Protein-mediated DNA loops: Effects of protein bridge size and kinks

Nicolas Douarche and Simona Cocco
CNRS-Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75005 Paris, France
(Received 11 August 2005; published 2 December 2005)

This paper focuses on the probability that a portion of DNA closes on itself through thermal fluctuations. We investigate the dependence of this probability upon the size r of a protein bridge and/or the presence of a kink at half DNA length. The DNA is modeled by the wormlike chain model, and the probability of loop formation is calculated in two ways: exact numerical evaluation of the constrained path integral and the extension of the Shimada and Yamakawa saddle point approximation. For example, we find that the looping free energy of a 100-base-pairs DNA decreases from 24 k_BT to 13 k_BT when the loop is closed by a protein of $r=10$ nm length. It further decreases to 5 k_BT when the loop has a kink of 120° at half-length.

DOI: 10.1103/PhysRevE.72.061902 PACS number(s): 87.15.−v, 82.35.−x, 46.70.−p, 36.20.−r

I. INTRODUCTION: DNA LOOPS IN GENE TRANSCRIPTION REGULATION

Gene expression is regulated by a wide variety of mechanisms. These activities as well as repression phenomena may occur at all expression steps (translation, transcription, etc) and do involve interactions between several biological molecules (DNA, RNAs, proteins, etc). For instance, proteins bound on specific DNA sequences may turn on or off gene transcription by interacting with each other. By bringing those proteins closer, DNA looping can ease their interactions [1–4]. Such looping events may be observed over a wide range of lengths spreading from hundreds to thousands of base pairs (bp). The looping probability was measured by cycling DNA segments in solution with cohesive ends. Once the loop is formed, proteins called ligases stabilize it [5,6]. It is then possible to count the circular DNAs with respect to the linear ones. The loop formation mediated by proteins has also been experimentally studied. Two examples are the loops formed by the LacR or the GalR transcriptional repressors [7]. Two units of such proteins bind at two specific positions along the same DNA and associate to form a complex when the binding sites come in contact. The formation of such loops has been recently studied using micromanipulation experiments on a single DNA molecule [1,2]. The study of the GalR mediated loop has shed light on the role of a third protein called HU that sharply bends (i.e., kinks) the DNA at half-length.

The DNA loop probability depends mainly on its length and flexibility. Long DNAs (typically longer than 1500 bp) essentially behave as Gaussian polymers (GP) [8,9]: the cyclization cost is mainly of an entropic nature [10]. On the contrary, for small lengths DNA cyclization is difficult mainly because of the bending energy cost. The computation of the elastic energy for the wormlike chain (WLC) model [8,9,11] can be analytically performed [14]; numerical methods have also been employed when electrostatic properties are included [15]. At intermediate length scales (in the range 50–2500 bp) elastic rigidity and entropic loss are both important. Several approximations have been developed to study this length's range [3,12,13,16], among them the calculations of the fluctuations around the lowest bending energy configurations performed by Shimada and Yamakawa [14]. Numerical approaches have also been developed: Monte Carlo [17] and Brownian dynamics-based simulations [4,18,19] as well as numerical calculations of the WLC path integral under the closed ends constraint [20,21]. This latter method allowed Yan, Kawamura, and Marko to study the elastic response of DNA subject to permanent or thermally excited bendings caused by binding proteins (such as HU) or inhomogeneities along the DNA double helix [21,22]. All these studies do provide a better understanding of the underlying regulation phenomena despite their overall complexity.

In this paper we study two processes that turned out to be important in DNA looping, namely, the size of the protein complex clamping the loop [4,18], acting as a bridge between the two DNA ends, and mechanisms implying DNA stiffness loss, which are taken into account in an effective way by kinking the WLC at half-contour length [4,18,19,23]. In Sec. II we define the model and the methods: we describe the numerical approach (Sec. II A) and the analytical saddle point approximation (SPA, Sec. II B) to calculate the r-dependent closure factor and the looping free energy. In Sec. III we compare the numerical and SPA results with previous experimental and theoretical results. In Sec. IV we extend the numerical and SPA approaches to a kinked loop; we discuss our results and we propose a simple formula that accounts for both the protein bridge and kink effects (Sec. IV C). We conclude (Sec. V) by sketching how to include omitted DNA properties, which may also play an important role in its closure such as twist rigidity or electrostactic interactions.

II. DEFINITIONS AND METHODS

We use the well-known wormlike chain (WLC) model [8,9,11]. The DNA polymer is described as an *inextensible* continuous differentiable curve of contour length L, with unit tangent vector $\vec{t}(s)$ $(0 \leq s \leq L)$. The polymer is characterized by the persistence length A beyond which tangent vectors lose their alignment: $\langle \vec{t}(s) \cdot \vec{t}(s') \rangle = \exp(-|s-s'|/A)$. The energy of a configuration of the polymer stretched under an external force $f\vec{e}_z$ reads

 ©2005 The American Physical Society

$$E[\vec{t};L,f]=\frac{1}{2}\frac{A}{\beta}\int_0^L\left[\frac{d\vec{t}(s)}{ds}\right]^2 ds-f\int_0^L \vec{e}_z\cdot\vec{t}(s)ds, \quad (1)$$

where we use $\beta=1/k_BT$. No twist elasticity or extensibility will be considered. The partition function is

$$Z(L,f)=\int \mathcal{D}\vec{t}\exp\{-\beta E[\vec{t};L,f]\}. \quad (2)$$

Notice that summation over all initial and final tangent vectors orientations, $\vec{t}(0)$ and $\vec{t}(L)$, is implicitly understood in this path integral.

In this paper we are interested in the formation of a loop in a DNA molecule, and the probability density functions (PDFs) of end-to-end distances play an important role. The quantities under study are denoted by Q, S, P, and J respectively and defined as follows: (i) the end-to-end extension $\vec{r}=(x,y,z)$ PDF at zero force,

$$Q(\vec{r},L)=\frac{1}{Z(L,f=0)}\int \mathcal{D}\vec{t}\,\delta\left[\int_0^L\vec{t}(s)ds-\vec{r}\right]\times\exp\{-\beta E[\vec{t};L,f=0]\}. \quad (3)$$

In the absence of force, Q depends on its argument $\vec{r}$ through its modulus $r=|\vec{r}|$ only, and we may introduce the radial PDF

$$S(r,L)=4\pi r^2Q(r,L). \quad (4)$$

(ii) The z extension PDF reads

$$P(z,L)=\int_{-L}^{L}dx\int_{-L}^{L}dyQ((x,y,z),L). \quad (5)$$

In the absence of external force, notice the x or y extensions PDFs are given by P too. Interestingly, the radial and z extensions PDF are related to each other through the useful identity [24],

$$S(r,L)=-2r\frac{dP}{dz}(r,L). \quad (6)$$

(iii) The cyclization factor

$$J(L)=Q(0,L), \quad (7)$$

defined as the density of probability that the two ends of the DNA are in contact with one another. (iv) The r-dependent closure factor is

$$J(r,L)=\left[\int_0^r S(r',L)dr'\right]\Big/\left(\frac{4}{3}\pi r^3\right). \quad (8)$$

It gives the density probability for the two ends of the chain to stay within a sphere of radius r. It is easy to check that $J(r,L)\to J(L)$ when $r\to 0$. For experimental convenience, units used for $J(L)$ and $J(r,L)$ are moles per liter: 1 M $\equiv$1 mol/L$\approx$0,6 nm^{-3}. In these units $J(r,L)$ gives directly the concentration of one binding site in proximity of the other. (v) Finally we consider the looping free energy cost

$$\beta\Delta G(r,L)=-\ln\left[J(r,L)\times\frac{4}{3}\pi r^3\right]. \quad (9)$$

Note that this definition does not include the details of the geometry or the affinities of the DNA-protein and protein-protein interactions. We actually assume all the sphere of radius r to be the reacting volume, i.e., that the loop will form if the DNA ends happen to be in this sphere.

Despite intense studies of the WLC model, no exact analytical expression is known for Q and the quantities of interest here, namely J and S. However, approximations expanding from the two limiting regimes (entropic [12,13] and elastic [14,16]) along with exact numerical computations are available. Hereafter, we have resorted to numerical as well as approximate analytical techniques (SPA) for calculating the cyclization factor J and the probability of almost closed DNA configurations.

A. Numerical calculation of the probabilities *P*, *S*, and *J*

Our starting point for the calculation of the z extension PDF is the Fourier representation of the Dirac δ-function in (3) and (5),

$$P(z,L)=\int_{-\infty}^{+\infty}\frac{dk}{2\pi}e^{+ikz}Z(L,f=-ik/\beta). \quad (10)$$

At fixed momentum k we are left with the calculation of the partition function $Z(L,f)$ at (imaginary) force $f=-ik/\beta$. The path integral (2) defining Z is interpreted as the evolution operator of a quantum system, the rigid rotator under an external imaginary field

$$Z(L,f)=\langle\text{final}|\exp[-L/A\times\hat{H}(f)]|\text{initial}\rangle. \quad (11)$$

The entries of its Hamiltonian $\hat{H}$ are easily expressed in the spherical harmonics $|\ell,m\rangle$ basis: $\langle\ell,m|\hat{H}(f)|\ell',m'\rangle=H_{\ell,\ell'}(f)\delta_{m,m'}$ with

$$H_{\ell,\ell'}(f)=\frac{\ell(\ell+1)}{2}\delta_{\ell,\ell'}-\beta Af\frac{\ell\,\delta_{\ell',\ell-1}+\ell'\,\delta_{\ell,\ell'-1}}{\sqrt{(2\ell+1)(2\ell'+1)}}. \quad (12)$$

The entries of $\hat{H}$ do not depend on the azimuthal number m due to cylindrical symmetry around the force axis. Finally, integration over all initial and final orientations for the tangent vectors at the ends of the polymer chain selects $|\text{initial}\rangle=|\text{final}\rangle=|0,0\rangle$.

A recent paper [24] used MATHEMATICA to compute the vacuum amplitude (11) through a direct matrix exponentiation. We have instead used the EXPOKIT library [25] since it proves to be more accurate and faster for intensive numerical calculations. We truncated Hamiltonian (12) in such a way that the outcome is insensitive to the cutoff on the harmonics order. The (inverse) Fourier transform (10) is then handled by a fast Fourier transform (FFT) algorithm [26]. This task is in particular facilitated thanks to the inextensibility constraint which makes the distribution bandwidth limited. Our results for the z extension PDF are shown in Fig. 1. The crossover from the rigid elastic regime ($L/A=0.1,0.5,1$) to

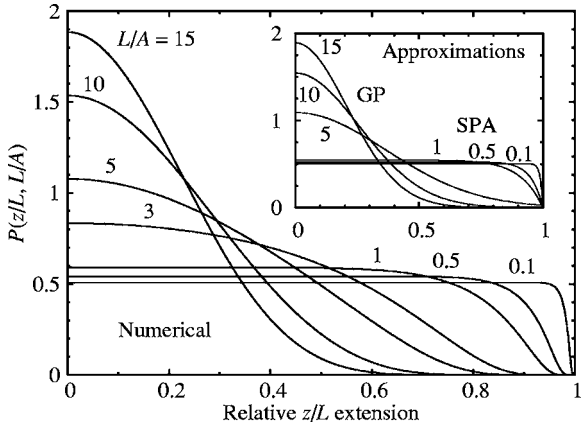

FIG. 1. Numerical computation of the z extension PDF over a wide range of contour lengths L. As expected, the agreement with the SPA prediction (see Sec. II B) improves as L decreases (upper bound $L/A \lesssim 1$). The long WLC behavior is caught by the GP approximation [8,9] as soon as, say, $L/A \gtrsim 5$.

the flexible entropic regime ($L/A=5,10,15$) is clearly visible. Using (6) then gives us access to S, the radial extension PDF (see Fig. 2). The most probable value for the distance $r_\star$ switches from full extension $r_\star \lesssim L$ for contour lengths $L \lesssim A$ (elastic dominated regime) to the GP most probable extension $\sqrt{4LA/3}$ for longer ones $L \gg A$ (entropic dominated regime). Also note that S always goes continuously to zero in the $r \to L$ limit due to WLC inextensibility. We have finally calculated the r-dependent closure factor J according to (8) by numerical integration of S, and the looping free energy cost ΔG defined in (9).

Let us now discuss the numerical errors that could be important when calculating probabilities of rare events. Main sources of error are the Hamiltonian (12) truncation and the integration step to compute the r dependent closure factor (8). As mentionned above, the cutoff on the harmonic order was systematically chosen to observe convergence. We have

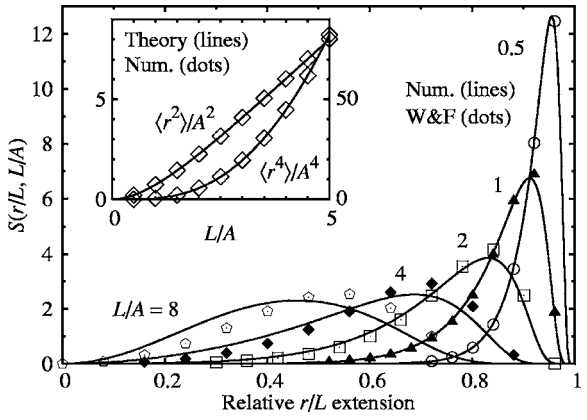

FIG. 2. Numerical computation of S, the radial r extension PDF. The outcomes of the numerical calculations are tested against exactly known values for the first even moments $\langle r^{2n} \rangle$ (inset) [8,9]. The shape of S compares very well to the widely used Wilhelm and Frey (W&F) expansion [16], valid up to $L/A \lesssim 2$. Similar tests were achieved with other popular approximation schemes [3] (data not shown).

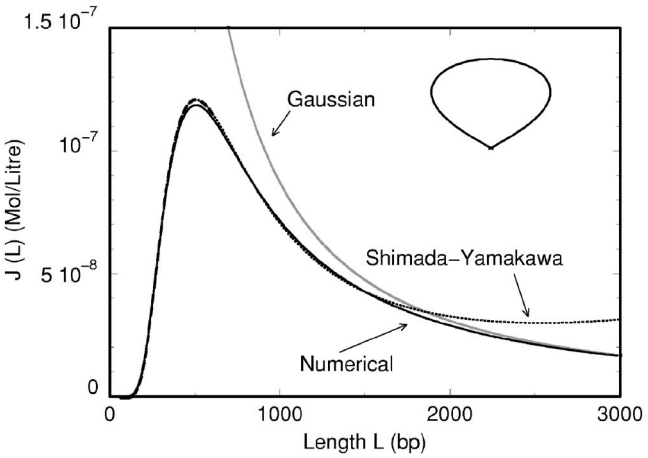

FIG. 3. Cyclization factor as a function of the DNA length with the Gaussian model (gray line); the WLC model with the Shimada and Yamakawa formula (dotted black line); the WLC model with the numerical calculation (full black line). The most probable length is 500 bp. Inset: the lowest bending energy configuration of a closed loop.

used $\ell=50$ after having verified that the result is unchanged for $\ell=100$. Concerning integration, limitation comes from the number of available data in the range $0 \leqslant r' \leqslant r$ which is directly related to the k sampling of the partition function Z. For $r=1$ nm the numerical integration is still reliable, but decreasing further r turns out to be critical. Other potential sources of error are negligible. Indeed, bandwidth limitation of $P(z;L)$ prevents any aliasing [26] during FFT (10) and the derivative of P (6) can actually be skipped by an integration by part of (8) to compute J. Further hypothetical errors would then come from the EXPOKIT library [25] itself but its routines were coded to accurately compute matrix exponentials over a broad range of matrices [27]. We have checked the numerical precision of our method by the comparison with the exact values for the first even moments of $S(r)$ (Fig. 2 inset); moreover, as shown in Fig. 2, $S(r)$ agrees with the Wilhelm and Frey expansion for small L/A values; finally, we will see in Sec. III that the numerically calculated cyclization factor $J(L)$ (8) is in agreement with the Shimada-Yamakawa and Gaussian approximation results for respectively small and large L (Fig. 3).

B. Saddle point approximation for J

In addition to the exact numerical calculations detailed above, we have carried out approximate calculations based on a saddle point estimate of the partition function (2). We follow the Shimada and Yamakawa calculation for the saddle point configuration [14], extending it for an opened DNA. The saddle point configuration for a closed loop is shown in Fig. 3 inset: it is a *planar loop*. The tangent vector at position s along the chain is characterized by its angle $\theta(s)$ with respect to the end-to-end extension $\vec{r}$. The optimal configuration is symmetric with respect to the perpendicular $\vec{e}_\perp$-axis while the half-length angle is $\theta(L/2)=180°$. The initial angle $\theta(0)$ is chosen by the minimization of the bending energy of

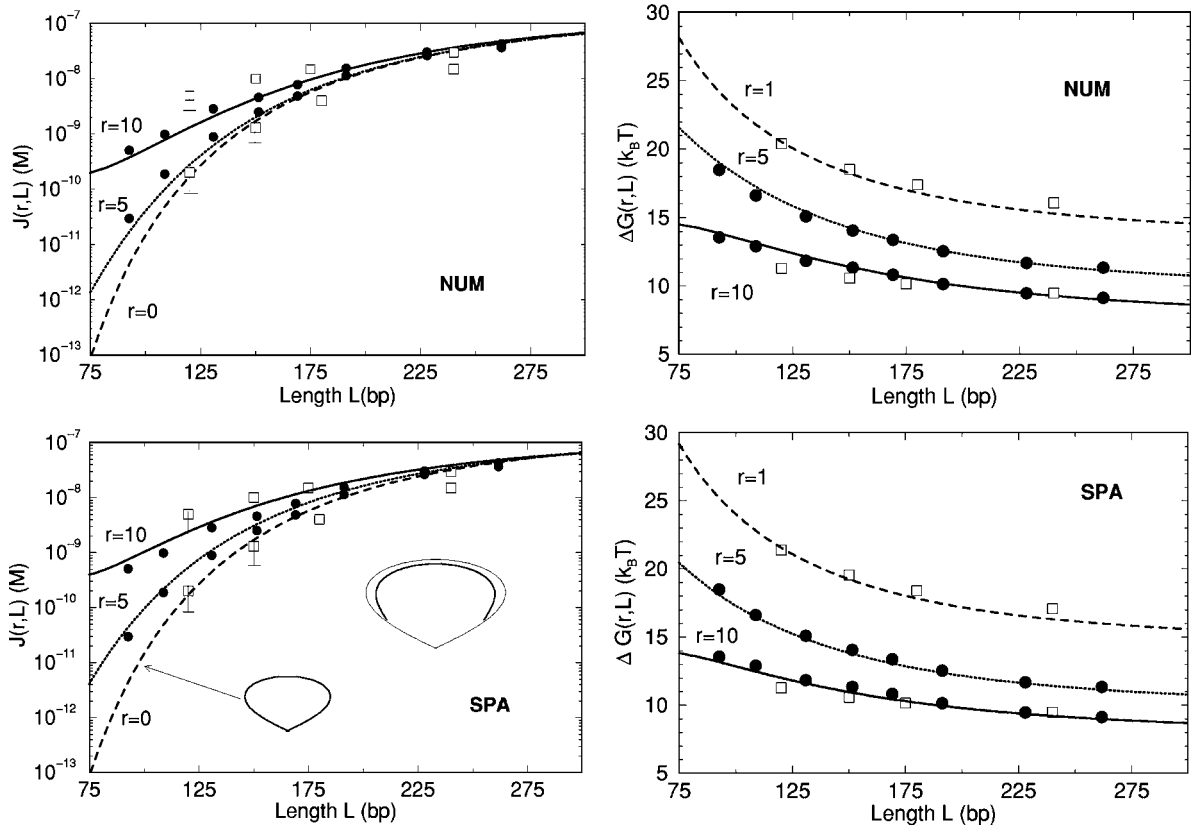

FIG. 4. Closure factor (left panel) and free energy (right panel) with a protein bridge of sizes: $r=1$ nm (dashed line); 5 nm (dotted line); 10 nm (full line). The error bars are shown when they are larger than the symbol sizes. Top: numerical calculation of the constrained path integral. The $r=1$ nm curve coincides with the $r=0$ cyclization factor. Bottom: closure factor obtained by the extension of the Shimada and Yamakawa calculation, which includes r. Theoretical results are in very good agreement with Monte Carlo simulations obtained by Podtelezhnikov and Vologodskii (filled circle, ●) [17], and in fair agreement with Brownian dynamics simulations obtained by Langowski *et al.* (empty square, □) [4,18]. Inset of the bottom left panel: lowest bending energy configurations for 100 bp and $r=0$ (bottom) or $r=10$ nm (top), the closure of the $r=10$ nm configuration is shown by a thin line.

the chain. The optimization gives rise to the following condition on the parameter $x=\cos[\theta(0)/4]$

$$(1+r/L)\hat{K}(x^2)=2\hat{E}(x^2), \tag{13}$$

where $\hat{K}(x^2)=K(\pi/2,x^2)$ and $\hat{E}(x^2)=E(\pi/2,x^2)$ are the complete elliptic integrals of the first and second kinds respectively [28]. The corresponding elastic energy is

$$\beta\Delta E(r,L)=\frac{4}{L/A}\hat{K}^2(x^2)\times(2x^2-1+r/L). \tag{14}$$

The end-to-end extension PDF is then approximated as

$$Q(r,L)=C(r,L)\exp[-\beta\Delta E(r,L)]. \tag{15}$$

The prefactor $C(r,L)$ should be calculated by taking into account quadratic fluctuations to the saddle point configuration. Since such calculation is quite involved, we will actually only extend the Shimada and Yamakawa results, which were computed considering fluctuations to a closed loop. In M=mol/L units this reads

$$C_{\mathrm{SY}}(L)=\frac{1.66}{A^3}\times\frac{112.04}{(L/A)^5}\exp(0.246\times L/A). \tag{16}$$

For this factor to fit the correct fluctuations (of the opened loop) we have to consider the fluctuations of a fake closed loop of similar bending energy. Such a loop may be obtained by considering the optimal configuration of a loop of contour length $(L+2r)$ instead of L, as shown in Fig. 4 (bottom, inset). The choice of the factor $2r$ derives from the following geometrical considerations: the closed saddle point configuration has an initial angle $\theta(0)=49.2°$, the ΔL closing the loop could be calculated for each value of r by requiring

$$\int_0^{\Delta L}\cos\,\theta(s)ds=r. \tag{17}$$

The angle $\theta(s)$ increases slightly on the first part of the trajectory $\theta(s)>\theta(0)$ and $\Delta L>1.53r$. We have chosen $\Delta L=2r$ as an approximate value; this approximation has the advantage that it can be directly put in the fluctuation expression $C(r,L)\approx C_{\mathrm{SY}}(L+2r)$ in Eq. (15) to obtain:

$$Q(r,L) = C_{SY}(L+2r)\exp[-\beta\Delta E(r,L)], \qquad (18)$$

where ΔE is given in formula (14) and C_{SY} is given in formula (16). The validity of this approximation for the fluctuations prefactor was checked out by comparing $J(r,L)$ obtained from $Q(r,L)$ through formula (8), with the numerical results. The good agreement shown in Fig. 4 allows us to obtain a semianalytical formula for the loop probability with a finite interacting volume, which is valid for molecules of up to 2 kb (kilo base pairs).

III. RESULTS: THE EFFECT OF A PROTEIN BRIDGE

In Fig. 3 we show the cyclization factor $J(L)$ for lengths L up to 3 kb. The most probable loop length is of about $L_\star = 500$ bp that is $L_\star/A \approx 3.5$ [4,14]. Both shorter and longer cyclized lengths are less probable: stiffness makes difficult the bending of smaller polymers while entropy makes longer polymers ends unlikely to meet. The numerical cyclization factor is compared with the saddle point calculation of Shimada and Yamakawa [14] and with the Gaussian polymer (GP) model. The former is in good agreement with the numerical results for lengths smaller than about 1.5 kb while the latter works well for loops larger than 2.5 kb.

The effects of the finite size r of the protein bridge are displayed in Fig. 4, which shows $J(r,L)$ (on the left) and $\Delta G(r,L)$ (on the right) for lengths L ranging from 75 bp (i.e., 25 nm) to 300 bp (i.e., 100 nm) and r respectively ranging from 1 nm to 10 nm. The numerical results (on the top) are in good agreement with the SPA results (on the bottom). Note that for small lengths L, $J(r,L)$ has a peak for the $L \approx r$ event corresponding to rigid rodlike configurations. Figure 4 does not show this peak, occurring for $r=10$ nm at about $L=30$ bp (or $L=10$ nm) because we focus on the cyclization events. As shown in this figure, r values ranging from 1 nm to 10 nm make no difference for the r-dependent closure factor, considering contour lengths L larger than 300 bp (or 100 nm). The cyclization factor $J(L)$ is evaluated as the limit $r \to 0$ of $J(r,L)$. In practice, convergence is reached as soon as r is about one order of magnitude smaller than L. In the range $L > 75$ bp (25 nm), $J(5\text{ nm}, 150\text{ bp})$ converges to $J(1\text{ nm}, 150\text{ bp})$. On the other hand, $J(r,L)$ is considerably different from $J(L)$ when r is of the same order of magnitude as L. For example, for loops of $L=100$ bp (34 nm) an end-to-end extension of 10 nm increases by two orders of magnitude the closure factor. Therefore proteins of size ≈ 10 nm are expected to produce drastic enhancements in looping short DNA sequences.

In terms of energetics (see Fig. 4, right) cycling a 100 bp DNA sequence costs 25 k_BT when the loop ends are required to stay within a sphere of radius $r=1$ nm. This cost decreases to 13 k_BT if the sphere has the typical protein size of $r=10$ nm. For loop lengths larger than 300 bp the only difference in the three curves of Fig. 4 (right) is a free energy shift due to the difference in the reacting volume. For instance, a reaction radius of 10 nm decreases the looping free energy of $3 \times \ln(10) \approx 7$ k_BT with respect to a 1 nm reaction radius.

Our results for the closure factor and the looping free energy are compared in Fig. 4 with the Monte Carlo (MC) and Brownian dynamics (BD) simulations results, obtained respectively by Podtelezhnikov and Vologodskii [17] (shown in the figure with filled circles, ●) and Langowski *et al.* [4,18] (displayed in the figure with empty squares, □). Numerical and SPA data (that for $r=1$ nm converge to the Shimada and Yamakawa curve) are in better agreement with the MC data than with the BD ones; indeed numerical, SPA and MC data are obtained with a simpler model than BD ones, which does not include twist rigidity and electrostatic effects.

Considerations about Lac operon repression energetics will help us illustrating our results and compare them with other previous results. Expression of proteins enabling bacteria *E. Coli* to perform the lactose metabolism can be prevented at the transcriptional level by cycling two different sequences [1], the smallest one including the operon promoter. Let $O_1O_3 = 76$ bp and $O_1O_2 = 385$ bp denote these two resulting DNA loops where the so-called operators $O_{1,2,3}$ actually are the small specific DNA sequences (10 bp or so) at which the tetrameric repressor protein LacR can bind thus clamping the desired loop. Notice both processes are needed for efficient repression: despite O_1O_3 containing the operon promoter, its cyclization is much less probable to occur than the O_1O_2 one (see cyclization factor $J(L)$ in Fig. 3). The LacR size is estimated from its crystallized stucture to $r \approx 13$ nm [7]. In the *in vitro* experiments many parameters are under control, among which are the operators' sequence and location. The distance between the two operators, which defines the length of the DNA loop, has been fixed in [2] to about 100 bp. Our results are in good agreement with the experimental measured stability of a 114 bp DNA loop mediated by a LacR protein, obtained by Brenowitz *et al.* in 1991 [7]. By measuring the proportion of looped complexes present in a solution with respect to the unlooped molecules they obtained a looping free enegy of 20.3 ± 0.3 k_BT to which they associated a closure factor of 8×10^{-10} M. From Fig. 4, the closure factor of a loop of 114 bp with a protein bridge of $r=10$ nm is $J(10\text{ nm}, 114\text{ bp}) = 10^{-9}$ M to which we associate, from formula (9) a cyclization free enegy $\Delta G(10\text{ nm}, 114\text{ bp}) = 12$ k_BT. Note that the very good agreement between the closure factor contrasts with the bad agreement for the cyclization free energy. The latter could have been calculated considering a different reaction volume or it could also include the competition with configurations that do not allow the formation of a loop (see Fig. 2 of [7]). To explain the high value found for the closure factor, Brenowitz *et al.* already included the size of the protein in the analysis of their results by comparing their J result with the value expected for the cyclization probability of a free DNA when the length of the protein is included in the size of the loop.

Another result on a DNA loop mediated by LacR protein has been obtained by Balaeff *et al.* [15] who have numerically calculated the elastic energy of the O_1O_3 loop from a WLC model also including: the twist rigidity, the short range electrostatic repulsion and the details of the LacR/DNA complex crystal structure. The elastic energy is estimated to be 23 k_BT in [15], of which 81% (that is 18 k_BT) is due to the bending and 19% is due to the unwinding. The bending energy of 18 k_BT is to be compared to the saddle point energy

(14) of 15 k_BT for a 75 bp loop with an end-to-end extension of $r=10$ nm. The corresponding free energy of loop formation obtained by (18) and after integration over the reacting volume (9) is $\Delta G(10\ \text{nm}, 75\ \text{bp}) = 14\ k_BT$ (see Fig. 4, bottom/right). For a 400 bp loop since LacR is only $\approx 10\%$ of the loop length, its size is expected to play a less important role. Indeed the cyclization probability does not depend on $r < 10$ nm and the free energy of forming such a loop decreases from 15 k_BT for $r=1$ nm to 8 k_BT for $r=10$ nm only because the reaction volume increases by a factor $3 \times \ln(10) \approx 7\ k_BT$.

IV. EFFECT OF THE PRESENCE OF A KINK

The previous protein mediated DNA looping modelization (Sec. III) assumes that the only intervening proteins are the ones clamping the loop ends (e.g., the Lac operon repressor LacR). Actually regulation phenomena involve several proteins which may bind along the whole DNA. Indeed naked DNA situations barely exist *in vivo*. For instance, single molecule manipulations [2] have shown that efficient Gal operon repression needs a stiffness loss of the 113 bp DNA portion to be looped. The HU protein produces such loss by kinking the sequence.

A. Numerical calculation of J for a bridged and kinked loop

Such stiffness loss may be taken into account in an effective way by kinking the WLC at half-length $L/2$. Let us call θ_- and θ_+ the angles of the DNA just before and after the kink respectively. We assume that the kink plane is vertical and choose it to define the origin of the azimuthal angle ϕ: $\phi_- = \phi_+ = 0$. Using the quantum language of Sec. II A, we replace the calculation of the evolution operator $Z(L,f)$ (11) with its kinked counterpart

$$\begin{aligned} Z_{\text{kinked}}(L,f) &= \langle \text{final} | \exp\left[-\frac{L}{2A}\hat{H}(f) \right] | \theta_+, \phi_+ \rangle \\ &\quad \times \langle \theta_-, \phi_- | \exp\left[-\frac{L}{2A}\hat{H}(f) \right] | \text{initial} \rangle \\ &= \sum_{\ell,\ell'} \langle 0,0 | \exp\left[-\frac{L}{2A}\hat{H}(f) \right] | \ell, 0 \rangle \langle \ell', 0 | \\ &\quad \times \exp\left[-\frac{L}{2A}\hat{H}(f) \right] | 0,0 \rangle \times Y^0_\ell(\theta_+, 0) Y^0_{\ell'}(\theta_-, 0), \end{aligned} \tag{19}$$

where we have used the change of basis from angles to spherical harmonics

$$|\theta_\pm, \phi_\pm = 0\rangle = \sum_{\ell \geq 0} Y^0_\ell(\theta_\pm, 0) |\ell, 0\rangle. \tag{20}$$

Although calculations are a little bit more involved, the evaluation scheme for the cyclization factor J remains unchanged in its principle (Sec. II A). We now have "$|\ell \neq 0, 0\rangle$ elements" corresponding to particular orientations arriving at $(-)$ and leaving from $(+)$ $s = L/2$. The kink angle κ is intuitively defined from these WLC tangent vectors at half-length (in spherical coordinates) through

$$\kappa = \theta_+ - \theta_-. \tag{21}$$

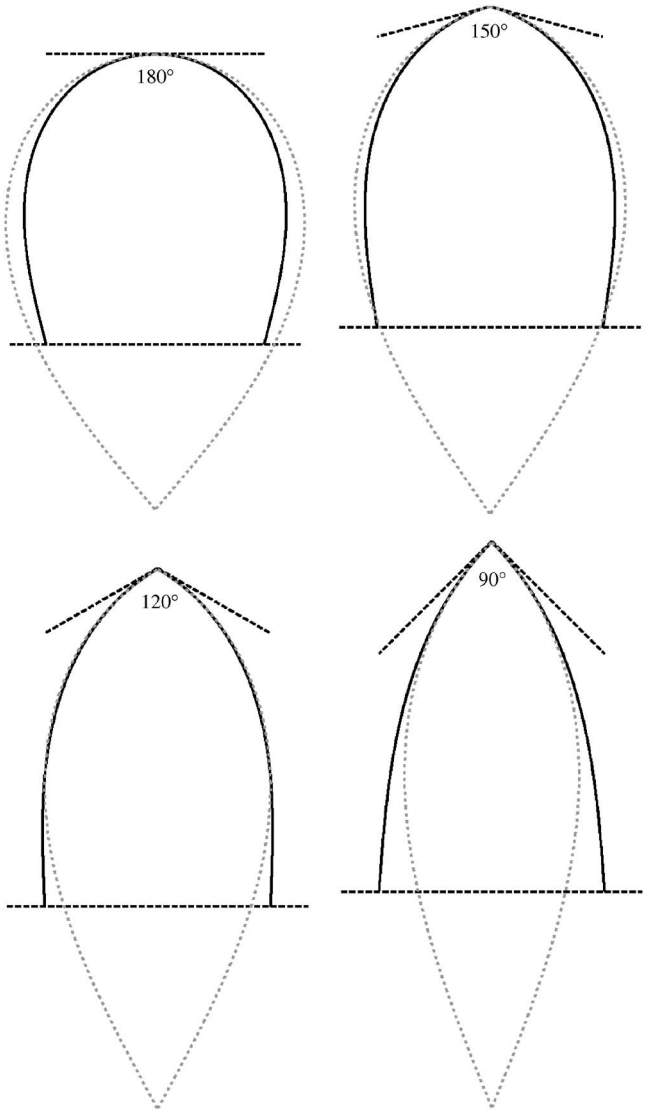

FIG. 5. Lowest bending energy configurations for a 100 bp DNA loop with an end-to-end extension $r=10$ nm and kinks $\kappa = 90°, 120°, 150°, 180°$, respectively. The gray configurations are the closed loops used in the calculation of the fluctuation.

B. Saddle point approximation of J for a bridged and kinked loop

The saddle point calculation of Sec. II B can be straightforwardly extended to the case of a kinked loop. In Fig. 5 we show the configurations with the lowest bending energy for a 100 bp DNA loop with an end-to-end extension $r=10$ nm. The kink is accounted for by a bending angle in the middle of the chain $\kappa = 2 \times \theta(L/2) - \pi$ *a priori* different from the previous trivial value $\kappa = \pi$, ranging from 150° to 90°. We introduce the phase $\psi = \arcsin[\sin(\pi - \kappa/4)/x]$. The parameter x is now obtained from equation

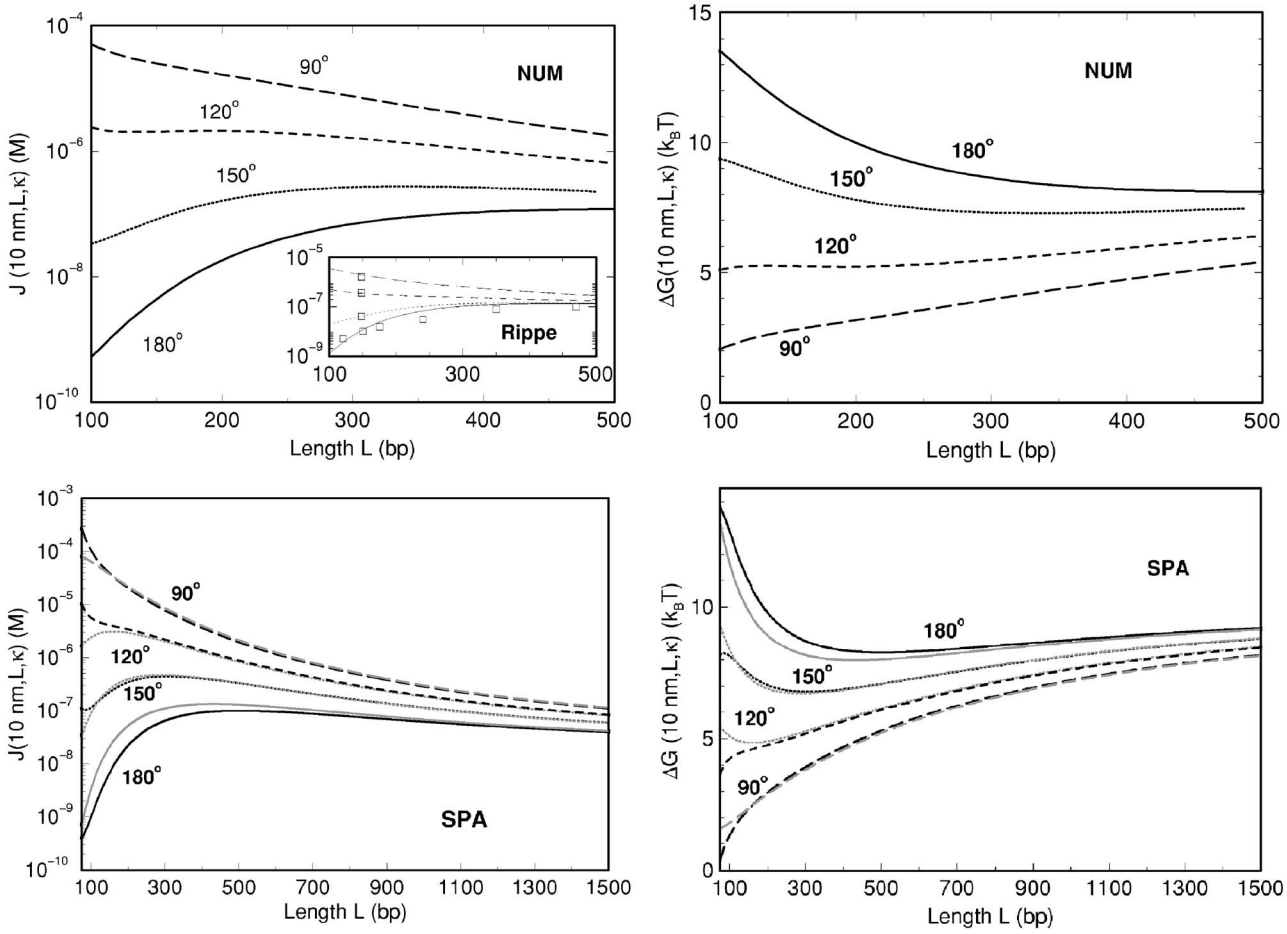

FIG. 6. Closure factor (left) and free energy for a loop with $r=10$ nm and a kink κ in the middle of the chain: $\kappa=180°$ (full lines); $\kappa=150°$ (dotted lines); $\kappa=120°$ (dashed line); $\kappa=90°$ (long dashed lines). Top: numerical calculation of the constrained path integral. Bottom: extension of the Shimada and Yamakawa calculation (black lines) and approximate formula (gray line) given in the text (25). Numerical results are in very good agreement with the SPA approximation and the approximate formula (25). Inset: Brownian dynamics simulations point obtained by the Langowski and collaborators (empty square, □) [4,18], fitted by a simple formula by Rippe [19].

$$(1 + r/L)[\hat{K}(x^2) - K(\psi,x^2)] = 2[\hat{E}(x^2) - E(\psi,x^2)] \quad (22)$$

and the total bending energy is

$$\beta\Delta E(r,L,\kappa) = \frac{4}{L/A}[\hat{K}(x^2) - K(\psi,x^2)]^2(2x^2 - 1 + r/L). \quad (23)$$

In analogy with (15) we obtain the end-to-end extension $\vec{r}$ PDF

$$Q(r,L,\kappa) = C_{\mathrm{SY}}(L + 2r)\exp[-\beta\Delta E(r,L,\kappa)], \quad (24)$$

where C_{SY} is given in (16), from which we calculate $J(r,L,\kappa)$ through formula (8). Note that (24) reduces to the loop probability given in [23] for a closed and kinked DNA. Again the good agreement obtained with numerical results allows us to establish a semianalytical formula for the loop probability with a finite interacting volume and kink.

C. Results for a kinked and bridged loop

In Fig. 6 results for the looping probability density (left) and free energy (right) are shown for a typical end-to-end extension $r=10$ nm and kinks $\kappa=150°$, 120°, and 90°. The curve with no kink ($\kappa=180°$) is also shown for comparison. The numerical results on the top of the figure are in very good agreement with the extension of the Shimada and Yamakawa saddle point approach on the bottom of the figure (black lines). In Fig. 6 (inset) we show the results obtained by BD simulations by Langowski *et al.* [4,18] for J (10 nm,L,κ) (empty squares, □) fitted by Rippe [19] with a simple formula containing one fitting parameter for each curve. The curves in the inset of Fig. 6 reproduce the same behavior with L and κ of our numerical (top of Fig. 6) or SPA curves (bottom of Fig. 6). The numerical gap with BD results increases for large kinks: at $\kappa=90°$ our closure factor is 10 times larger than the closure factor obtained by BD simulations. The electrostatic and twist rigidity effect could indeed play a more important role when the chain is kinked. Note that the lengths' range of the numerical results are from 100 bp to 500 bp, while the lengths' range of the saddle point results is from 75 bp to 1500 bp. The lengths' range of the SPA is limited by the validity of the approximation $\approx$1500 bp shown in Fig. 3. As an example, the closure factor J

$(10\ \mathrm{nm}, 113\ \mathrm{bp}, \kappa)$ of a 113 bp fragment, obtained by the numerical calculations, increases from the value of $10^{-9}\ M$ for a nonkinked loop ($\kappa=180°$) to $4\times10^{-8}\ M$, $2\times10^{-6}\ M$, and $4\times10^{-5}\ M$ for, respectively, $\kappa=150°$, 120°, and 90°. The corresponding looping free energy is $\Delta G(10\ \mathrm{nm}, 113\ \mathrm{bp}, \kappa)=13\ k_BT$, $9\ k_BT$, $5\ k_BT$ and $2.5\ k_BT$ for $\kappa=180°$, 150°, 120°, and 90° respectively. With the saddle point approach we find similar results: $J(10\ \mathrm{nm}, 113\ \mathrm{bp}, \kappa)=2\times10^{-9}\ M$, $10^{-7}\ M$, $5\times10^{-6}\ M$, and $8\times10^{-5}\ M$ while $\Delta G(10\ \mathrm{nm}, 113\ \mathrm{bp}, \kappa)=11\ k_BT$, $8\ k_BT$, $4.4\ k_BT$, and $1.7\ k_BT$. As it is shown in Fig. 6, the presence of the kink becomes irrelevant for DNA segments larger than about 1500 bp. The example of a 113 bp DNA segment has been chosen to compare the results with the single- molecule experiments on the GalR mediated loop of an ≈113 bp DNA portion between the two operators. From the kinetics of the loop formation mediated by the GalR and HU proteins, Lia *et al.* [2] have deduced a looping free energy of $12\ k_BT$, that with respect to our values should correspond to a kink angle of more than 150° or to an end-to-end extension smaller than $r=10$ nm. Another significant change in the cyclization probability is that the stiffness loss induced by κ reduces the most probable loop length from 500 bp for a nonkinked DNA to 340 bp and 190 bp (from the numerical calculation) or to 300 bp and 150 bp (from SPA) for kinks of respectively 150° and 120°. Note that for a kinked loop with an end-to-end extension r the minimal length L_0 corresponding to the rigid rodlike configuration fulfills the relation $L_0\sin(\kappa/2)=r$ and therefore it is of ≈42 bp for $\kappa=90°$ instead of ≈30 bp for $\kappa=180°$. For $\kappa=90°$ the most probable loop length is the rigid *kinked* rodlike configuration of the two half-DNA portions. To catch both kink and protein bridge effects in a simple formula, we have calculated the cyclization factor with the extension of Shimada and Yamakawa formula for a kinked closed loop of length $(L+2r)$. This approach is similar to what was suggested in 1991 by Brenowitz *et al.* [7] to interpret their experimental data, i.e., to directly consider the protein as part of the length of the loop. A linear fit [26] of the bending energy (23) for the optimal closed configuration ($r=0$) in the presence of a kink (expressed in degrees), $\beta\Delta E(r=0, L, \kappa)\approx(-7.1+0.1155\kappa)/(L/A)$, is shown in Fig. 7. It gives the following approximated formula for the closure factor as a function of the protein size r, the length L of the DNA, and the kink angle κ

$$J_{\mathrm{approx}}(L,r,\kappa)=C_{\mathrm{SY}}(L+2r)\exp\left[\frac{7.1-0.1155\kappa}{(L+2r)/A}\right], \quad (25)$$

where C_{SY} is given in (16). Notice the integration step (8) has been skipped since it does not make any significant difference. Formula (25) allows us to obtain a simple prediction for the loop probability in the presence of a kink in the middle of the sequence and a finite separation between the extremities. As shown in Fig. 6 this formula (gray lines) is in good agreement with the loop probability obtained with the exact calculation of the saddle point energy of the open configuration (full line). In particular, Fig. 6 shows that for kink angles in the range $90°<\kappa<150°$ this simple formula works remarkably well for lengths L larger than about $5r$, that is 150 bp for $r=10$ nm. For smaller lengths the optimal configuration is more a rigid rodlike configuration and cannot be approximated by a closed loop. Similar simple formulas that includes a kink angle κ and a finite end-to-end distance r in an effective way have also been written down by Rippe or Ringrose to fit their Brownian dynamics simulation [19] or experimental data [29], but these formulas contain a parameter that must be fitted for each values of r and κ from the data points (Fig. 6).

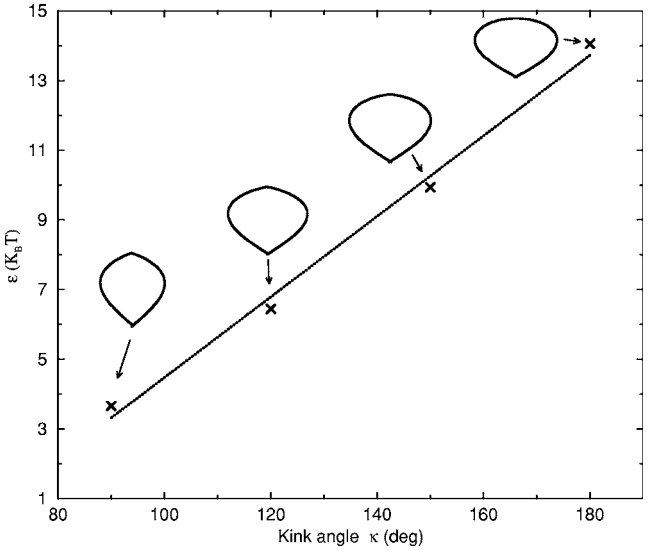

FIG. 7. The "×" points: saddle point energy in units of L/A (that is, $\varepsilon=\Delta E\times L/A$) for the saddle point configurations (also displayed in the figure) with kink angles of $\kappa=90°, 120°, 150°$, 180°. Dotted line: linear interpolation used in formula (25).

V. CONCLUSION

We performed both numerical and analytical calculations of the closure factor J, even in the presence of a protein bridge and of a protein-mediated kink. More precisely we have numerically calculated the path integral of the WLC polymer model under the constraints of a fixed end-to-end distance r and a kink κ in the middle of the DNA portion. Analytically we have extended the Shimada and Yamakawa saddle point approximation [14] to the case of a bridged and kinked loop. We have seen that the formation of DNA loops is significantly sensitive to the size of the protein bridge when this size r is more than 10% of the loop length L, that is, 300 bp (or 100 nm) for a typical protein bridge size of $r=10$ nm. To give an example, the closure factor for a 100 bp DNA segment increases from $J(100\ \mathrm{bp}, 0)\approx10^{-11}\ M$ to $J(100\ \mathrm{bp}, 10\ \mathrm{nm})\approx10^{-9}\ M$. Correspondingly, looping free energy decreases from $\Delta G(100\ \mathrm{bp}, 0)=24\ k_BT$ to $\Delta G(100\ \mathrm{bp}, 10\ \mathrm{nm})=13\ k_BT$. A kink ranging from 150° to 90° produces a significant increase of J for DNA fragments of lengths up to about 2500 bp. For instance, the closure factor for a 100 bp DNA segment with a protein bridge $r=10$ nm and a kink of 90° is $J(100\ \mathrm{bp}, 10\ \mathrm{nm}, 90°)\approx10^{-4}\ M$, and the corresponding looping free energy $\Delta G(100\ \mathrm{bp}, 10\ \mathrm{nm}, 90°)\approx2\ k_BT$. A kink also changes the most probable

loop length from 500 bp (no kink) to about 175 bp for a kink of 120°, going to the rigid *kinked* rodlike configuration for smaller κ values. This is an interesting mechanism because the loop lengths implied in *in vivo* DNA processing by proteins spread on a large lengths range. Our results were compared to previous analytical approximations (in particular the results of the Gaussian model, the Wilhelm and Frey expansion [16] and the Shimada and Yamakawa formula [14]) and numerical calculations (in particular the Monte Carlo simulations data obtained by Podtelezhnikov and Vologodskii [17], and the Brownian dynamics simulations data obtained by Langowski *et al.* [4,18,19]) as well as experimental results [1,2,7] (e.g., the ones obtained by Lia *et al.* on the looping dynamics mediated by the Gal and HU proteins). Finally a simple formula (25) including both the protein bridge and kink effects has been proposed. This formula has the advantage of not containing adjustable parameters with respect to the existing formulas that include both these effects [19].

Still, many effects omitted in this work can be included without significant changes in the numerical algorithm. The kink we considered is actually permanent (that is, not thermally excited), site specific (at half-length), and rigid (κ fixed). Although this rigidity seems relevant to most protein bindings to DNA at first glance [30], it was pointed out in the works of Yan *et al.* [20,22] that kinks may also be semiflexible (exhibiting higher or lower rigidities than the bare DNA) or even fully flexible [21,31]. For instance, the HU/DNA complex was recently observed to be very flexible under specific experimental conditions [32]. Flexible hinges were also stated to occur along the DNA due to the opening of small denaturation bubbles [21,33], such as the one needed by HU to fit in the double helix [2]. Such flexibility could be taken into account [30] in our model. These kinds of defects could also be thermally activated, occurring at multiple nonspecific sites along the DNA [20,22,30,31,34]. Both effects may be included in our model. Note that using effective persistence lengths could turn out convenient, despite little information about the kink properties (number, location, rigidity, etc.). Actually this would be equivalent to study DNA stiffness loss due to sequence effects [35] by cutting WLC in different stiff fragments, depending on the CG or AT content of the whole sequence to cycle. The same approach may allow an approximative study of the DNA polyelectrolyte nature too [36]; otherwise, electrostatic potential has to be included in WLC energy. Twist elasticity leads to slight modifications of the quantum analog we used although requiring some care [37]. This is expected to play an important role in looping, especially when specific alignment of the loop extremities are required. Finally cyclization dynamics could be modeled using a simple two-states model where DNA is "closed" or "opened," that is, cycled or not. Such study relies on the (statics) cyclization factor we computed in this article [38]. Direct comparison to experimental lifetimes measures would be possible [1,2].

ACKNOWLEDGMENTS

The authors would like to thank D. Chatenay, R. Monasson, and F. Thalmann for useful discussions and a critical reading of the manuscript.

[1] L. Finzi and J. Gelles, Science **267**, 378 (1995).
[2] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. A. Lewis, S. Adhya, and L. Finzi, Proc. Natl. Acad. Sci. U.S.A. **100**, 11373 (2003).
[3] J.-F. Allemand, S. Cocco, N. Douarche, and G. Lia, Eur. Phys. J. B (to be published).
[4] K. Rippe, P. H. von Hippel, and J. Langowski, Trends Biochem. Sci. **20**, 500 (1995).
[5] D. Shore, J. Langowski, and R. L. Baldwin, Proc. Natl. Acad. Sci. U.S.A. **78**, 4833 (1981).
[6] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **102**, 5397 (2005).
[7] M. Brenowitz, A. Pickar, and E. Jamison, Biochemistry **30**, 5986 (1991).
[8] H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions* (Springer-Verlag, Berlin, 1997).
[9] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World Scientific Publishing Co., Singapore, 2004).
[10] A. Hanke and R. Metzler, Biophys. J. **85**, 167 (2003).
[11] O. Kratky and G. Porod, Recl. Trav. Chim. Pays-Bas **68**, 1106 (1949).
[12] H. E. Daniels, Proc. - R. Soc. Edinburgh, Sect. A: Math. **63**, 290 (1952).
[13] W. Gobush, H. Yamakawa, W. H. Stockmayer, and W. S. Magee, J. Chem. Phys. **57**, 2839 (1972).
[14] J. Shimada and H. Yamakawa, Macromolecules **17**, 689 (1984).
[15] A. Balaeff, L. Mahadevan, and K. Schulten, Phys. Rev. Lett. **83**, 4900 (1999).
[16] J. Wilhelm and E. Frey, Phys. Rev. Lett. **77**, 2581 (1996).
[17] A. A. Podtelezhnikov and A. V. Vologodskii, Macromolecules **33**, 2767 (2000).
[18] H. Merlitz, K. Rippe, K. V. Klenin, and J. Langowski, Biophys. J. **74**, 773 (1998).
[19] K. Rippe, Trends Biochem. Sci. **26**, 733 (2001).
[20] J. Yan and J. F. Marko, Phys. Rev. E **68**, 011905 (2003).
[21] J. Yan and J. F. Marko, Phys. Rev. Lett. **93**, 108108 (2004).
[22] J. Yan, R. Kawamura, and J. F. Marko, Phys. Rev. E **71**, 061905 (2005).
[23] S. Sankararaman and J. F. Marko, Phys. Rev. E **71**, 021911 (2005).
[24] J. Samuel and S. Sinha, Phys. Rev. E **66**, 050801(R) (2002).
[25] R. B. Sidje, ACM Trans. Math. Softw. **24**, 130 (1998). The EXPOKIT library is available for download at URL http://www.maths.uq.edu.au/expokit
[26] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1992).
[27] C. Moler and C. van Loan, SIAM Rev. **45**, 3 (2003).

[28] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover Publications Inc., New York, 1970). The elliptic integrals of first and second kinds respectively read in Legendre definition $K(\psi,x^2)=\int_0^{\psi}dy/\sqrt{1-x^2\sin^2 y}$ and $E(\psi,x^2)=\int_0^{\psi}\sqrt{1-x^2\sin^2 y}dy$.

[29] L. Ringrose, S. Chabanis, P.-O. Angrand, C. Woodroofe, and A. F. Stewart, EMBO J. **18**, 6630 (1999).

[30] Y. O. Popov and A. V. Tkachenko, Phys. Rev. E **71**, 051905 (2005).

[31] P. A. Wiggins, R. Phillips, and P. C. Nelson, Phys. Rev. E **71**, 021909 (2005).

[32] J. V. Noort, S. Verbrugge, N. Goosen, C. Dekker, and R. T. Dame, Proc. Natl. Acad. Sci. U.S.A. **101**, 6969 (2004).

[33] P. Ranjith, P. B. Sunil Kumar, and G. I. Menon, Phys. Rev. Lett. **94**, 138102 (2005).

[34] B. Chakrabarti and A. J. Levine, Phys. Rev. E **71**, 031905 (2005).

[35] P. Nelson, Phys. Rev. Lett. **80**, 5810 (1998).

[36] R. Zandi, J. Rudnick, and R. Golestanian, Phys. Rev. E **67**, 021803 (2003).

[37] C. Bouchiat and M. Mézard Phys. Rev. Lett. **80**, 1556 (1998); V. Rossetto and A. C. Maggs, *ibid.* **88**, 089801 (2002) (comment); C. Bouchiat and M. Mézard, *ibid.* **88**, 089802 (2002) (reply).

[38] S. Jun, J. Bechhoefer, and B.-Y. Ha, Europhys. Lett. **64**, 420 (2003).