Eur. Phys. J. E **19**, 293–302 (2006)
DOI 10.1140/epje/i2005-10073-y

THE EUROPEAN PHYSICAL JOURNAL E

Loops in DNA: An overview of experimental and theoretical approaches

J.-F. Allemand, S. Cocco[a], N. Douarche, and G. Lia

CNRS - Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75005 Paris, France

Received 11 August 2005 /
Published online: 23 March 2006 –

Abstract. DNA loop formation plays a central role in many cellular processes. The aim of this paper is to present the state of the art and open problems regarding the experimental and theoretical approaches to DNA looping. A particular attention is devoted to the effects of the protein bridge size and of protein induced sharp DNA bending on DNA loop formation enhancement.

PACS. 87.14.Gg DNA, RNA – 87.15.La Mechanical properties – 87.15.He Dynamics and conformational changes

1 Importance of loop formation in biological processes

DNA processing is not always the result of a single protein going to a single target sequence. It can also be the result of a single or multimeric proteins connecting two distant sites on DNA thus forming a loop (see Fig. 1). Indeed protein-mediated DNA loops play a central role in many biological processes of DNA life [1,2].

Looping is widely used in gene regulation, in particular various genetic events of transcriptional regulation (RNA synthesis), including activation but also repression, can involve formation of such DNA structure. One of the most known example of transcription regulation which is accomplished through DNA looping is the regulation of the *gal* operon (see Fig. 1) [3]. In this genetic system transcription can be blocked by loop formation between two distinct sites of defined sequences distant of 113 base pairs (bp) that encompass promoter region. When a sugar, the galactose, is present in the surounding medium, looping is inhibited, and transcription can occur. Other typical examples of gene regulation by looping are the *ara*, *lac* and *deo* operon in *E. Coli*, the lysogenic-to-lytic switch in phage λ, the protein/DNA complex between c-Myb and C/EBP in the mim-1 promoter, and the protein p53, to name just a few [4–9].

Nucleoprotein interaction resulting in a loop formation can also transiently occur between proteins and the RNA polymerase. In this case, looping regulates the quantity of RNA produced by RNA polymerase. Example are the NtrC system (nitrogen regulatory protein C) in *E. Coli*, the Jun activator with the pre-initiation

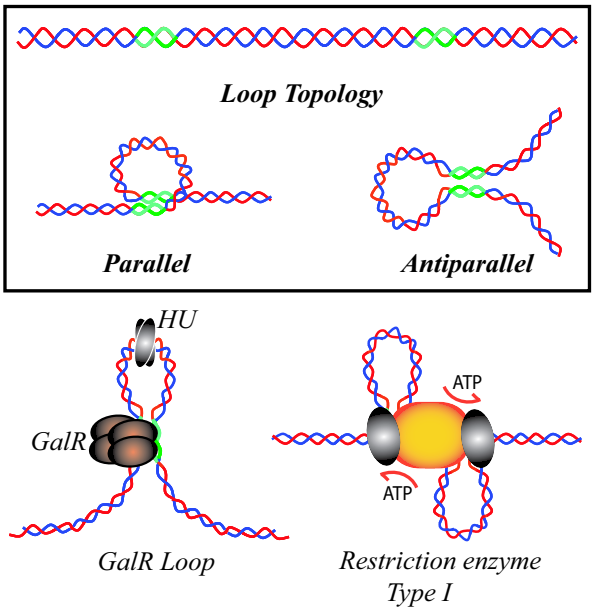

Fig. 1. Upper panel: looping between two distant sites can lead to different forms. Two major types of conformations are the parallel and antiparallel ones. Energy cost is different for every conformation. Angles may vary, local bends might be present inside the loop, phasing of the two helical pitches may also be an important parameter. Lower panel: In *E. Coli* a tetramer of galR, in the presence of the structural protein HU, forms a loop between two sites 113-bp distant. This loop inhibits the transcription of the gene whose origin of replication lies inside the loop. Type-I restriction enzymes bind to a specific site and, using ATP as an energy source, extrude loops. Loop size before cutting depends on conditions. Initiation of the loop, that at first glance seems energy consumable, is still an unknown phenomenon.

[a] e-mail: simona.cocco@lps.ens.fr

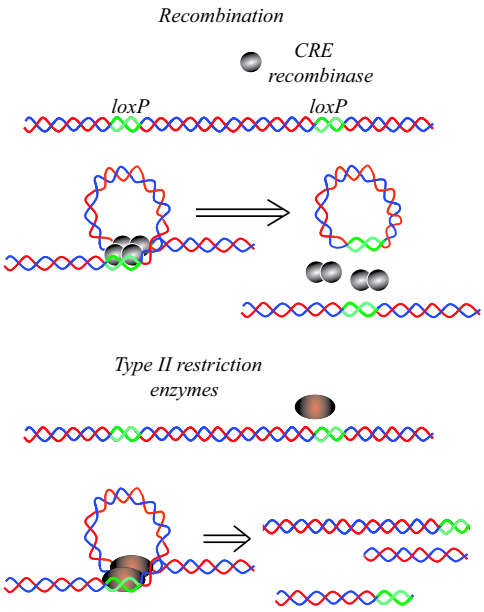

Fig. 2. Cre recombinase is one of the most well known of the site-specific recombinases. This family of enzymes binds to two specific sites, the loxP sites (34-bp long) in the case of Cre, and excise the DNA in between after the formation of a transient loop, called synapse. Loop formation may enhance the activity of the well-known type-II restriction enzymes, the ones used in gene cloning to cut DNA at specific positions.

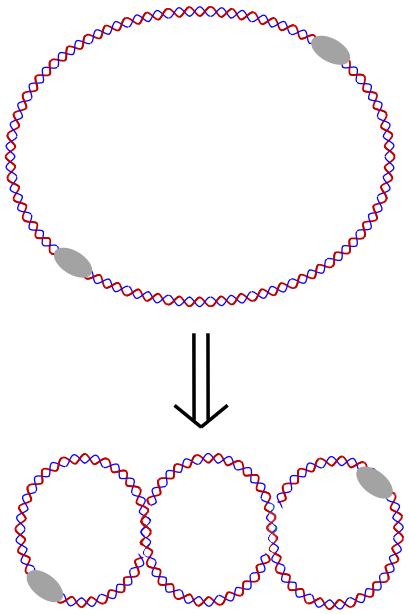

Fig. 3. DNA isolated from living organisms always present some torsional strain, *i.e.* torque. This is required for DNA processing and this strain is highly regulated by various enzymes in the cell. This torsional stress leads to the formation of structures called plectonemes. In the figure two turns have been added to the relaxed, *i.e.* without torsional stress, circular DNA drawn on top giving the lower form. These plectonemes will strongly affect the probability and kinetics of loop formation between two or more specific sites (bold ellipsoids). In this case superposition of the two sites, necessary for loop formation, requires slithering.

complex and the eukaryotic silencers and enhancers, like the human *β-globin* locus [10–13].

DNA looping in these processes may involve loop sizes from about fifty to several thousands of base pairs. The proteins that activate transcriptional factors are located 50–200 bp upstream the promoter, while enhancers and silencers sequences are several thousands base pairs upstream or downstream of theirs promoters; the lambda repressor switch can act looping a sequence of roughly 2300 bp; the lactose repressing act by looping a 305-bp sequence between the two DNA binding sites called O_1 and O_2 or a sequence of 76 bp between O_1 and a third site O_3; the two binding site of the Gal repressor are at a distance of 113 bp.

Such a loop between two sequence-specific sites is a common event in site-specific recombination [14,15], a process by which the DNA fragment between the two sites can be excised (or inserted in an inverse process used to transpose DNA fragments, see Fig. 2). The most popular example is the Cre protein that recognizes two 24-bp sites, named loxP, makes loop between them and deletes the DNA fragments between those sites. When the surrounding DNA is correctly designed this reaction allows switching on a gene. This is rutinely used for artificially switch a gene on or off *in vivo* [16,17].

Even the proteins used in DNA cloning, the type-II restriction enzymes, that recognize a sequence and cut the DNA at a specific position in the same region, can be involved in DNA looping (see Fig. 2). Some of them, like SfiI or NgoMIV [18, 19], have an enhanced cutting activity when bound to two target sites (in the same or in two different DNA molecules) instead of just one. In these cases looping is likely to happen because of random collisioning.

DNA looping is then a very common phenomenon, that generally is generated passively. Nevertheless, loop can also be formed by more active processes.

In fact some motors use ATP to form loop (see Fig. 1). Type-I restriction enzymes are a good example [20, 21]. Those enzymes bind to a single specific sequence and then start extruding one or two loops from this sequence until, for still an unknown reason, it cuts the DNA at some blocking point. As a consequence the loop size before the cutting event displays a large distribution.

At a higher level of organization chromosomes are also organised in topologicaly isolated loops (see Fig. 3). In *E. Coli* the average size of these loops has been measured at about 10 kb (that is 1000 bp) [22]. In eukaryotes some structural proteins interact with chromatin (DNA compacted by histone proteins) organizing it in very large loop (between 30 and 140 kb) [23].

2 Experimental approaches

Experimentally various approaches has been developed to study DNA looping. Since loops will affect DNA shape,

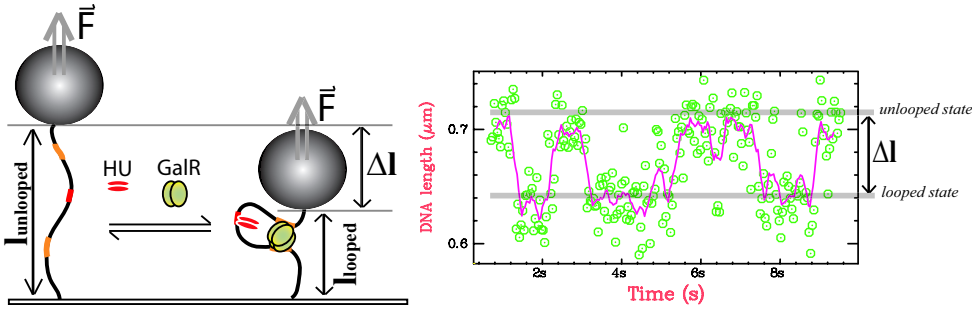

Fig. 4. Illustration of the experiment of Lia *et al.* A magnetic bead submitted to a magnetic-field gradient is submitted to a force (left panel). The force will disfavor the closed loop form. If F is the force and δl the change in extension due to loop formation: when $F \times \delta l$ is equal to the energy of the loop formation, then the proportion of the time spend in the closed or open loop is equal. This two-level system then alternates between the two conformations as illustrated on the time trace on the right. Without the force the loop would remain in the closed state almost all the time. As described in [34] supercoiling (torsion) is important in this process. The experimental measured value gives $\delta l = 55\,\text{nm}$. This distance is bigger than the distance between the sites incating an antiparallel conformation. The force where the loop is formed during 50% of the time is 0.9 pN thus leading to an energy of $-12\ k_\text{B}T$ (where $k_\text{B}T$ is the thermal energy) for the loop formation in the experimental conditions. Measuring the average duration of each state leads to the kinetic constants k_loop and k_unloop. These depends on the force with an Arhenius factor $\exp(-\frac{F\times\delta l}{k_\text{B}T})$. Using this factor one may extrapolate the looped time to 21 s and the unlooped time 0.1 ms at zero force.

it is possible to detect loop formation in gel electrophoresis (see for instance [24]). As it brings two sites in close distance to each other, loop formation can also be detected by fluorescence energy transfer (FRET) [25–27] between two fluorophores located close to the two sites involved in looping. Other bulk methods imply the use of additional proteins [28–30]. The most known method uses the ligase [28], an enzyme that is able to ligate two DNA extremities. To perform this operation it requires that the two extremities are placed at the same position, *i.e.* a loop is formed. In fact it also requires an angular phasing between the two extremities because of the requirement of having a number of helical pitches in agreement with the ~ 10.5 bp turn for the circular DNA. Recombination efficiency [29], detected by the efficiency of the excision of the DNA fragment between the two sites, can also be used for that purpose. These techniques, can be very sensitive but are subjected to protein concentration and activity [31]. Of course high-resolution microscopy techniques, like electron microscopy or AFM [32], are useful tools to image loops.

Single-molecule techniques [25, 33–35] have also been applied to this problem. The use of these techniques made the real-time observation of the looping possible, allowing information about the kinetics of the looping. Moreover, with the micromanipulation techniques, a new control parameter is introduced: the force (see Fig. 4). Applying a force will disfavor the formation of the loop. So the higher the force the smaller the time spent in the closed-loop conformation and the longer the time spent in the open-loop form. In fact the mechanical constrains allow the modification of the kinetics barriers of the looping reaction without modifying the physico-chemistry parameters. Simple thermodynamics allows then to have access to some kinetics, thermodynamics and structural parameters [34].

3 DNA looping: the problematics

The study of DNA loops has been an active field of research since the eighties. It was always characterized by a very strong interaction between experiments and theory.

The experimental and theoretical main goal is to calculate the probability that two sites of the same DNA molecule will meet by looping. For experimental convenience the looping probability is commonly expressed in terms of the local molar concentration of one site with respect to the other one (j, or j_M in $\text{M} = \text{mol} \cdot \text{L}^{-1}$). j_M is also related to the free energy of loop formation. The loop stability depends on the comparison of j_M with the dissociation constant of the protein-protein or protein-DNA interaction that stabilize the loop formation.

The difficulty for the DNA to make a loop are twofold, and depend on the loop length. The formation of a small loop gives rise to an elastic problem because of the bending energy, while the formation of a large loop leads to an entropic problem because far apart sites have to come into collision.

A central point in the past studies of DNA looping has been to derive the probability distribution of loop formation as a function of the loop length. The most probable loop size for a given DNA molecule is encountered when the elastic energy starts to decrease (as the length increases) and the entropic loss is not too large yet; this corresponds to about 500 bp for a bare DNA. However, as we have underlined in the introduction, observed loop lengths *in vivo* spread over a large range of sizes. A central question is therefore to determine what kinds of protein action can modify the most probable loop length to make a given DNA processing possible.

4 On the cyclisation factor

4.1 Some historical background

Loop formation was first studied by Shore *et al.* through cyclisation or ring closure experiments [36]. These experiments are the simplest ones to analyse for two reasons. First, a linear DNA is entirely cyclized in the loop; secondly, the ligase proteins are not involved in the loop formation itself but only stabilize the loop once it is formed. In these experiments the loop length is varied from 130 to 4300 bp from the preparation of different linear DNA. These experiments have been first used to determined the elastic properties of DNA, *e.g.*, the persistence length, the twist modulus, and the pitch of the double helix.

From a theoretical point of view the theory of elasticity of a stiff chain was developed after the 1950s. The Wormlike Chain (WLC) of Kratky and Porod [37] turned out to be a very good model in this context, and was widely studied. The absence of exact analytical solution for the end-to-end distribution of the WLC model pushed workers in the field to develop various expansion approaches [38–51].

Shimada and Yamakawa applied in 1984 [42] the calculation of the end-to-end distribution of the WLC model to DNA cyclisation and compare with Shore *et al.* experiments. They derived a simple formula for j_{M} with numerical coefficients, which can be directly used by experimentalists. In the same period Shore *et al.*'s experimental results and Shimada and Yamakawa theoretical results have also been compared to Monte Carlo simulations by Hagermann *et al.* [52], Levene and Crothers *et al.* [28, 53, 54] and Frank-Kamenetskii *et al.* [55].

Very recently cyclisation experiments and theory have known a renewal of interest, stemming from Cloutier and Widom's results on an unexpectedly high cyclisation probability for small DNA molecules (94–116 bp) [56]. Though these experimental findings were merely an artifact due to an erroneous concentration of ligases [30], they have triggered a lively debate on the issue. Several theoretical works have been developed based on the idea that for short DNA the formation of a few base pairs of single-stranded DNA can facilitate loop formation; single-stranded (ss) DNA is indeed very flexible and therefore can provide a "flexible hinge". The WLC model has therefore been modified to include the possibility for formation of such thermally fluctuating bubbles [57–61]. In this context the sequence should play a crucial role [61]. The debate came to a conclusion with the work of the group of Vologodski [30, 62] who experimentally reinvestigated the cyclization of 100–130 bp DNA, and showed that the j_{M} factor is in full agreement with theoretical prediction and Monte Carlo simulations [30, 63] based on the WLC model.

4.2 Theoretical approaches to cyclisation probability

The Wormlike Chain model [37] has been successfully used to describe the elasticity of a single DNA molecule under a force. The correctness of the theoretical prediction has been verified by the experimental force extension curve obtained by stretching of single DNA molecules [64–66].

Numerical calculations of the WLC partition function, combined with expansions at small and high forces have provided simple formulas, widely used by experimentalists [67]. The WLC model is therefore a good model to calculate the loop formation probability too. This calculation is very difficult to perform because it amounts to the evaluation of a path integral, the action of which is the bending energy of the polymer with closed end. The latter condition introduces a global constraint which can be expressed through the introduction in the action of an imaginary force field (Lagrange multiplier vector). The calculation of the path integral with this imaginary force field is in fact identical to the one of the WLC model with a real force, with the additional step that the path integral is then integrated over the Lagrange multiplier values. A second difficulty comes from the fact that loop lengths of the order of 100 bp are too short for the thermodynamic limit to hold in the calculation of the constrained partition function. As we have already discussed two effects contribute to the difficulty that DNA has to make a loop, and their relative importances depend on the length of the molecule.

For long molecules, of at least about 10 persistence lengths, the contact of the two ends of the molecules is difficult due to the entropy loss expressing the decrease in the number of configurations accessible when the loop is formed. The looping probability can be simply calculated with the Gaussian polymer model and reads

$$j_{\mathrm{M}}(L) = 1.66 \left(\frac{3}{4\pi L A}\right)^{3/2}, \tag{1}$$

where $A = 50\,\mathrm{nm}$ is the DNA persistence length. Note that $1\,\mathrm{nm}^{-3} = 1.66\,\mathrm{M} = 1.66\,\mathrm{mol}\cdot\mathrm{L}^{-1}$.

Short molecules (up to some persistence lengths), are much more rigid. The energetic cost to bend these molecules and put in contact the two ends is very large. Optimally closed configuration, with a ring shape, have a bending energy given by

$$\Delta E_{\mathrm{c}}(L) = AL/R^2 = 4\pi^2 \frac{A}{L}, \tag{2}$$

where $R = L/(2\pi)$ is the ring radius. The real configuration is indeed not a circle if, as is generally the case, the tangent vector of the first and the last bases are not aligned along the same direction. The optimal configuration and its energy can be calculated in the continuous limit because they are related to the solution of the Mathieu equation for the motion of the pendulus [41]. The resulting bending energy is somewhat smaller than the ring one (2), and reads

$$\Delta E(L) = 14.055\ \frac{A}{L}. \tag{3}$$

The optimal configuration for a 100-bp molecule is shown in Figure 5 (inset).

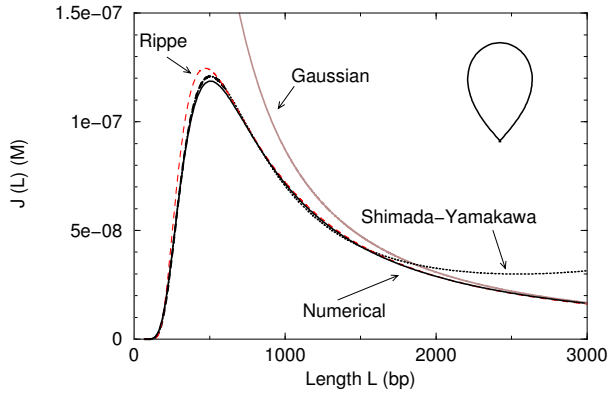

Fig. 5. Cyclization factor as a function of the length of DNA with the Gaussian model (gray line), WLC model and the Shimada-Yamakawa formula (dotted black line) [42], WLC model and a numerical calculation (full black line) [69], Rippe-Ringrose approximation (thin gray dashed line) [29,73]. The most probable length is 500 bp. Inset: the low-bending-energy configuration of a closed loop.

The contribution of quadratic fluctuations around the optimal configuration to the cyclisation factor has been calculated by Shimada and Yamakawa with the outcome

$$C_{\rm SY}(L) = \frac{1.66 \times 112.04}{A^3} \left(\frac{L}{A}\right)^{-5} e^{0.246\ \frac{L}{A}}, \qquad (4)$$

leading to the final expression for the cyclisation factor:

$$j_{\rm M}(L) = C_{\rm SY}(L)\ e^{-\Delta E(L)}. \qquad (5)$$

Expression (4) for $C_{SY}(L)$ has been obtained through a very involved calculation, in part analytical and in part numerical. While the saddle point configuration and its related elastic energy are easy to re-derivate [68,69], the calculation of the fluctuations is tricky, mainly because the fluctuations are calculated in the discrete space while the saddle point configuration is calculated in the continuous space limit. It would be very interesting to have an alternative derivation for this entropic cost factor. A similar calculation for the quadratic fluctuations was recently done by Kulić *et al.* around the optimal configuration of a looped DNA stretched at high force [70]. The looping free energy can be separated into a bending and an entropic contributions:

$$\Delta G(L) = \Delta E(L) - \Delta S(L). \qquad (6)$$

The entropic part $\Delta S(L)$ (in units of $k_{\rm B}T$) in the looping formation is given by $\Delta S(L) = -\ln[C_{\rm SY}(L) \times dv/1.66]$ with $dv = \frac{4}{3}\pi r^3$ (in $\rm nm^{-3}$) the interaction volume in which the ring closes. As a numerical illustration, for a 100-bp loop, the elastic cost is of $\Delta E(100\,{\rm bp}) = 21\ k_{\rm B}T$; the entropic loss is of $\Delta S(100\,{\rm bp}) = -3\ k_{\rm B}T$ for an interaction radius of $r = 1\,{\rm nm}$ (typical dimension of a binding site). Note that the entropic loss strongly depends on the interaction volume. For example if one considers the typical protein linear dimension of $r = 10\,{\rm nm}$ a contribution $3 \times \ln 10 = 7\ k_{\rm B}T$ has to be added to the entropy $\Delta S(L)$. In Figure 5 we show the Gaussian polymer formula (1) valid for long loops (full gray line) and the short-loop Shimada-Yamakawa formula (5) (full black line).

Figure 5 also shows the curve obtained with two alternative approaches: the Brownian dynamics simulations [71,72] (dotted line) and the numerical calculation of the path integral under the global constraint of zero end-to-end distance [57,61,69]. The cyclisation probability goes to zero for small and long molecules and exhibits a maximum in between. This maximum is located at about 3.3 persistence lengths (500 bp), see Figure 5.

As DNA is made of two strands, it not only has a bending modulus but also a torsional modulus. This was included in the Shimada-Yamakawa framework [42]. In fact cyclisation also requires an angular phasing between the two extremities for the number of helical pitches to agree with the $\sim 10.5\,{\rm bp}$ turn of circular DNA. The linking-number-dependent $j_{\rm M}$ factor increases and oscillates up to about 3 persistence lengths, then decreases monotonically.

As shown in Figure 5 the Shimada-Yamakawa saddle point calculation works very well for lengths up to 1500 bp, while the Gaussian approximation is valid for DNA having more than 3000 bp. An approximate formula including both short- and large-length-scale behavior has been obtained by combining the Gaussian behavior at large length scales with an exponential decay at short length scales [29,73,74]. This formula is shown for comparison in Figure 5 (gray dashed line). The Shimada-Yamakawa approach is so far the best expansion in the short-stiff chain limit; Wilhelm and Frey have also developed an expansion in this limit [48], but the validity domain of their approach is limited to lengths of the molecule up to about 300 bp (two persistence lengths). Highest-order corrections have been calculated by perturbation theory but do not substantially improve Wilhelm and Frey's result [49–51]. On the other hand, expansions around the long-flexible limit have also been developed by Daniels in 1952 [39]. The first-order correction to the Gaussian chain works well for molecule lengths larger than about 1000 bp (10 persistence lengths). Other approaches consist in relaxing the inextensibility constraint of the chain [44–47] but so far, did not succed in catching the rigid rod-like regime in a systematic way (Ha *et al.* works give an accurate estimate for the most probable end-to-end, but not for the whole distribution for instance).

5 Protein-mediated loops

5.1 Some historical background

A step forward in complexity and biological interest was the study of the protein-protein interactions mediated by DNA loops [1]. The aforementioned *lac*, *gal* [33, 34, 75] and recombinase [29] *in vitro* experiments were carried out in the last decade, through the preparation of DNA fragments with a distance between the *lac* operators of about 300 and 100 bp in the first two cases, and a distance

between the recombinase binding sites variyng from 74 to 1500 bp in the third case.

Protein-mediated loop experiments can be more complex than cyclisation experiments because of the presence of both protein-protein and DNA-proteins interactions. Each interaction has its own affinity constant, and competitions between these interactions with different possible populations of complexes are possible [1, 33, 75]. An exemple is the GalR-DNA complex shown in Figure 1, and made of two dimers bound to each other to form a tetramer and to DNA. If, for exemple, the two DNA operators carry, respectively, a dimer and a tetramer, the loop cannot be formed.

Proteins may in addition impose geometrical constraints to DNA (see for instance the parallel or antiparallel configurations in Fig. 1) [1]. These constraints have to be taken into account at different stages of simplification in theoretical approaches.

Theoretical investigations have mainly focused on how proteins can enhance the looping probability [72]. Three important factors have been studied with Brownian dynamic and Monte Carlo simulations by the Langowski and Vologodskii groups since 1992 [31, 71, 76].

First a purely geometrical factor is relevant for small, *i.e.* up to about 300 bp, DNA loops: the protein has a linear size r, typically of the order of 10 nm. This is a non-negligible distance with respect to the DNA persistence length ($A = 50$ nm). As firstly pointed out in the work of Brenowitz *et al.* the protein can act as a bridge in the loop formation. The loop gets clamped when the end-to-end distance of the molecule is about r, and not zero. The bending energy in the presence of this protein bridge is therefore reduced with respect to a closed loop.

Secondly, the presence of additional proteins that sharply bend the DNA fragment can also facilitate the loop formation (see the HU protein in Fig. 1). Proteins can also produce a kink on the DNA in an indirect way by twisting it. This twist can give rise to a denaturation bubble and therefore a zone of high flexibility on the DNA segment. This action has been modeled in Brownian dynamical simulations [71, 72, 76] by introducing a kink, of 120° or 80°, in the middle of the chain. The presence of a kink has been shown to greatly enhance the looping probability for molecular lengths of less than roughly 500 bp.

Thirdly when a torque is applied, DNA may also buckle in much the same way as an elastic tube, and create plectonemes. These structures, which also exist *in vivo* may strongly affect the end-to-end collision probability, as first pointed out by Vologodskii *et al.* [31]. The presence of superhelical conformation can enhance the loop probability for loop lengths up to about 1000 bp [72].

The geometrical and topological constraints due to the proteins, in general, make analytical treatements more difficult. A first, numerical method, has been proposed by Balaeff *et al.* to calculate the bending energy given precise boundary conditions obtained from the knowledge of the LacR protein structure [77–79]. In these calculations the twist and electrostatic energies are included but the entropic contribution to the free energy is neglected. A second method recently developed by Yan *et al.* [57] and Douarche *et al.* [69], consist in the numerical calculation of the WLC path integral with the r end-to-end distance constraint. It allowed the authors to calculate the loop probability with a protein bridge, and to impose particular boundary conditions (for example, the parallel or anti-parallel confguration in Fig. 1) [57], or a kink in the middle of the sequence. A third approach that we detail in the next section is the extension of the Shimada-Yamakawa calculation to the case of an open and kinked loop.

5.2 Theoretical approaches for the effects of protein bridge and kink at half-length

A protein bridge can easily be accounted for Shimada and Yamakawa saddle point calculation [69]. The optimal configuration for a 100-bp (34 nm) DNA with an end-to-end separation $r = 10$ nm is shown in Figure 6. For the entropic contribution C due to quadratic fluctuations around the optimal configuration a simple and surprisingly accurate estimate is obtained by adapting Shimada-Yamakawa formula (4), and approximating

$$C(L, r) = C_{\mathrm{SY}}(L + 2r). \tag{7}$$

This approximation can be understood if one compares (see Fig. 6) the open configuration with the closed configuration with a length increased by twice the end-to-end separation. The closing part has no elastic cost because it is essentially not curved, and contributes to fluctuations only. The choice of the factor 2 can be explained in geometrical terms [69].

As shown in Figure 6 the cyclisation factor for a 100-bp DNA is enhanced by a factor 10^2 in the presence of a protein bridge of 10 nm between the DNA extremities. For a DNA length more than about 10 times the end-to-end separation the curve goes to the Shimada-Yamakawa closed loop formula (5). Figure 6 also shows a good agreement between the saddle point extension of Shimada and Yamakawa formula (top panel) numerical calculation (bottom panel), and the Brownian dynamics (open squares) and Monte Carlo simulation (filled circles) points. Note that the Brownian dynamics points are obtained from a model which includes twist rigidity and electrostatic effects. The free energy of loop formation is shown in Figure 6 (left). A kink in the middle of the molecule can be treated in an analogous manner. The saddle point kinked configuration is shown in Figure 7 (top). A kink of 90° increases the loop probability by a factor 10^5. The most probable loop length decreases also from 500 bp to the minimal length of 35 bp for a kink of 90°. The free energy of loop formation is shown on the left of Fig. 7. A loop of 100 bp with a kink of 150° in the middle of the sequence and an end-to-end distance $r = 10$ nm has a free energy cost $\Delta G = 9\ k_{\mathrm{B}}T$ to be compared with the value of $\Delta G = 12\ k_{\mathrm{B}}T$ found from the loop formation mediated by the GalR and HU proteins [34] (see Fig. 4).

An even simpler approach is to directly calculate the cyclization factor with the Shimada-Yamakawa formula on

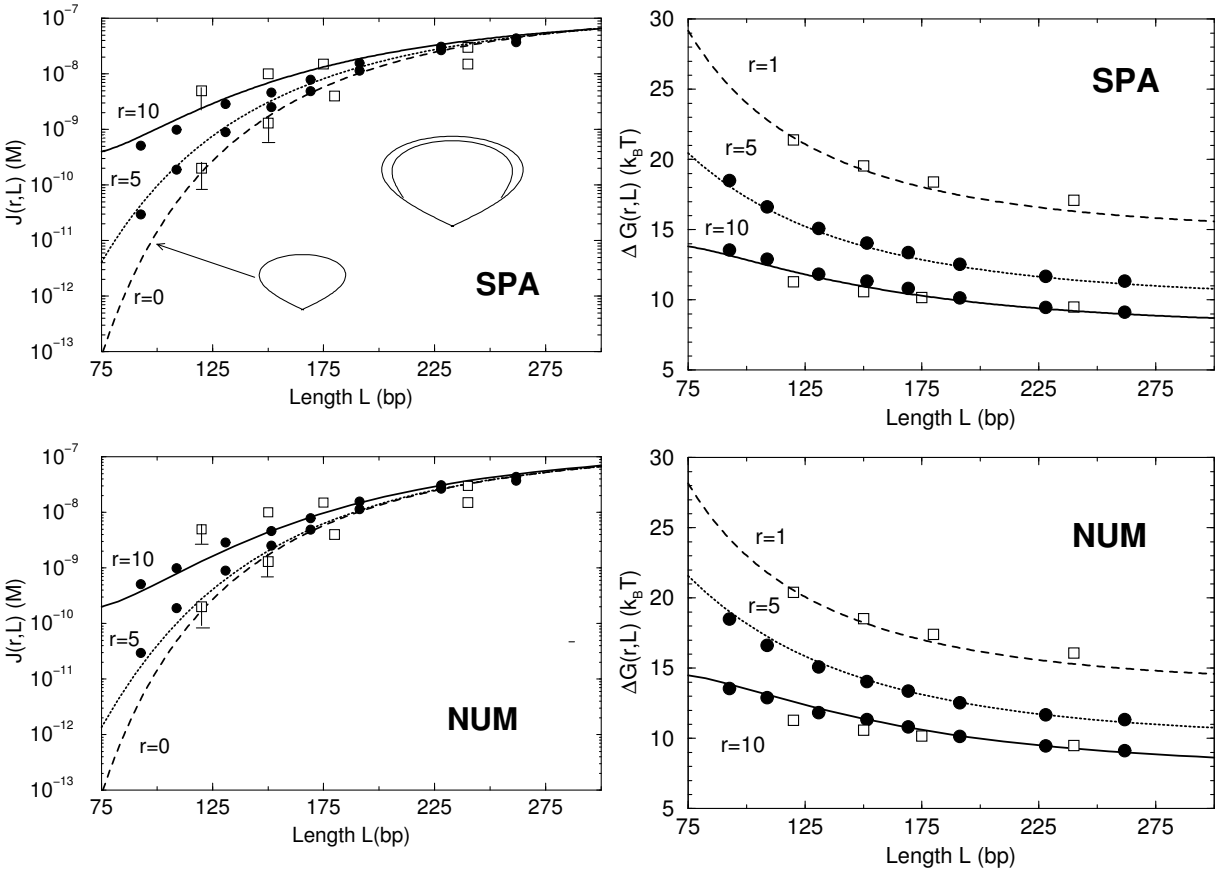

Fig. 6. Looping probability density (left panel) and free energy (right panel) within an interaction radius $r = 1\,\text{nm}$ (dashed line), 5 nm (dotted line), 10 nm (full line). Top: extension of the Shimada-Yamakawa calculation to finite r. Bottom: numerical calculation of the constrained path integral. The error bars are shown when they are larger than the symbol sizes. The $r = 1\,\text{nm}$ curve coincides with the $r = 0$ cyclization factor. Theoretical results are in good agreement with Monte Carlo simulations obtained by Vologodskii and Podtelezhnikov (full circle) [63] and Brownian dynamics simulation obtained by Langowski and collaborators (open square) [71]. Inset of the top left panel: lowest-bending-energy configurations for 100 bp and the $r = 0\,\text{nm}$ (bottom), and $r = 10\,\text{nm}$ (top), the closure of the $r = 10\,\text{nm}$ configuration being shown by a thin line.

the equivalent loop of length $(L + 2r)$. These approach is similar to what was suggested in 1991 by Brenowitz *et al.* to interpret their experimental data, *i.e.* to directly consider the protein as part of the length of the loop, directly in the Shimada-Yamakawa formula. The proposed formula for the cyclisation factor as a function of the protein size r, the length L of the DNA, and the kink angle κ reads

$$j_{\text{M}}(L, r, \kappa) = C_{\text{SY}}(L + 2r)\, e^{-\Delta E(L+2r,\kappa)}. \qquad (8)$$

A linear fit of the bending energy for the optimal configuration in the presence of a kink is

$$\Delta E(L, \kappa) = \left(-7.1 + 20.8\,\frac{\kappa}{\pi}\right)\frac{A}{L}\,. \qquad (9)$$

Formula (8) 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 Figure 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). Moreover the results are in good agreement with numerical calculations (graphics in the bottom). Rippe has also written down a simple formula that includes a kink angle κ and a finite end-to-end distance r in an effective way, but his formula contains a parameter that is fitted, for each values of r and κ, from the curve obtained by Brownian dynamic simulations [73,74]. The Brownian dynamic data points, obtained from a model which includes twist rigidity and elecrtostatic effects, and the Rippe interpolation formula are shown in the inset of Figure 6.

It is important to notice that protein-mediated site-to-site interaction differs from the end-to-end interactions because the DNA fragment is longer than the fragment involved in the loop. The extra DNA should reduce the contact probability between the two sites because it excludes certain conformations. The factor j_{M} calculated above through the end-to-end distribution is therefore an upper limit to the concentration of site-to-site interactions. These entropic effects have been included in theoretical

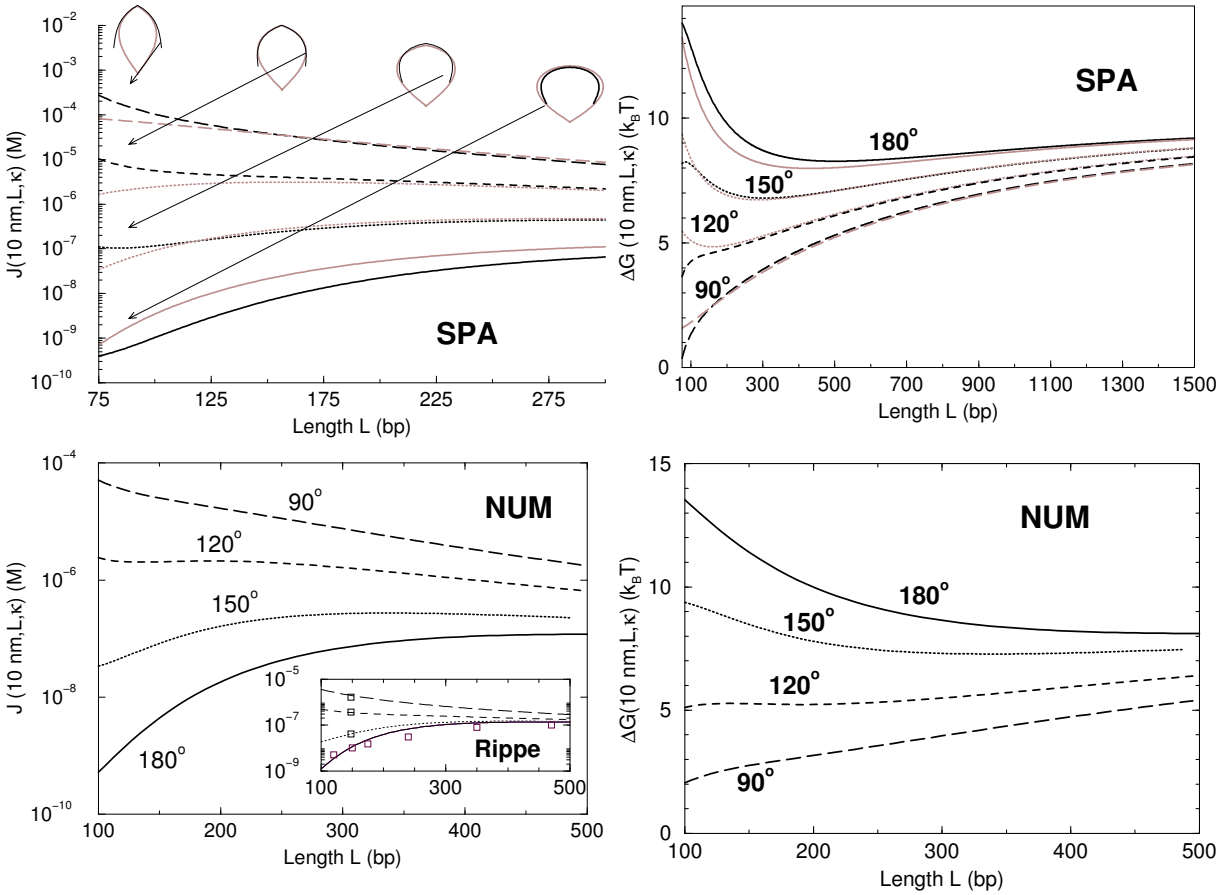

Fig. 7. Probability density (left) and free energy for a loop with $r = 10\,\text{nm}$ and a kink κ in the midle of the chain. $\kappa = 180°$ (full lines), $\kappa = 150°$ (dotted lines), $\kappa = 120°$ (dashed line), $\kappa = 90°$ (long-dashed lines). Top: extension of Shimada-Yamakawa calculation (black lines), approximate formula (gray line) given in the text. Bottom: numerical calculation of the constrained path integral. Inset on the bottom left panel: Brownian dynamics simulation points obtained by Langowski and collaborators (open square) [71], fitted by Rippe formula [73].

approaches for long loops [80]. Moreover this extra DNA has been, for example, taken into account in the calculation of the formation of loop under an applied tension [68,70,81] described in the next section. Another effect that has been underlined in theoretical works is that the DNA stiffness depends on the sequence [82] and can be modified by proteins attached to DNA [83].

6 Concluding remarks: recent and further works

6.1 Loop probability and dynamics under a stretching force

The single-molecule experiments, started with Lia *et al.* work in 1995 [33,34], of loop formation under a stretching force have recently stimulated several theoretical works, in particular by the Marko group [68,81,84]. In their analysis the loops give rise to a length loss in the stretching work done by the force. The loop geometry has to be taken into account; for example in the antiparallel configuration of Figure 1 the extra DNA cannot be fully stretched. The extension of Shimada-Yamakawa calculation of the quadratic fluctuations around the saddle point configuration has been performed in presence of a large external force by Kulić *et al.* [70]. The dynamics of loop formation, which can been directly studied by single strand experiments has been investigated by Monte Carlo simulations [63, 71]. From a theoretical point of view a Langevin dynamics on the end-to-end distance has been studied in [85], and calculation of the mean first passage of cyclisation has been performed in [86]. Clearly the application of polymer dynamics to the problem of loops formation deserves more investigations.

6.2 Towards in vivo modeling

Of course the *in vivo* problem is more complicated because DNA is not free but attached to chromatin structural proteins and the uncontrolled presence of proteins may affect DNA elasticity. The experiments of Ringrose [29] show that the most probable loop length is smaller in chromatin than in bare DNA. Considering its importance this problem clearly deserves some work even

at a simplified level. A very recent work of Sankararaman and Marko has studied the interaction between multiple loops on the same DNA. A simple effect for non–site-specific loops is that the most probable loop length can decrease due to the rearrangement entropy between multiple loops on the same DNA [87].

References

1. S.E. Halford, A.J. Welsh, M.D. Szczelkun, Annu. Rev. Biophys. Biomol. Struct. **33**, 1 (2004).
2. R. Schleif, Annu. Rev. Biochem. **61**, 199 (1992).
3. S. Adhya, Annu. Rev. Genet. **23**, 227 (1989).
4. T.M. Dunn, S. Hahn, S. Ogden, R.F. Schleif, Proc. Natl. Acad. Sci. U.S.A. **81**, 5017 (1984).
5. K.S. Matthews, Microbiol. Rev. **56**, 123 (1992).
6. M. Amouyal, L. Mortensen, H. Buc, K. Hammer, Cell **58**, 545 (1989).
7. M. Ptashne, Cold Spring Harbor Laboratory Press, 3rd edition (2004).
8. T. Tahirov, K. Sato, E. Ichikawa-Iwata, M. Sasaki, T. Inoue-Bungo, M. Shiina, K. Kimura, S. Takata, A. Fujikawa, H. Morii *et al.*, Cell **108**, 57 (2002).
9. P. Jackson, I. Mastrangelo, M. Reed, P. Tegtmeyer, G. Yardley, J. Barrett, Oncogene **16**, 283 (1998).
10. W. Su, S. Porter, S. Kustu, H. Echols, Proc. Natl. Acad. Sci. U.S.A. **87**, 5504 (1990).
11. C. Wyman, I. Rombel, A.K. North, C. Bustamante, S. Kustu, Science **275**, 1658 (1997).
12. J.C. Becker, A. Nikroo, T. Brabletz, R.A. Reisfeld, Proc. Natl. Acad. Sci. U.S.A. **92**, 9727 (1995).
13. B. Tolhuis, R.J. Palstra, E. Splinter, F. Grosveld, W. de Laat, Mol. Cell **10**, 1453 (2002).
14. J. Mahillon, M. Chandler, Microbiol. Mol. Biol. Rev. **62**, 725 (1998).
15. Y. Voziyanov, S. Pathania, M. Jayaram, Nucleic Acids Res. **27**, 930 (1999).
16. A. Nagy, Genesis **26**, 99 (2000).
17. D.L. Hamilton, K. Abremski, J. Mol. Biol. **178**, 481 (1984).
18. M.L. Embleton, V. Siksnys, S.E. Halford, J. Mol. Biol. **311**, 503 (2001).
19. M.L. Embleton, S.A. Williams, M.A. Watson, S.E. Halford, J. Mol. Biol. **289**, 785 (1999).
20. N.E. Murray, Microbiol. Mol. Biol. Rev. **64**, 412 (2000).
21. D. Dryden, N. Murray, D. Rao, Nucleic Acids Res. **29**, 3728 (2000).
22. L. Postow, C.D. Hardy, J. Arsuaga, N.R. Cozzarelli, Genes Dev. **18**, 1766 (2004).
23. J.R. Marsden, U.K. Laemmli, Cell **17**, 849 (1979).
24. Y.L. Lyubchenko, L.S. Shlyakhtenko, B. Chernov, R.E. Harrington, Proc. Natl. Acad. Sci. U.S.A. **88**, 5331 (1991).
25. Z. Katiliene, E. Katilius, N.W. Woodbury, Biophys. J. **84**, 4053 (2003).
26. L.J. Parkhurst, K.M. Parkhurst, R. Powell, J. Wu, S. Williams, Biopolymers **61**, 180 (2003).
27. S. Chatterjee, Y.N. Zhou, S. Roy, S. Adhya, Proc. Natl. Acad. Sci. U.S.A. **94**, 2957 (1997).
28. J.D. Kahn, D.M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **89**, 6343 (1992).
29. L. Ringrose, S. Chabanis, P. Angrand, C. Woodroofe, A.F. Stewart, EMBO J. **18**, 6630 (1999).
30. Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **102**, 5397 (2005).
31. A. Vologodskii, N.R. Cozzarelli, Biophys. J. **70**, 2548 (1996).
32. K. Rippe, M. Guthold, P.H. von Hippe, C. Bustamante, J. Mol. Biol. **270**, 125 (1997).
33. L. Finzi, J. Gelles, Science **267**, 378 (1995).
34. G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. Lewis, S. Adhya, L. Finzi, Proc. Natl. Acad. Sci. U.S.A. **100**, 11373 (2003).
35. D. Skoko, J. Yan, R.C. Johnson, J.F. Marko, Phys. Rev. Lett. **95**, 208101 (2005).
36. D. Shore, J. Langowski, R.L. Baldwin, Proc. Natl. Acad. Sci. U.S.A. **78**, 4833 (1981).
37. O. Kratky, G. Porod, Rel. Trav. Chim. **68**, 1106 (1949).
38. J. J. Hermans, R. Ullman, Physica **18**, 951 (1952).
39. H.E. Daniels, Proc. R. Soc. Edinburgh, Sect. A **63**, 290 (1952).
40. W. Gobush, H. Yamakawa, W.H. Stockmayer, W.S. Magee, J. Chem. Phys. **57**, 2839 (1972).
41. H. Yamakawa, W.H. Stockmayer, J. Chem. Phys. **57**, 2843 (1972).
42. J. Shimada, H. Yamakawa, Macromolecules **17**, 689 (1984).
43. H. Yamakawa, *Helical Wormlike Chains in Polymer Solutions* (Springer Verlag, Berlin, 1997).
44. R.A. Harris, J.E. Hearst, J. Chem. Phys. **44**, 2595 (1966). J.E. Hearst, R.A. Harris, E. Beals, J. Chem. Phys. **45**, 3106 (1966).
45. J.B. Lagowski, J. Noolandi, B. Nickel, J. Chem. Phys. **95**, 1266 (1991).
46. B.-Y. Ha, D. Thirumalai, J. Chem. Phys. **103**, 9408 (1995). J.K. Battacharjee, D. Thirumalai, J.D. Bryngelson, cond-mat/9709345 (1997).
47. R.G. Winkler, P. Reineker, L. Harnau, J. Chem. Phys. **101**, 8119 (1994). R.G. Winkler, J. Chem. Phys. **118**, 2919 (2003).
48. J. Wilhelm, E. Frey, Phys. Rev. Lett. **77**, 2581 (1996).
49. S. Stepanow, G.M. Schütz, Europhys. Lett. **60**, 546 (2002). S. Stepanow, J. Phys. Condens. Matter **17**, S1799 (2005).
50. A.J. Spakowtiz, Z.-G. Wang, J. Chem. Phys. **119**, 13113 (2003); Macromolecules **37**, 5814 (2004).
51. H. Kleinert, A. Chervayakov, cond-mat/0503199 (2005).
52. P.J. Hagerman, V.A. Ramadevi, J. Mol. Biol. **212**, 351 (1990).
53. D.M. Crothers, J. Drak, J.D. Kahn, S.D. Levene, Methods Enzymol. **212**, 3 (1992).
54. S.D. Levene, D.M. Crothers, J. Mol. Biol. **189**, 61 (1986), J. Mol. Biol. **189**, 73 (1986).
55. M.D. Frank-Kamenetskii, A.V. Lukashin, V.V. Anshelevich, A.V. Vologodskii, J. Bimol. Struct. Dyn. **2**, 1005 (1985).
56. T.E. Cloutier, J. Widom, Mol. Cell **14**, 355 (2004).
57. J. Yan, J.F. Marko, Phys. Rev. E **68**, 011905 (2003); Phys. Rev. Lett. **93**, 108108 (2004); J. Yan, R. Kawamura, J.F. Marko, Phys. Rev. E **71**, 061905 (2005).
58. P.A. Wiggins, R. Phillips, P.C. Nelson, Phys. Rev. E **71**, 021909 (2005).
59. B. Chakrabarti, A.J. Levine, Phys. Rev. E **71**, 031905 (2005).
60. Y.O. Popov, A.V. Tkachenko, Phys. Rev. E **71**, 051905 (2005).

61. P. Ranjith, P.B. Sunil Kumar, G.I. Menon, Phys. Rev. Lett. **94**, 138102 (2005).
62. Q. Du, M. Vologodskaia, H. Kuhn, M. Frank-Kamenetskii, A. Vologodskii, Biophys. J. **88**, 4137 (2005).
63. A.A. Podtelezhnikov, A.V. Vologodskii, Macromolecules **33**, 2767 (2000).
64. C. Bustamante, J.F. Marko, E.D. Siggia, S. Smith, Science **265**, 1599 (1994).
65. J.F. Marko, E.D. Siggia, Macromolecules **28**, 8759 (1995).
66. C. Bouchiat, M.D. Wang, S.M. Block, J.-F. Allemand, T. Strick, V. Croquette, Biophys. J. **76**, 409 (1999).
67. M. Wang, H. Yin, R. Landick, J. Gelles, S.M. Block, Biophys. J. **72**, 1335 (1997).
68. S. Sankararaman, J. Marko, Phys. Rev. E **71**, 021911 (2005)
69. N. Douarche, S. Cocco, Phys. Rev. E. **72**, 061902 (2005).
70. J.M. Kulić, H. Mohrbach, V. Lobaskin, R. Thaokar, H. Schiessel, Phys. Rev. E. **72**, 041905 (2005).
71. H. Merlitz, K. Rippe, K. Klenin, J. Langowski, Biophys. J. **74**, 773 (1998).
72. K. Rippe, P.H. von Hippel, J. Langowski, Trends Biochem. Sci. **20**, 500 (1995).
73. K. Rippe, Trends Biochem. Sci. **26**, 733 (2001).
74. V.A. Bloomfield, D.M. Crothers, J.I. Tinoco, *Physical Chemistry of Nucleic Acids* (Harper and Row, 1974).
75. M. Brenowitz, A. Pickar, E. Jamison, Biochemistry **30**, 5986 (1991).
76. K. Klenin, H. Merlitz, J. Langowski, Biophys. J. **74**, 780 (1998).
77. M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, P. Lu, Science **271**, 1247 (1996).
78. A. Balaeff, L. Mahadevan, K. Schulten, Phys. Rev. Lett. **83**, 4900 (1999); Structure **12**, 123 (2004); A. Balaeff, C.R. Koudella, L. Mahadevan, K. Schulten, Philos. Trans. R. Soc. London, Ser. A **362**, 1355 (2004).
79. E. Villa, A. Balaeff, K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **102**, 6783 (2005).
80. A. Hanke, R. Metzler, Biophys. J. **85**, 167 (2003).
81. S. Blumberg, A.V. Tkachenko, J.C. Meiners, Biophys. J. **88**, 1692 (2005).
82. D. Garrivier, B. Fourcade, Europhys. Lett. **49**, 390 (2000).
83. J. Rudnick, R. Bruinsma, Biophys. J. **76**, 1725 (1999).
84. J.F. Marko, E. Siggia, Biophys. J. **73**, 2173 (1997).
85. S. Jun, J. Bechhoefer, B.-Y. Ha, Europhys. Lett. **64**, 420 (2003).
86. A. Dua, B. Cherayl, J. Chem. Phys. **117**, 7765 (2002); **116**, 399 (2002).
87. S. Sankararaman, J.F. Marko, Phys. Rev. Lett. **95**, 078104 (2005).