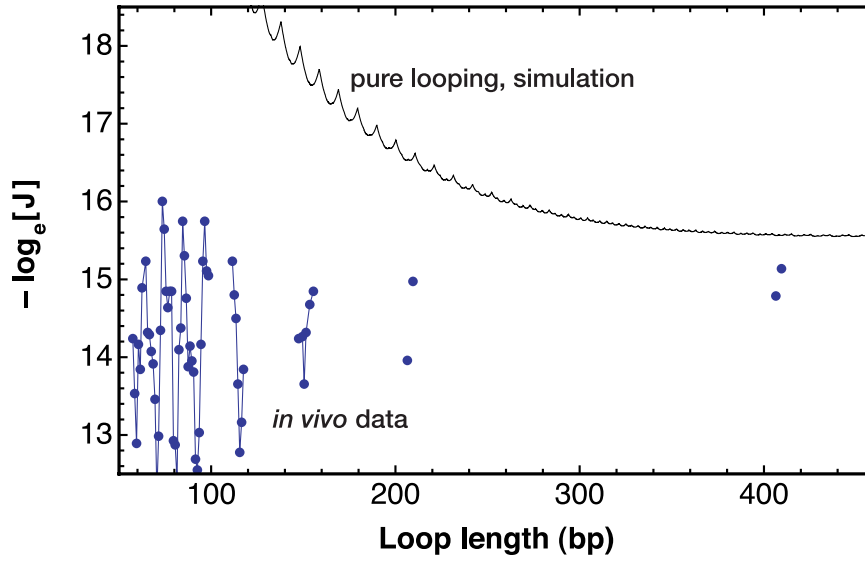
Figure 10. Comparison of our Monte Carlo results for pure looping to experimental data on *in vivo* repression. Experimental data from *in vivo* gene repression experiments [13] were converted to J factor values using a formula developed in Ref. [42] (see Sect. S5.5) and are shown in *blue*. The *black line* is an interpolation of our Monte Carlo results and is identical to the one in Fig. 9.

looping is far more prevalent than predicted from our simple model. The next subsection presents qualitatively similar results for the case of *in vitro* TPM experiments.

5.4.2. TPM looping   For the situation relevant to TPM experiments, the bead and wall must be taken into account. This necessitates use of the algorithm described in Sect. S7, to obtain an estimate of the looping J factor at each loop length and for each topology. The bead and wall affected the overall looping J factor, generally reducing it by about 30% for loops of length 100–300 bp. We can interpret this reduction in terms of the slight entropic stretching force generated by the bead and the wall [21]. We also found that the presence of the bead and wall significantly changes the relative weights of the various loop topologies from the corresponding pure looping case. For example, consider the "short" constructs. Even when the simulation generates a DNA conformation that qualifies as a type P1 loop, there is some chance that the conformation may be discarded because it violates one of the steric constraints;

the chance of retaining a P1 loop was found to be about twice the corresponding probability for an A1 loop.

We can understand this phenomenon qualitatively as follows: Due to the relatively short length of tether between the wall and the first operator, the DNA is generally pointing away from the wall when it *enters* the loop (at the first operator), thus favoring loops (P1 and P2) that maintain this directionality and keep the bead away from the wall. This bias is significant because the length of DNA from the second operator to the bead is relatively short. Presumably the reason P1 exhibits a larger shift than P2 is because the P1 topology *exits* the loop about $7\,\mathrm{nm}$ in front of where it *enters* the loop, whereas P2 *exits* about $7\,\mathrm{nm}$ behind where it *enters*.

Overall magnitude of J We first examine the overall magnitude of J. To minimize the effects of statistical experimental error, we computed the average quantity $\bar{J}$ (see Eqn. (7)) for both the "long" and "short" constructs. Our Monte Carlo calculation yielded the value $\bar{J}(\mathrm{long,\ theory}) = 100\,\mathrm{nM}$ and the ratio

$$\bar{J}(\mathrm{short,\ theory})/\bar{J}(\mathrm{long,\ theory}) = 2.0/100 \approx 0.020. \tag{8}$$

That is, harmonic elasticity theory makes the qualitative prediction that the short loop should be strongly penalized for its high elastic energy cost.

Turning next to the experimental values, we faced the problem that TPM data yielding an absolute number for J are so far available only for one loop length (see Sect. S5.4). However, Eqn. (S9) shows that this one point can be used to normalize all the others. With this procedure, we found that $\bar{J}(\mathrm{long,\ exp})$ lies in the range 24–$45\,\mathrm{nM}$. Thus the predicted overall magnitude of the J factor for long loops, computed with no fit parameters, lies within a factor of 2.2–4 of experiment, or equivalently our simulation found the free energy of looping ΔG_{loop} in agreement with our experimental determination to within about $k_{\mathrm{B}}T \ln 3 \approx 1\,k_{\mathrm{B}}T$.

Our uncertainty in overall normalization drops out of ratios such as $\bar{J}(\text{short, exp})/\bar{J}(\text{long, exp}) \approx$ 0.35. Comparing to Eqn. (8) shows that our theoretical model cannot account for the relation between short- and long-loop J factors: In the short-loop regime, looping is much easier than predicted by harmonic elasticity theory. The next paragraph gives more details.

Variation of J Fig. 11 shows the behavior of J as we scanned through two ranges of loop lengths ("short" and "long"). Because of the large experimental uncertainty in the overall magnitude of J, we divided both theory and experimental values of J by their respective averages $\bar{J}(\text{long})$, thus forcing both the solid (our theory) and dashed (experiment) black curves in panel B to be centered on zero. Fig. 11*b* shows that, although individual looping topologies have significant phasing effects, these nearly cancel in our simulation results, because in this paper we assume that all four operator binding orientations have the same binding energy (see Fig. 5). (Similar phenomena were discussed in Refs. [98, 99].) Fig. 11*a* shows that the harmonic elasticity theory, embodied by our simulation, was unable to account for the relative free energy of looping of long versus short loops, overestimating $\Delta G_{\text{loop}}(94\,\text{bp}) - \overline{\Delta G_{\text{loop}}}(\text{long})$ by up to about $3.7\,k_{\text{B}}T$. This observed excess of looping for short DNAs joins other signs of non-classical elastic behavior, which also begin to appear at short length scales [32, 100]. However, it could instead be explainable in terms of other effects neglected in our model (see Sects. 1.3.3 and 8).

5.4.3. Open LacI conformation Sect. 5.4.2 showed that the hypotheses of harmonic elasticity, a rigid V-shaped LacI tetramer, and no nonspecific DNA–repressor interactions, cannot explain the high looping incidence seen in our experiments for short DNA. One possible explanation, for which other support has been growing, is the hypothesis of DNA elastic breakdown at high curvature [27, 28, 100]. Indeed, Ref. [93] showed that such elastic breakdown can accommodate both enhanced looping at short lengths, and normal DNA behavior observed for loops longer than $300\,\text{bp}$.
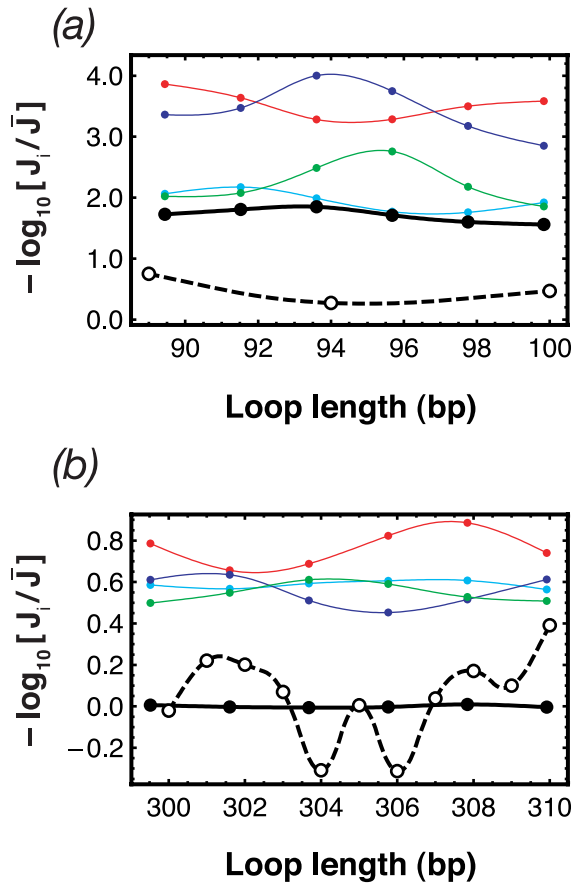
Figure 11. Comparison of the relative J factor from our Monte Carlo results (*solid, heavy black curves*) and TPM data of Han et al. [26] on random-sequence DNA (*open circles with dashed black curves*). (*a*) Relative J factors for the "short" DNA constructs (see Eqn. (4)), based on about $8 \cdot 10^9$ simulated chains. (*b*) Relative J factors for the "long" constructs, based on about 10^{10} simulated chains. All the experimental J factors are quoted relative to $\bar{J}(\text{long exp})$ defined by Eqn. (7) for the experimental data in panel (*b*); similarly, the theory values are relative to $\bar{J}(\text{long theory})$. The blue, red, green, and cyan solid lines represent contributions from P1, P2, A1, and A2 respectively; the heavy black solid line represents their sum.

An alternative hypothesis is that for our shorter loops, looping is actually dominated by the contribution from a distinct, "open" conformation of the repressor tetramer. Accordingly, we repeated our simulation for one particular representative version of the open conformation, the one discussed in [24]. Here each dimer is assumed to be rigid, but the opening

angle of the hinge where the dimers join has spread to $180°$. This time we found $\bar{J}_{\text{loop}}(95\,\text{bp})/\bar{J}_{\text{loop}}(305\,\text{bp}) \approx 0.13 \exp(-\Delta G_{\text{open}})$, where ΔG_{open} is the free energy cost of opening the tetramer. There are a wide variety of estimates of ΔG_{open}, but we see that even if it were equal to zero, the hypothesis of an open conformation still would not be consistent with our results.

6. Effect of looping on bead excursion

Sect. 8 will discuss the status of the results in the previous section, but clearly the agreement between theory and experiment is rather rough. We now turn to a much more striking comparison. In addition to studying the total probability of looping, TPM yields more detailed information about the effect of looping on bead excursion (see Figs. 12–13). A common experimental practice is to bin the data into finite sample windows, giving rise to a probability distribution of bead excursion. In this section we describe how we modeled this situation theoretically; for more details see Sect. S3. Figs. 12–13 show the degree to which our model successfully predicts the experimental observations.

6.1. Mimicking looped, doubly-bound DNA tethers

In the absence of LacI proteins, our procedure is straightforward: We generate chains as in Sect. 4.2, divide them into batches representing $4\,\text{s}$ windows (see also Sect. S3), and compute the RMS excursion in each batch. Instead of computing the mean of these $\rho_{\text{RMS},\,4\,\text{s}}$ values, however, we instead histogram their distribution. We will now apply essentially this same procedure to the more elaborate calculation of Sect. 5 to obtain the looping distributions we want, with one important modification.

Sect. 5 considered the *potential* for binding the O_1 operator. That is, we computed the fraction of time for which an unbound O_1 operator was positioned close to a pose suitable for binding to take place. Once binding does occur, however, the geometry of O_1 alters: It

develops a kink. Modeling the bead excursion for looped states requires that we account for the geometry of the fully bound complex, not the about-to-bind state. (See Sect. S4 for more on this distinction.)

Because we model the LacI tetramer as a rigid object, it may seem that for each selected looping topology we need only consider the DNA outside the loop region, replacing the entire loop by a single rigid Euclidean motion from the *entry* to the *exit* poses determined by the tetramer (see Fig. 4). Eliminating the loop region from the simulation would certainly speed up calculations, and indeed, it is nearly correct. However, this procedure would miss the possibility of steric clashes between the loop region and the bead and wall, potentially skewing the reported distribution of bead excursions. As a compromise between speed and accuracy, we simulated the regions wall→*entry*, and *exit*→bead, as usual, but, for each looping topology, inserted a *representative* loop between them. The representative was an actual loop configuration stored from the more exhaustive simulation in Sect. 5. The entire chain was then checked for steric clashes as usual, and the distribution of bead excursions for each of the four looping topologies was built up.

To find the appropriate relative weights to assign each of these four distributions, we used the simulations described in Sect. 5. Finally, we combined the resulting overall distribution for looped states with the corresponding one for the unlooped state. In principle, we could have found the appropriate relative weighting by using our computed looping J factor and the binding constant K_{d} extracted from experimental data in Sect. S5. In practice, however, the experimental data do not yet yield very precise values for K_{d}. Moreover the theoretical prediction for overall weighting depends very sensitively on the value of DNA persistence length, which is not precisely known. Accordingly (and in the spirit of Fig. 11), we chose the overall relative weighting between looped and unlooped states by hand for one value of L_{loop}. That is, we chose a common, constant value of this factor for all curves shown in Fig. 12. Our justification for this step is that our adjustment does not modify the locations, widths, nor

relative strengths of the looped-state peaks, which are thus no-parameter predictions of the theory.*

6.2. Bead excursion results

We followed the method of the previous subsection, including applying a modified blur correction appropriate for looped tethers (see Sect. S2.2). In order to obtain smooth distributions, we ran the Monte Carlo code until a total of 2.5×10^4 observations (independent values of $\rho_{\text{RMS}, 4\,\text{s}}$) were obtained. The results were then binned with bins of width $2\,\text{nm}$ and normalized.

Figs. 12 and 13 show the experimentally determined probability density functions for bead excursion as blue dashed lines. The rightmost peaks in these distributions correspond to our expectations for unlooped tethers (Fig. 3). One might think that at least one of the remaining peaks would be located at a value corresponding to an imagined tether that is unlooped, but shortened by the number of basepairs between the two operators; on the contrary, this expected peak location generally falls *between* the two peaks seen in the data [26]. Figs. 12–13 show that in contrast to this naive expectation, our simulation does a fairly good job of predicting *both* peak locations. Indeed, the figures show significant resemblance between the theory and experiment, including the trends as L_{loop} is varied. Specifically, the theory automatically yields two looped peaks, and roughly gives their observed locations and widths (each to within about $10\,\text{nm}$). The theory also predicts that the middle peak is small at 302–$304\,\text{bp}$, and bigger elsewhere, and that the lower peak is not modulated as much as the middle one, all of which are in agreement with the experimental data. All of these qualitatively satisfactory conclusions emerge without the hypothesis of any major conformational change of LacI.

* More precisely, we started with the predicted bead excursion histograms $P_{\text{loop}}(\rho)$ and $P_{\text{noloop}}(\rho)$. Then we chose a constant λ and displayed $(\lambda P_{\text{loop}} + P_{\text{noloop}})/(1 + \lambda)$. Choosing a value for λ that is smaller than unity thus enhances the relative contribution of the unlooped states.
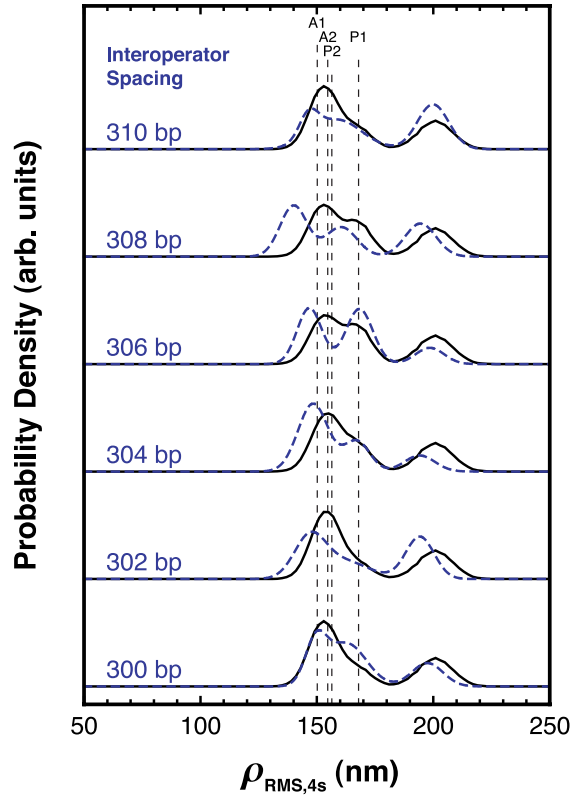
Figure 12. Theory and experiment for the probability density functions of RMS bead excursion for six of the "long chain" constructs ($L \approx 900\,\text{bp}$, $L_{\text{loop}} \approx 306\,\text{bp}$) studied by Han et al. [26]. *Blue dashed curves* show experimental TPM data on random-sequence DNA. *Black curves* show our theoretically predicted distributions for the corresponding interoperator spacings. The model yields these histograms as the sum of five contributions, corresponding to the four looped topologies (A1, A2, P1, and P2), and the unlooped state. Because our simulation results were not fits to the data, they did not reproduce perfectly the ratio of looped to unlooped occupancies. For visualization, therefore, we have adjusted this overall ratio by a factor common to all six curves (see main text). This rescaling does not affect the locations of the peaks, the relative weights of the two looped-state peaks, nor the dependences of weights on loop length L_{loop}, all of which are zero-fit-parameter predictions of our model. The separate RMS displacements for each individual loop topology, for the 300 bp case, are also shown as *vertical dashed lines*.
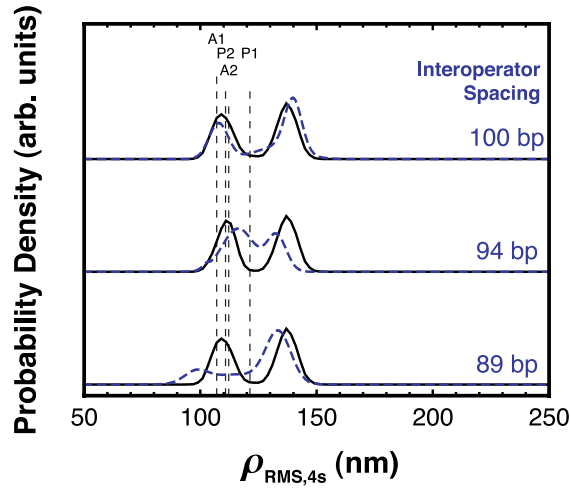
Figure 13. Theory and experiment for the probability density functions of the finite-sample RMS bead excursion for our three "short chain" constructs. The separate RMS displacements for each individual loop topology, for the $89\,\mathrm{bp}$ case, are also shown as *vertical dashed lines*. In other respects the figure is similar to Fig. 12.

The dissection of bead excursion into distinct peaks is also relevant to the question of a possible open conformation of the LacI tetramer. Even if we suppose that the middle looped peak in Fig. 13 reflects an open LacI conformation, as proposed by Wong et al., nevertheless those authors also proposed that the lower peak reflected the closed (V-shaped) conformation [24]. The experiments of Han et al. show that these peaks have comparable strength, and so in particular the lower one is inconsistent with the assumption of harmonic DNA elasticity. (One could instead propose that *both* peaks reflect an open conformation, but Sect. 5.4.3 argued that even this new hypothesis is unlikely to explain the experimental results.)

7. Relation to others' results

The calculational approach in this paper is Gaussian sampling Monte Carlo (see Sect. 4). Here we make just a few specific comments about representative examples of other calculational methods. The reader may wish to pass directly to the discussion in Sect. 8.

7.1. Analytic methods

Some methods do not involve the generation of random matrices:

Diffusion equation on $E(3)$: A series of independent steps defines a random walk. Thus the successive bends between chain segments can be regarded as defining a walk on the group manifold of $SO(3)$, the rotation group; more generally, the successive poses of segments define a walk on the Euclidean group $E(3)$. The probability density function of segment poses evolves as we move along the chain, in a way that can be calculated by using a set of orthonormal functions [90, 101], a procedure mathematically similar to some calculations in quantum mechanics. In some cases the resulting series can be summed and represented as a continued fraction [87, 88, 94, 102]. Another approach to the evolution of a distribution is via matrix exponentiation [91], or other transfer-matrix approaches [89]. These approaches converge slowly, however, for the case of short chains, and none accommodates easily the sort of nonlocal constraints arising in TPM experiments.

Gaussian approximation: The looping J factor is essentially a probability for configurations satisfying a global constraint. As such it can be represented as a functional integral, which in turn can be approximated by an expansion of its integrand around its critical points (maxima). Keeping the value of the integrand only at the critical points recovers the equations of elastic rod equilibrium; however, it omits entropic contributions to the free energy (see for example [49, 74, 103]). An improvement to this procedure approximates the integrand as a sum of Gaussians about its maxima; the integral may then be done, yielding the entropic contribution as the log of a functional determinant [43, 75, 79, 85, 92, 104]. Unfortunately, this approximation breaks down when any eigenvalue of the fluctuation operator becomes small, and in particular when the loop becomes bigger than a few hundred basepairs. Also, it is again difficult to generalize this approach to incorporate nonlocal constraints such as bead–wall exclusion. And although the method is efficient for comparing the free energies of

different looped topologies, the direct comparison of looped to unlooped is difficult. Perhaps for this reason, we do not know of any work using this approach that has attempted to include the dependence on LacI concentration; that analysis was crucial to the present work in order to disentangle the effects of J and K_d in the experimental data.

Note, however, that Zhang et al., using this approach, have introduced a more detailed model of the LacI conformation than the one in the present paper [43, 92]. Also, they and Swigon et al. introduced a detailed model of sequence dependence in their calculations, unlike the present work [75].♯ Like the present work, Refs. [43,75,92] neglected possible "wrapping" effects. Swigon et al. considered very low salt concentrations (and so had to account for long-range electrostatic effects), so their results cannot be directly compared to ours; moreover they considered only the situation we have called "pure looping" (Sect. 5.4.1). Nevertheless, it is interesting that for pure looping, our results agree qualitatively with their finding that, assuming the V-conformation of LacI, the antiparallel loop has lowest looping free energy.

7.2. Monte Carlo methods

We chose a Monte Carlo method because it is computationally tractable, it calculates the quantities actually observed in TPM experiments, it covers the entire range of loop lengths of interest to us, and it allows a simple implementation of all steric constraints relevant to our problem. Additionally, the method can readily be generalized to include sequence dependence [65] or nonlinear models of DNA elasticity [93]. As a bonus, Monte Carlo methods give a direct visualization of which chain conformations, and in particular which topologies, dominate the statistical sum, unlike the diffusion equation or matrix-exponentiation methods. It also gives us the distribution of near-miss configurations, allowing us to quantify the importance of the stereospecificity of binding (Sect. 5.3.1). Also, we are not obliged to find every critical point of the elastic energy, an error-prone search in a high-dimensional space

♯ Popov and Tkachenko also studied statistical properties of sequence-dependence effects [105] in models of this type. See also Refs. [86, 103, 106–108].

of configurations. For example, it is easy to miss stability bifurcations as the loop length is stepped through a range. Our Monte Carlo code stumbles upon the dominant configurations, including all topologies, without requiring human insight. (It also automatically lumps together all distinct topologies that cannot be experimentally distinguished, without the need to find their individual critical points, then add the corresponding contributions by hand.) Finally, unlike some methods, Monte Carlo simulation easily allowed us to work out the distribution of bead excursions (Sect. 6).

Among prior Monte Carlo methods we mention:

Gaussian Sampling: The work in this paper extends prior work in Refs. [15, 21]. As mentioned before, Czapla et al. also applied this method to cyclization (but not looping) [65]. Sect. S8.2 makes a specific point about a side calculation in that work.

Metropolis Monte Carlo, Brownian Dynamics: These powerful and general methods can in principle handle chains of any length, with arbitrarily complex constraints, and in some forms can also study kinetics (see, e.g., Ref. [80–84,109–111]). We only note that for the equilibrium calculations of interest to us, Gaussian sampling Monte Carlo is a simpler alternative method, which completes in a reasonable time on current laptop computers. Moreover, because in GS Monte Carlo every chain is independent of every other one, we have fewer worries about pre-equilibration, coverage of the allowed phase space, and so on than in Markov-chain MC methods.

All-atom simulation: Schulten and couthors have developed a hybrid method that represents the DNA as an elastic continuum, coupled to an all-atom simulation of the LacI tetramer [97, 112–115]. They did not apply this method to TPM experiments, and their simulation appears to neglect twist–bend coupling and entropic contributions from the DNA chain. However, simulations of this type did identify a "locking" mechanism, which apparently

prevents opening of the tetramer (i.e. maintains its overall V-form), even in the presence of significant external stress.

7.3. Fitting

Some insight into mechanisms can be gleaned by fitting data to phenomenological parameters roughly representing DNA stiffnesses etc., essentially obtaining interpolation formulas summarizing the data [42, 99]. For example, the anomalously high looping compared to theoretical expectations (Sect. 5.4.1) was previously noted in Ref. [42], and the possibility of partial cancellation of phasing modulation (Sect. 5.4.1) in Ref. [98, 99]. Ref. [42] also called attention to an asymmetric modulation in their graphs of looping free energy versus L_{loop}; we agree with their later observation that such asymmetries can arise from the sum of different loop topologies (fine structure in Fig. 10).

A drawback of the fitting approach, however, is that the inferred values of fit parameters do not have a literal interpretation as elastic constants; for example, their numerical values depend on extraneous variables [99]. The present work sought instead to predict the data from first principles, using fixed values of elastic constants. Also, many prior works studied *in vivo* data, whereas we have focused on TPM experiments for reasons described earlier.

7.4. Estimates and other analyses

The analysis of Wong et al. attempted to estimate the positions of the peaks in their TPM data from simple geometry applied to assumed configurations for the repressor tetramer [24]. The present work sharpens and corrects some qualitative estimates made, for example, in their Supplement.

Vanzi et al. obtained a looping J factor ≈ 0.1 nM from their data and noted that this value is much lower than those measured in other types of experiments [16]. Both our theory and our experimental results obtained from data in Ref. [26] agree that J is much larger than 0.1 nM,

albeit with large uncertainties for now. Vanzi et al. suggested that it would be important to calculate effects such as the entropic force generated by the bead; the present work gives the needed calculations.

8. Discussion

Our theoretical model and its main results were summarized in Sect. 1.3.

As mentioned earlier, DNA cyclization and looping have been the focus of many prior calculations. Broadly speaking, the novelty of the present work lies in the combination of a number of features: We have calculated looping, not cyclization; we have calculated quantities relevant to TPM experiments; and we have presented detailed comparisons to experimental data, with no fitting parameters. (Sect. 2 explained why we found TPM to be a particularly revealing class of experiments.) Finally, we know of no prior work that predicts the detailed distribution of bead excursions in looping as functions of loop length and LacI concentration. TPM experiments yield such data, and with it the prospect of experimentally separating the contributions of different loop topologies.

Fig. 3 shows that our simple model adequately captures much of the physics of equilibrium tethered particle motion without looping. Alternatively, the effective bead size may be slightly different from the nominal value, or effectively different due to surface irregularity. Despite these reservations, clearly our understanding of TPM is more than adequate to distinguish changes in effective tether length of $100\,\mathrm{bp}$ or more.

Sect. 5.4 gave a determination of the absolute magnitude of the looping J factor as a function of operator spacing, and a comparison to experiment. These comparisons were only approximately successful. We may point out, however, that the experimental determination had large uncertainties, because the available data are still somewhat sparse. In particular, all our absolute values currently depend on a single measurement, of the fraction of time spent looped for a $306\,\mathrm{bp}$ loop at $[\mathrm{LacI}] = 100\,\mathrm{pM}$ (see Eqn. (S9)). Also the slope of the titration

curve, and hence the parameter J_*, is still poorly characterized by available data (see Fig. S2 and Eqn. (S9)). Additional experiments would help improve this situation.

On the theory side, we note that existing measurements of the DNA persistence length ξ are subject to uncertainties, and that rather small changes in the assumed value of ξ make significant changes in our prediction for the overall J factor. Thus an accurate absolute prediction of J is perhaps demanding too much at this time; and in any case we also just argued that experiments, too, do not yet yield an accurate absolute experimental determination of J. To address both of these concerns, we noted that the *relative* J factors for different situations are better determined by experiments than the absolute magnitude, and so we scaled the theory and experimental results by their respective averages for loop lengths near 305 bp. Then we compared the predicted and observed relative J factors for several individual lengths near 305 bp, and also for a few lengths near 94 bp. Here the results were that *(i)* near 305 bp, neither theory nor experiment were strongly modulated by phasing, though experiment was more modulated than theory (Fig. 11*b*); *(ii)* near 94 bp, our theory predicts far less looping (lower J) than was observed (Fig. 11*a*).

The fact that harmonic elasticity theory underestimates P_{loop} for short loops cannot simply be attributed to our neglect of sequence information. Indeed, special sequences are observed to cyclize [116] and to loop [67, 117, 118] even more avidly than the random-sequence constructs studied here and in Ref. [26]. Nor can we simply suppose that we overestimated the true value of the DNA persistence length; reducing the value in the simulation would increase looping at both short and long lengths, leading to a worse discrepancy at the long end. Instead, we must look to our other physical hypotheses to see which one is breaking down for short DNA (see Sect. 1.3.3).

The very small phasing modulation observed in our calculations results partly from the near cancellation of out-of-phase modulations in the contributions from individual looping topologies. Certainly we could have obtained greater modulation had we been willing to

adjust our model's twist stiffness *ad hoc,* but our goal was to see how well the model predicted the data without fitting. It is possible that the degeneracy we assumed between the binding in each of the four topologies is an oversimplification, and that therefore the cancellation is not as complete as what we found in our simulation. (For example, we neglected the strain on the tetramer exerted by the DNA, which could differentially affect the different looping topologies.)

Our results were much more striking when we turned to the detailed distributions of bead excursions. These results are complementary to the ones on absolute and relative J factors, because the locations of the peaks are not affected by any possible elastic breakdown of the DNA within the loops; instead, they depend sensitively on the assumed geometry of the repressor tetramer. Fig. 12 shows that the distinctive three-peak structure, including the positions, and even the relative strengths, of the two lower peaks, emerges as a natural consequence of the four discrete looping topologies for LacI and its known geometry. We also found reasonable agreement with the less extensive available data on the peak positions for the short-loop bead excursion distribution (see Fig. 13).

Our calculations did not systematically study alternate, "open," conformation of LacI such as the one proposed in Ref. [119], although we did simulate one somewhat artificial model (a rigid, $180°$ conformation [24]). Although certainly such a conformational switch is possible, we note that Villa et al. have argued against it on grounds of molecular mechanics [97]. On the other hand, Wong et al. observed *direct* transitions between their two looped states (i.e., without an intervening unlooped episode), and they argued that this rules out any interpretation of the different looped states in terms of the distinct binding topologies. Our work has not resolved this issue. But in any case, the four loop topologies we studied must exist with any LacI conformation, and we showed that the closed conformation alone gives a surprisingly detailed account of the observations of Han et al. [26]. We also found that when we simulated the maximally advantageous, $180°$, conformation, the resulting looping J factor

still fell short of the value observed in experiments, even if we supposed zero free energy cost to pop into that state. And if desired, our calculation scheme may easily be extended to incorporate any desired hypothesis for LacI opening.

Finally, to illustrate the generality of our method, we also calculated J factors for a DNA with no bead or wall ("pure looping"). This calculation also gave us a quantitative estimate of the effect of the bead and wall on looping.

8.1. Future experiments

More extensive titration experiments will help pinpoint the values of J better, and eliminate the need to base all determinations of J on a single titration curve, as we have been compelled to do. Also, taking data with a fast camera shutter would eliminate, or at least minimize, the role of the blur correction (see also Sect. S2).

Our ability to resolve bead excursion distributions into distinct contributions from different looping topologies in Sect. 6 involved a subtle tradeoff. The finite sample RMS bead excursion, $\rho_{\mathrm{RMS},t}$, is more sharply peaked for longer sampling time t. Thus, using larger values of t could in principle clarify the structure of the distributions in Figs. 12–13. But increasing t also increases the fraction of incidents when a tether changes its looping state in the middle of a sample. One could imagine instead increasing the video frame rate, but Sect. S3 points out that the bead diffusion time sets a limit to what can be gained in this way. Thus it would be desirable to do experiments with smaller beads, hence faster bead diffusion times, and correspondingly faster video frame rate.

A second advantage to using smaller beads is that this would minimize the perturbation to looping caused by the bead, for instance the entropic force pushing the bead away from the wall, which slightly stretches the DNA.

Finally, it would be interesting to try experiments of the sort considered here but using other regulatory proteins, particularly ones not suspected to be as labile as LacI, in order to

see if the effects we calculated really are generic, as we believe they are.

8.2. Future theory

Certainly we could improve agreement with experiment by introducing two fitting parameters, which could be considered as effective twist and bend stiffnesses.†† Alternatively, the method in this paper could trivially be adapted to a detailed model of sequence-specific, harmonic DNA elasticity. But such detailed models may miss a larger point, that harmonic elasticity itself seems to break down at high bend and/or twist strain. An advantage of our Monte Carlo scheme is that it does not depend on the assumption of a Gaussian distribution of elementary bends; any distribution of interest may be substituted in the code, for example the one proposed in Ref. [100]. Finally, any desired characterization of repressor flexibility, for example the one outlined in Ref. [24], can be incorporated by drawing the matrices **M**, **N**, etc. (see Sect. S4) from appropriate distributions.

8.3. Conclusion

Our goal was to go the entire distance from an elasticity theory of DNA to new, quantitative experimental results. To cast the sharpest light on the comparison, we chose to study experiments that are free from confounding elements present *in vivo*, e.g., molecular crowding and DNA bending proteins other than the repressor of interest. We developed a number of techniques that will also be relevant for other experiments involving tethering of particles near surfaces. Although even this simplified setting presents some daunting geometrical complications (the effects of the tethered bead and wall), nevertheless it really is possible to understand much of the available experimental data (e.g., the three-peak structure of the bead excursion distribution) with a physically simple model. This level of detailed agreement will be helpful when trying to identify the many peaks in future experiments involving more

††The other two elastic constants, involving anisotropy and the twist–bend coupling, had small effects on our results (see Sect. S8.1).

than two DNA-binding sites. It also gives us strong evidence that our experiments are working as intended; for example, if bead sticking events were corrupting our data, it would be quite a coincidence if nevertheless we found agreement with theory for the full histogram of bead excursions. We did find significant deviations between theory and experiment, at short loop lengths. Here the fact that our underlying physical model had very few assumptions lets us focus attention more specifically on what modifications to those assumptions may be needed.

Acknowledgments

References

[1] K. S. Matthews. DNA looping. *Microbiol. Revs.*, 56(1):123–136, 1992.

[2] B. Müller-Hill. *The lac operon: A short history of a genetic paradigm*. Walter de Gruyter, Berlin, 1996.

[3] K. S. Matthews and J. C. Nichols. Lactose repressor protein: Functional properties and structure. *Prog. Nucleic Acid Res. Mol. Biol.*, 58:127–164, 1998.

[4] M. Lewis. The lac repressor. *C. R. Biologies*, 328:521–548, 2005.

[5] M. Ptashne. *A genetic switch: Phage lambda revisited*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 3rd edition, 2004.

[6] T. M. Dunn, S. Hahn, S. Ogden, and R. F. Schleif. 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(16):5017–5020, 1984.

[7] M. C. Mossing and M. T. Record. Upstream operators enhance repression of the *lac* promoter. *Science*, 233(4766):889–892, 1986.

[8] H. Krämer, M. Niemöller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Müller-Hill. *lac* repressor forms loops with linear DNA carrying 2 suitably spaced *lac* operators. *EMBO J.*, 6(5):1481–1491, 1987.

[9] W. T. Hsieh, P. A. Whitson, K. S. Matthews, and R. D. Wells. Influence of sequence and distance between two operators on interaction with the lac repressor. *J. Biol. Chem.*, 262(30):14583–14591, 1987.

[10] S. Adhya. Multipartite genetic control elements: Communication by DNA loop. *Annu. Rev. Genetics*, 23:227–250, 1989.

[11] B. Revet, B. von Wilcken-Bergmann, H. Bessert, A. Barker, and B. Müller-Hill. Four dimers of lambda repressor bound to two suitably spaced pairs of lambda operators form octamers and DNA loops over large distances. *Curr. Biol.*, 9(3):151–154, 1999.

[12] J. M. G. Vilar and S. Leibler. DNA looping and physical constraints on transcription regulation. *J. Mol. Biol.*, 331(5):981–989, 2003.

[13] J. Muller, S. Oehler, and B. Muller-Hill. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.*, 257(1):21–29, 1996.

[14] L. Finzi and J. Gelles. Measurement of lactose repressor-mediated loop formation and breakdown in single DNA-molecules. *Science*, 267(5196):378–380, 1995.

[15] P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap. Tethered particle motion as a diagnostic of DNA tether length. *J. Phys. Chem. B*, 110(34):17260–17267, 2006.

[16] F. Vanzi, C. Broggio, L. Sacconi, and F. S. Pavone. Lac repressor hinge flexibility and DNA looping: Single molecule kinetics by tethered particle motion. *Nucl. Acids Res.*, 34(12):3409–3420, 2006.

[17] J. F. Beausang, C. Zurla, C. Manzo, D. Dunlap, L. Finzi, and P. C. Nelson. DNA looping kinetics analyzed using diffusive hidden Markov model. *Biophys. J.*, 92:L64–L66, 2007.

[18] J. F. Beausang and P. C. Nelson. Diffusive hidden Markov model characterization of DNA looping dynamics in tethered particle experiments. *Phys. Biol.*, 4:205–219, 2007.

[19] M. Hogan, J. LeGrange, and B Austin. Dependence of DNA helix flexibility on base composition. *Nature*, 304:752–754, 1983.

[20] O. V. Tsodikov, R. M. Saecker, S. E. Melcher, M. M. Levandoski, D. E. Frank, M. W. Capp, and M. T. Record. Wrapping of flanking non-operator DNA in lac repressor-operator complexes: Implications for DNA looping. *J. Mol. Biol.*, 294(3):639–655, 1999.

[21] D. E. Segall, P. C. Nelson, and R. Phillips. Volume-exclusion effects in tethered-particle experiments:

Bead size matters. *Phys. Rev. Lett.*, 96(8):088306–(1–4), 2006.

[22] P. C. Nelson. Colloidal particle motion as a diagnostic of DNA conformational transitions. *Curr. Op. Colloid Interf. Sci.*, 12:307–313, 2007.

[23] L. Han, K. Towles, J. F. Beausang, H. Garcia, R. Phillips, and P. Nelson. Tethered particle motion: Theory and experiment for looped and unlooped DNA. *Biophys. J.*, 94:2573, 2008.

[24] O. K. Wong, M. Guthold, D. A. Erie, and J. Gelles. Interconvertible lactose repressor–DNA looped complexes revealed by single-molecule experiments. *PLoS Biology*, 6:e232–(1–15), 2008.

[25] D Normanno, F Vanzi, and FS Pavone. Single-molecule manipulation reveals supercoiling-dependent modulation of *lac* repressor-mediated DNA looping. *Nucl. Acids Res.*, 36:2505–2513, 2008.

[26] L. Han, H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips. Concentration and length dependence of DNA looping in transcriptional regulation. Submitted, 2008.

[27] J. Yan and J. F. Marko. Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.*, 93:108108–(1–4), 2004.

[28] P. A. Wiggins, P. C. Nelson, and R. Phillips. Exact theory of kinkable elastic polymers. *Phys. Rev. E*, 71:021909–(1–19), 2005.

[29] D. Shore, J. Langowski, and R. L. Baldwin. DNA flexibility studied by covalent closure of short fragments into circles. *Proc. Natl. Acad. Sci. USA*, 170:4833–4837, 1981.

[30] D. Shore and R. L. Baldwin. Energetics of DNA twisting 1. Relation between twist and cyclization probability. *J. Mol. Biol.*, 170(4):957–981, 1983.

[31] J. D. Kahn and D. M. Crothers. Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. USA*, 89:6343–6347, 1992.

[32] T. E. Cloutier and J. Widom. DNA twisting flexibility and the formation of sharply looped protein–DNA complexes. *Proc. Natl. Acad. Sci. USA*, 102:3634–3650, 2005.

[33] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. USA*, 102(15):5397–5402, 2005.

[34] N J Agrawal, R Radhakrishnan, and P K Purohit. Geometry of mediating protein affects the probability of loop formation in DNA. *Biophys. J.*, 94:3150–3158, 2008.

[35] Q. Du, M. Vologodskaia, H. Kuhn, M. Frank-Kamenetskii, and A. Vologodskii. Gapped DNA and cyclization of short DNA fragments. *Biophys. J.*, 88(6):4137–4145, 2005.

[36] J. M. Vilar and L. Saiz. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.*, 15(2):136–144, 2005.

[37] N. A. Becker, J. D. Kahn, and L. J. Maher. Bacterial repression loops require enhanced DNA flexibility. *J. Mol. Biol.*, 349(4):716–730, 2005. Erratum ibid. 353(4): 924–926, 2005.

[38] S.M. Law, G. R. Bellomy, P. J. Schlax, and M. T. Record. 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(1):161–173, 1993.

[39] N. A. Becker, J. D. Kahn, and L. J. Maher III. Effects of nucleoid proteins on DNA repression loop formation in Escherichia coli. *Nucl. Acids Res.*, 35:3988–4000, 2007.

[40] L. Bintu, N. E Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: Models. *Curr. Op. Genetics Devel.*, 15:116–124, 2005.

[41] L Bintu, NE Buchler, HG Garcia, U Gerland, T Hwa, J Kondev, T Kuhlman, and R Phillips. Transcriptional regulation by the numbers: Applications. *Curr. Op. Genetics Devel.*, 15(2):125–135, 2005.

[42] L. Saiz, J. M. Rubi, and J. M. Vilar. Inferring the in vivo looping properties of DNA. *Proc. Natl. Acad. Sci. USA*, 49:17642–17645, 2005.

[43] Y. Zhang, A.E. McEwen, D.M. Crothers, and S.D. Levene. Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS ONE*, 1(1):e136–(1–13), 2006.

[44] G. Lia, D. Bensimon, V. Croquette, J. F. Allemand, D. Dunlap, D. E. A. Lewis, S. C. Adhya, and L. Finzi. Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc. Natl. Acad. Sci. USA*, 100(20):11373–11377, 2003.

[45] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of RNA–polymerase observed by light-microscopy. *Nature*, 352(6334):444–448, 1991.

[46] C. Zurla, A. Franzini, G. Galli, D. D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi. Novel tethered particle motion analysis of cI protein-mediated DNA looping in the regulation of bacteriophage lambda. *J. Phys.: Condens. Matter*, 18(14):S225–S234, 2006.

[47] B. van den Broek, F. Vanzi, D. Normanno, F. S. Pavone, and G. J. L. Wuite. Real-time observation of DNA looping dynamics of type IIE restriction enzymes NaeI and NarI. *Nucl. Acids Res.*, 34(1):167–174, 2006.

[48] L. Han, B. H. Lui, S. Blumberg, J. F. Beausang, P. C. Nelson, and R. Phillips. Calibration of tethered particle motion experiments. In C.J. Benham et al., editor, *Mathematics of DNA Structure, Function and Interactions*. Springer, New York, 2008. In press.

[49] P. K. Purohit and P. C. Nelson. Effect of supercoiling on formation of protein-mediated DNA loops. *Phys.*

Rev. E, 74(6):061907–(1–14), 2006. Erratum ibid. 75:039903, 2007.

[50] N. Pouget, C. Dennis, C. Turlan, M. Grigoriev, M. Chandler, and L. Salome. Single-particle tracking for DNA tether length monitoring. *Nucl. Acids Res.*, 32(9):e73–(1–7), 2004.

[51] A. V. Vologodskii. *Topology and physics of circular DNA*. CRC Press, Boca Raton, 1992.

[52] P. Nelson. *Biological Physics: Energy, Information, Life*. W. H. Freeman and Co., New York, updated first edition, 2008.

[53] C. Bouchiat, M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophy. J.*, 76(1):409–413, 1999.

[54] J.F. Marko and E.D. Siggia. Bending and twisting elasticity of DNA. *Macromolecules*, 27:981–988, 1994. Erratum *ibid.* 29:4820, 1996.

[55] B. Eslami-Mossallam and M. R. Ejtehadi. Stretching an anisotropic DNA. *J. Chem. Phys.*, 128(12):125106–(1–11), 2008.

[56] J. D. Moroz and P. Nelson. Entropic elasticity of twist-storing polymers. *Macromolecules*, 31(18):6333–6347, 1998.

[57] W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, 95(19):11163–11168, 1998. The covariance matrices are available at `http://rutchem.rutgers.edu/~olson/cov_matrix.html`.

[58] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia. Extrinsic and intrinsic nucleosome positioning signals. http://arxiv.org/abs/0805.4017.

[59] T. R. Strick, V. Croquette, and D. Bensimon. Homologous pairing in stretched supercoiled DNA. *Proc. Natl. Acad. Sci. USA*, 95:10579–10583, 1998.

[60] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block. Stretching DNA with optical tweezers. *Biophys. J.*, 72(3):1335–1346, 1997.

[61] F. Vanzi, S. Vladimirov, C. R. Knudsen, Y. E. Goldman, and B. S. Cooperman. Protein synthesis by single ribosomes. *RNA*, 9:1174–1179, 2003.

[62] Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, and M. D. Betterton. Elasticity of short DNA molecules: Theory and experiment for contour lengths of 0.6–7 μm. *Biophys. J.*, 93:4360–4373, 2007.

[63] C. Bouchiat. Entropic elasticity of double-strand DNA subject to simple spatial constraints. *J. Stat. Mech. Th. and Exp.*, 2006. article P03019-(1–28).

[64] C. Bouchiat. Spatial constraint corrections to the elasticity of double stranded DNA measured with

magnetic tweezers. *J. Stat. Mech. Th. and Exp.*, 2007. article P06001-(1–22).

[65] L. Czapla, D. Swigon, and W. K. Olson. Sequence-dependent effects in the cyclization of short DNA. *J. Chem. Theor. Comp.*, 2(3):685–695, 2006.

[66] J. Y. Li, P. C. Nelson, and M. D. Betterton. Entropic elasticity of DNA with a permanent kink. *Macromolecules*, 39(25):8816–8821, 2006.

[67] L. Han et al. Unpublished, 2008.

[68] M. M. Levandoski, O. V. Tsodikov, D. E. Frank, S. E. Melcher, R. M. Saecker, and M. T. Record. Cooperative and anticooperative effects in binding of the first and second plasmid Osym operators to a LacI tetramer: Evidence for contributions of non-operator DNA binding by wrapping and looping. *J. Mol. Biol.*, 260(5):697–717, 1996.

[69] J. K. Barry and K. S. Matthews. Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. *Biochemistry*, 38(20):6520–6528, 1999.

[70] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, 271(5253):1247–1254, 1996.

[71] S. Oehler, E. R. Eismann, H. Krämer, and B. Muller-Hill. The 3 operators of the lac operon cooperate in repression. *EMBO J.*, 9(4):973–979, 1990.

[72] C. E. Bell, J. Barry, K. S. Matthews, and M. Lewis. Structure of a variant of *lac* repressor with increased thermostability and decreased affinity for operator. *J. Mol. Biol.*, 313(1):99–109, 2001.

[73] C. E. Bell and M. Lewis. Crystallographic analysis of *lac* repressor bound to natural operator O1. *J. Mol. Biol.*, 312(5):921–926, 2001.

[74] M. Geanacopoulos, G. Vasmatzis, V. B. Zhurkin, and S. Adhya. Gal repressosome contains an antiparallel DNA loop. *Nature Struct. Biol.*, 8(5):432–436, 2001.

[75] D. Swigon, B. D. Coleman, and W. K. Olson. Modeling the Lac repressor-operator assembly: The influence of DNA looping on Lac repressor conformation. *Proc. Natl. Acad. Sci. USA*, 103(26):9879–9884, 2006.

[76] H. Jacobson and W. H. Stockmayer. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.*, 18:1600–1606, 1950.

[77] P. J. Flory, U. W. Suter, and M. Mutter. Macrocyclization equilibria. 2. Poly(dimethylsiloxane). *J. Am. Chem. Soc.*, 98:5733–5739, 1976.

[78] N. L. Marky and W. K. Olson. Loop formation in polynucleotide chains. I. Theory of hairpin loop closure. *Biopolymers*, 21:2329–2344, 1982.

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

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

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

[82] H Merlitz, K Rippe, KV Klenin, and J Langowski. Looping dynamics of linear DNA molecules and the effect of DNA curvature: A study by Brownian dynamics simulation. *Biophys. J.*, 74(2, Part 1):773–779, 1998.

[83] A. A. Podtelezhnikov, N. R. Cozzarelli, and A. V. Vologodskii. Equilibrium distributions of topological states in circular DNA: Interplay of supercoiling and knotting. *Proc. Natl. Acad. Sci. USA*, 96(23):12974–12979, 1999.

[84] A. A. Podtelezhnikov, C. D. Mao, N. C. Seeman, and A. Vologodskii. Multimerization-cyclization of DNA fragments as a method of conformational analysis. *Biophys. J.*, 79:2692–2704, 2000.

[85] Y. L. Zhang and D. M. Crothers. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.*, 84(1):136–153, 2003.

[86] S Rappaport and Y Rabin. Effect of spontaneous curvature and sequence disorder on cyclization of fluctuating filaments. *Macromolecules*, 37(21):7847–7849, 2004.

[87] A. J. Spakowitz and Z. G. Wang. Exact results for a semiflexible polymer chain in an aligning field. *Macromolecules*, 37(15):5814–5823, 2004.

[88] A. J. Spakowitz and Z. G. Wang. End-to-end distance vector distribution with fixed end orientations for the wormlike chain model. *Phys. Rev. E*, 72(4):041802–(1–13), 2005.

[89] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon. Distribution functions, loop formation probabilities, and force-extension relations in a model for short double-stranded DNA molecules. *Phys. Rev. Lett.*, 94:138102–(1–4), 2005.

[90] J. Yan, R. Kawamura, and J. F. Marko. Statistics of loop formation along double helix DNAs. *Phys. Rev. E*, 71:061905–(1–17), 2005.

[91] N. Douarche and S. Cocco. Protein-mediated DNA loops: Effects of protein bridge size and kinks. *Phys. Rev. E*, 72(6):061902–(1–10), 2005.

[92] Y. Zhang, A. E. McEwen, D. M. Crothers, and S. D. Levene. Statistical-mechanical theory of DNA looping. *Biophys. J.*, 90(6):1903–1912, 2006.

[93] P. A. Wiggins and P. C. Nelson. Generalized theory of semiflexible polymers. *Phys. Rev. E*, 73:031906–

(1–13), 2006.

[94] A. J. Spakowitz. Wormlike chain statistics with twist and fixed ends. *Europhys. Lett.*, 73:684–690, 2006.

[95] Wu-Ki Tung. *Group theory in physics*. World Scientific, Singapore, 1985.

[96] S. Oehler, M. Amouyal, P. Kolkhof, B. Von Wilcken-Bergmann, and B. Muller-Hill. Quality and position of the three lac operators of *E. coli* define efficiency of repression. *EMBO J.*, 13(14):3348–3355, 1994.

[97] E. Villa, A. Balaeff, and K. Schulten. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc. Natl. Acad. Sci. USA*, 102(19):6783–6788, 2005.

[98] H. Krämer, M. Amouyal, A. Nordheim, and B. Müller-Hill. DNA supercoiling changes the spacing requirement of two *lac* operators for DNA loop formation with *lac* repressor. *EMBO J.*, 7(2):547–556, 1988.

[99] L. Saiz and J. M. G. Vilar. Multilevel deconstruction of the in vivo behavior of looped DNA-protein complexes. *PLoS ONE*, 2(4):e355–(1–7), 2007.

[100] P. A. Wiggins, T. Van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotech.*, 1(2):137–141, 2006.

[101] G. S. Chirikjian and Y. Wang. Conformational statistics of stiff macromolecules as solutions to partial differential equations on the rotation and motion groups. *Phys. Rev. E*, 62:880–892, 2000.

[102] S. Mehraeen, B. Sudhanshu, E. F. Koslover, and A. J. Spakowitz. End-to-end distribution for a wormlike chain in arbitrary dimensions. *Phys. Rev. E*, 77:0601803–(1–16), 2008.

[103] S. Goyal, T. Lillian, S. Blumberg, J.-C. Meiners, E. Meyhofer, and N. C. Perkins. Intrinsic curvature of DNA influences LacR-mediated looping. *Biophy. J.*, 93(12):4342–4359, 2007.

[104] I. M. Kulic, H. Mohrbach, V. Lobaskin, R. Thaokar, and H. Schiessel. Apparent persistence length renormalization of bent DNA. *Phys. Rev. E*, 72(4):041905–(1–5), 2005.

[105] Y. O. Popov and A. V. Tkachenko. Effects of sequence disorder on DNA looping and cyclization. *Phys. Rev. E*, 76:021901–(1–8), 2007.

[106] RS Manning, JH Maddocks, and JD Kahn. A continuum rod model of sequence-dependent dna structure. *J. Chem. Phys.*, 105(13):5626–5646, 1996.

[107] W. K. Olson, D. Swigon, and B. D. Coleman. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. *Phil. Trans. Roy. Soc. Lond. A*, 362(1820):1403–1422, 2004.

[108] Z. C. Zhou and B. Joos. Sequence-dependent effects on the properties of semiflexible biopolymers. *Phys. Rev. E*, 77:061906–(1–5), 2008.

[109] A. Vologodskii and N. R. Cozzarelli. Effect of supercoiling on the juxtaposition and relative orientation

of DNA sites. *Biophys. J.*, 70(6):2548–2556, 1996.

[110] Zhang Yang, Zhou Haijun, and Ou-Yang Zhong-Can. Monte Carlo implementation of supercoiled double-stranded DNA. *Biophys. J.*, 78:1979–1987, 2000.

[111] J. Huang, T. Schlick, and A. Vologodskii. Dynamics of site juxtaposition in supercoiled DNA. *Proc. Natl. Acad. Sci. USA*, 98(3):968–973, 2001.

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

[113] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten. Multiscale method for simulating protein-DNA complexes. *Multiscale Modeling Sim.*, 2(4):527–553, 2004.

[114] A. Balaeff, C. R. Koudella, L. Mahadevan, and K. Schulten. Modelling DNA loops using continuum and statistical mechanics. *Phil. Trans. Roy. Soc. London A*, 362(1820):1355–1371, 2004.

[115] A. Balaeff, L. Mahadevan, and K. Schulten. Modeling DNA loops using the theory of elasticity. *Phys. Rev. E*, 73:031919–(1–23), 2006.

[116] T. E. Cloutier and J. Widom. Spontaneous sharp bending of double-stranded DNA. *Molec. Cell*, 14(3):355–362, 2004.

[117] R. A. Mehta and J. D. Kahn. Designed hyperstable Lac repressor: DNA loop topologies suggest alternative loop geometries. *J. Mol. Biol.*, 294:67–77, 1999.

[118] M. A. Morgan, K. Okamoto, J. D. Kahn, and D. S. English. Single-molecule spectroscopic determination of Lac repressor-DNA loop conformation. *Biophys. J.*, 89(4):2588–2596, 2005.

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