A generalized theory of DNA looping and cyclization

D. P. Wilson1(a), A. V. Tkachenko2 and J.-C. Meiners1,3

1 Department of Physics, University of Michigan - 450 Church St., Ann Arbor, MI 48109, USA
2 Center for Functional Nanomaterials, Brookhaven National Laboratory - Bldg 735, Upton, NY 11973, USA
3 LSA Biophysics, University of Michigan - 930 N University Ave, Ann Arbor, MI 48109, USA

received 3 November 2009; accepted in final form 2 March 2010
published online 30 March 2010

PACS 87.14.gk – DNA
PACS 87.15.La – Mechanical properties
PACS 87.15.ad – Analytical theories

Abstract – We have developed a generalized semi-analytic approach for efficiently computing cyclization and looping $J$ factors of DNA under arbitrary binding constraints. Many biological systems involving DNA-protein interactions impose precise boundary conditions on DNA, which necessitates a treatment beyond the Shimada-Yamakawa model for ring cyclization. Our model allows for DNA to be treated as a heteropolymer with sequence-dependent intrinsic curvature and stiffness, yet faithfully reproduces the results of Shimada and Yamakawa for the ring and unconstrained loop. In this framework, we independently compute enthalpic and entropic contributions to the $J$ factor and show that even at small length scales ($\sim \ell_p$) entropic effects are significant. We propose a simple analytic formula to describe our numerical results for near planar loops of homogenous DNA, which can be used to predict experimental cyclization and loop formation probabilities as a function of loop size and binding geometry. We also introduce an effective torsional persistence length that describes the coupling between twist and bending of DNA when looped.

Copyright © EPLA, 2010

Introduction. – Calculating the probability that contact will occur between two distant ends of a polymer under prescribed orientations is a long-standing question of considerable significance in polymer physics. This problem was rigorously defined in the context of polyelectrolyte condensation as the ratio of equilibrium constants for cyclization and bimolecular association by introduction of the Jacobson Stockmayer ($J$) factor [1]. Yamakawa and Stockmayer expanded on this work using the Kratky-Porod wormlike chain model (WLC) to compute the $J$ factor of angle-independent DNA ring-closure probabilities [2]. Shimada and Yamakawa then included twist alignment of the end points [3], known as phasing, to explain the measured oscillatory cyclization rates by Shore and Baldwin on DNA shorter than 500 base pairs [4]. Shimada and Yamakawa calculated the $J$ factor for the ring and unconstrained loop, by treating DNA as a homo-polymer with coincident end points and parallel tangent vectors, as well as with coincident end points with unconstrained tangent vectors, respectively, see fig. 1. Our work generalizes this closure probability to include arbitrary end point locations, binding orientations, sequence-dependent curvature and elasticity while reproducing the earlier results of Shimada and Yamakawa for the ring and unconstrained loop [3].

We numerically calculate $J$ factors based on a semi-analytic continuous elastic-rod formulation that includes as inputs specified end point locations and orientations of the DNA. This formulation goes beyond the homogeneous straight elastic rod of Balaeff et al. [5–7] by allowing the inclusion of intrinsic curvature and stiffness based upon sequence-dependent effects. We also compute thermal fluctuations, which contribute non-trivially to the free-energy cost of loop formation. This aspect has some similarities to the recent work of Zhang and Crothers [8,9] who used a discrete model to compute thermal fluctuations, although their $J$ factors disagreed with the previous results of Shimada and Yamakawa [3].

Previous works by Olson et al. have modeled DNA as an elastic rod with sequence specific properties using individual base pairs as their elements to examine normal modes [10–17]. While Monte Carlo methods have been successfully used to compute $J$ factors [18,19], they are in general computationally taxing, making it potentially

(a)E-mail: dawilson13@gmail.com
difficult to separate out the individual effects of curvature and stiffness, or make the distinction between enthalpic and entropic contributions; by contrast, our computation of the $J$ factor based on a desired equilibrium shape takes only minutes on a desktop computer.

Our model independently calculates enthalpic and entropic contributions to the free energy of the DNA loop. The numerical results show that boundary-condition-dominated entropic contributions are important even for very short DNA on the order a persistence length ($\ell_p$). Within a cell, DNA is normally constrained by histones and other binding constraints, leaving this length scale as the typical size of locally fluctuating DNA.

Many DNA-binding proteins impose very specific boundary conditions on DNA loop formation. Previous results by Swigon et al., Segall et al. and Purohit et al. have shown that boundary condition constraints on the DNA end points play a significant role in the facilitation of loop formation [17,20,21]. Boundary conditions have also been suggested by Tkachenko [22] as an explanation for the striking disagreements between the cyclization rates measured by Du et al. [23,24] and Cloutier et al. [25,26]. Therefore, any useful model for these interactions must accommodate such arbitrary boundary conditions. Thus, the $J$ factor framework gives quantitative insights into the mechanics of protein-mediated DNA loop formation and is important for multi-scale models of larger DNA-protein assemblies such as chromatin and nucleosomes.

**Theory.** To model the enthalpic and entropic contributions to the looping $J$ factor, we use a coarse-grained elastic-rod model for the DNA polymer to calculate the Hamiltonians describing thermal fluctuations about the open ($H^o$) and looped ($H^l$) states. The $J$ factor is then calculated by comparing the probability densities of finding the DNA in a configuration that corresponds to the looped state with enforced boundary conditions for the end points, described by three angles ($\Theta$, $\Phi$, $\Psi$) and three positions ($x$, $y$, $z$) (see fig. 1), to that of the open state without such constraints:

$$J = \frac{8\pi^2 \int [d\xi] e^{-\beta H^o(\xi)} \delta(\bar{u}(L)) \delta(\theta_1(L)) \delta(\theta_2(L)) \delta(\psi(L))}{\int [d\xi] e^{-\beta H^l(\xi)}}$$

(1)

The integration is over the amplitudes $d\xi_i$ of the normal modes with eigenvalue $\lambda_i$ of the respective Hamiltonians, and $\beta = 1/(k_B T)$, the inverse product of the Boltzmann constant $k_B$, and temperature $T$. The DNA is parameterized by the arc length parameter $s$, where $s = 0$ and $s = L$ are taken to be the end points. The end point tangent vectors have three angular constraints $\theta_1(L)$, $\theta_2(L)$, $\psi(L)$ and a relative displacement vector $\bar{u}$ which are imposed by $\delta^0(\bar{u})$, $\delta(\theta_{1,2})$, and $\delta(\psi)$, respectively. The $\delta$-functions assume rigid constraints, although when boundary conditions are less rigid, such as in cases of head group or C-terminus protein flexibility, we are able to capture these effects by replacing the $\delta$-functions with appropriate boundary potentials as in Swigon et al. [17] and in Zhang et al. [9].

The open state is characterized by three input local curvature components $\bar{\kappa}^o = (\kappa_1^o(s), \kappa_2^o(s), \tau^o(s))$, which represent intrinsic curvature caused by sequence-dependence. We include as inputs, two bending persistence lengths $\ell_1(s), \ell_2(s)$, corresponding to bending elasticity along the major and minor grooves of DNA, respectively, as well as a torsional persistence length $\ell_3(s)$.

The three equilibrium looped state curvature components $\bar{\kappa}^l = (\kappa_1^l(s), \kappa_2^l(s), \tau^l(s))$ are found by minimizing the strain energy of DNA under specified orientations and positions of the end point tangent vectors, while tracking the DNA cross-sections, as demonstrated by Goyal et al. [27]. This tracking allows the use of the body fixed vectors $(\ell(s), \hat{n}_1(s), \hat{n}_2(s))$ for a basis to define angular deformations $(\theta_1(s), \theta_2(s), \psi(s))$ from the equilibrium open and looped states. The angles $\theta_{1,2}(s)$ are defined as rotations about the two open-state normal vectors and $\psi(s)$ is defined as a rotation about the tangent vector of the open state. The computation of this equilibrium loop does not contain electrostatics as did Balaeff et al. [6], although it does explicitly avoid self-contact [28]. While these interactions may affect equilibrium conformations in which DNA segments come within $1–2$ nm, generally electrostatic screening under typical physiological conditions makes these effects quite small.

The deformation-induced curvatures $\bar{\kappa}^{o,l} = (\kappa_1^{o,l}(s), \kappa_2^{o,l}(s), \tau^{o,l}(s))$ are computed separately for the open ($\kappa^o$) and looped ($\kappa^l$) DNA states.
and looped ($\tilde{\kappa}^l$) states, respectively. The looped-state Hamiltonian is expressed as $H^l = E^l + \delta H(\theta_1, \theta_2, \psi)$, where $E^l$ is the strain or enthalpic energetic cost of loop formation. The open-state equilibrium is the intrinsic curvature induced by sequence-dependence. As DNA is in an aqueous solution, its mobility is severely overdamped and the kinetic-energy contributions to the Hamiltonian are neglected. The deformation Hamiltonian is then

$$\beta H^{o,l} = \frac{1}{2} \int_0^L ds \left( \tilde{\kappa}^{o,l} - \tilde{\kappa}^o \right)^T B(s) \left( \tilde{\kappa}^{o,l} - \tilde{\kappa}^o \right). \quad (2)$$

In general we are capable of treating $B(s)$ as a symmetric matrix with 6 independent components as was suggested by Olson et al. [12] and Towles et al. [18]. The components of $B(s)$ may vary with sequence along the DNA as suggested by Colemen, Swigon and Olson [15,16]. Such non-homogeneous parameters are necessary to elucidate the results of Cloutier and Widom [26], who argued sequence-dependent elasticity in nucleosome binding sequences contributed to greater ring cyclization probability. For the purposes of this paper we take $B(s)$ to be diagonal with components $\ell_1(s)$, $\ell_2(s)$ and $\ell_3(s)$ to demonstrate that our work reproduces the previous results of Shimada et al. [3], as well as to illustrate the differences of our results to those of Zhang et al. [8].

The curvature components contained in the Hamiltonian are made non-dimensional by scaling with the overall DNA length, $L$.

When constructing the Hamiltonian, the Galerkin method is used to numerically solve for the normal modes of the open and looped states. Each deformation variable $(\theta_1(s), \theta_2(s), \psi(s))$ is expanded in terms of $N$ orthogonal comparison functions, which are then used to create a $3N \times 3N$ Hamiltonian matrix for the open $H^o$ and the looped $H^l$ states. The comparison functions satisfy the angular boundary constraints imposed by $\delta(\theta_1(L)), \delta(\theta_2(L)), \delta(\psi(L))$ in eq. (1). The remaining looped boundary condition $\delta^3(\bar{u})$ is satisfied by Fourier expanding the delta functions, and then integrating over the eigenvector amplitudes $\xi_i$, leading to the constraint matrix $V$. An additional integration for the open state is required to cover the modes which cause displacements of the end points. The $J$ factor can then be expressed as

$$J = \frac{1}{\ell_p^2} \sqrt{\frac{\det H^o}{2\pi^3 \det H^l}} \left( \frac{\ell_p}{L} \right)^{11/2} e^{-\frac{1}{2} \frac{\ell_p}{L} \int s^2 ds - \frac{L}{4 \ell_p}} \right), \quad (5)$$

which is a product of two functions, one describing the entropic contributions, henceforth referred to as the entropic coefficient $\Lambda(\Theta)$, and the exponential term containing the enthalpic contributions. While the product of eigenvalues is formally infinite, the ratio of determinants of the Hamiltonians for the open and looped state are finite. In the limit of large $N$, the differences between the open and looped eigenmodes vanish, as the spatial resolution of the eigenmodes is below the curvature of the states. This fact allows us to truncate the ratio of modes after the $M$-th eigenmode, which in practice, $M \sim 20$, which for DNA of the order of $\ell_p$ corresponds to spatial node distances of 7 base pairs, which can safely be regarded as continuous. The ratio of eigenvalues is

$$\frac{\det H^o}{\det H^l} = \frac{\lambda^o_1 \cdots \lambda^o_M}{\lambda^l_1 \cdots \lambda^l_M}. \quad (6)$$

The ratio of eigenvalues describes how the space accessible to thermal fluctuations of the DNA is reduced upon loop formation, which in turn allows us to quantify the entropic change of the system.

Results. – The results presented here are for near planar DNA loops with coincident end points and arbitrary loop tangent angle $\Theta$. In this paper we present results for DNA of the length 50 nm or approximately 14 helical repeats, so we will assume $\ell_1 = \ell_2$. We treat the DNA as a homogeneous polymer with bending and torsional persistence lengths of 50 nm and 75 nm, respectively [29–31]. While DNA is a heteropolymer with anisotropic bending persistence lengths $\ell_1(s) \neq \ell_2(s)$, this anisotropy largely averages out after a few helical repeats (10.5 base pairs) as demonstrated by Kehrbaum and Maddocks [32].

The entropic coefficient $\Lambda(\Theta)$ is computed for a torsionally unconstrained DNA loop with overall length $L = \ell_p$ and loop tangent angles ranging from a Ring $\Theta = 0$, to a Teardrop $\Theta \sim 0.54\pi$, to a Hairpin $\Theta = \pi$ and is given in fig. 2. The lowest eigenvalue of the in-plane loops can be well approximated as $\lambda^o_1 = 2\pi \sigma_\Theta$. Factoring this contribution from $\Lambda(\Theta)$ reveals a slowly varying function $\gamma(\Theta)$. We are able to fit our numerical results to within 1% by using a modified Bessel Function

$$J(\Theta) = \left[ I_0(2\pi \Theta) e^{-2\pi \Theta} \right] \gamma(\Theta) \frac{1}{\ell_p^2} \left( \frac{\ell_p}{L} \right)^{11/2} \left[ \frac{\ell_p}{L} E(\Theta) - \frac{L}{4 \ell_p} \right], \quad (7)$$

$$\gamma(\Theta) = 365\Theta^2 - 525\Theta + 32\pi^3, \quad (8)$$
Fig. 2: (Colour on-line) The entropic coefficient $\Lambda(\Theta)$ is largely dominated by contributions from the lowest eigenmode of the loop, $\lambda_1(\Theta)$. To illustrate this dependence, we write $\Lambda(\Theta) = f(\lambda_1(\Theta)) \gamma(\Theta)$, where $f(\lambda_1)$ contains only the contributions of the lowest eigenmode $\lambda_1$. The function $f(\lambda_1)$ is represented by the bracketed quantity in eq. (7), and $\gamma(\Theta)$ is given in eq. (8). It is then clear that $\gamma(\Theta)$ is a slowly varying function on the interval $\Theta = 0$ to $\Theta = 0.54\pi$, and then steadily increases on the interval from $\Theta = 0.54\pi$ to Hairpin ($\Theta = \pi$). The shift in behavior of $\gamma(\Theta)$ occurs after the lowest eigenmode changes from symmetric to antisymmetric. Even for relatively short DNA $\Lambda(\Theta)$ is shown to effect the $J$ factor by an order of magnitude in fig. 3. Note the dimensions are in molarity rather than concentration as in eq. (5).

\[
E(\Theta) = \frac{1}{2} \int \kappa_2^2 ds = 2.02(\Theta - 0.54\pi)^2 + 14.05, \tag{9}
\]
where $\Theta$ is the loop tangent angle in radians. The fit is accurate for all angles $\Theta$ although the dimensional scaling of eq. (7) needs to be modified to $\left( \ell_p/L \right)^6$ when $\Theta = 0$, as the ring has a zero mode [3]. The unconstrained loop by contrast has dimensional scaling of $\left( \ell_p/L \right)^5$, due to integrating over the orientations of the tangent vectors.

The Teardrop shape has the lowest strain (enthalpic) energy, $14.44\ell_p^2 k_B T$, of any of the in-plane shapes, and is where the end point curvatures of the loop vanish. Enthalpic considerations demonstrate which loop tangent angle $\Theta$ will produce the maximum $J$ factor, although for small angles, as well Hairpin structures, entropic considerations are required to demonstrate the absolute behavior of the $J$ factor, as seen in fig. 3.

Shimada and Yamakawa provided two special cases for the in-plane $J$ factors [3]: the Ring defined by aligned tangents, $\Theta = 0$, and the unconstrained loop. Zhang and Crothers [8] arrived at different results for these $J$ factors using their base pair harmonic mode computations, although the reason for the difference remains unclear.

We have reproduced the ring and unconstrained loop results to within 0.01% using our continuum rod model in agreement with Shimada and Yamakawa. The two models should not yield significantly different results since the radius of curvature is much greater than the DNA bond length. We also compute nearly the same ring and loop energy as Zhang and Crothers. In addition, the highest spatial frequency modes, which would probe the shortest length scales do not contribute as seen from eq. (6). Finally, our theory treats the delta functions in eq. (1) differently than Zhang and Crothers, which may be the cause of the discrepancy in their results.

The ring offers a unique challenge when computing the normal modes as it contains a zero-eigenvalue mode which causes the entire ring to rotate about the two DNA end points under our $\delta(\vec{u}, \theta)$-function conditions. This rotation can be related to the amplitude $\xi_0$ and the subsequent integration yields a finite contribution when integrated over this angle instead of $\xi_0$. The unconstrained loop formation assumes that $E(\Theta)$ given above in eq. (9) is symmetric about the Teardrop. As most biologically relevant cases do not fit neatly into one of these special cases, our generalized results allow a more accurate prediction of the $J$ factor. To illustrate the effect of angular dependence on the $J$ factor, we plot three $J$ factors with
A generalized theory of DNA looping and cyclization

Fig. 4: (Colour on-line) (a) The effective torsional persistence length, $\ell^*_\tau$ in units of $\ell_p$ as a function of loop formation angle, $\Theta$. The Ring has pure torsional modes with stiffness $\lambda_i\ell_\tau$ and as $\Theta$ increases the bending and torsional modes become coupled, reducing the effective torsional persistence length. (b) The torsion-bending coupling $\alpha(\Theta)$ shown as open circles, is quadratic from the Ring to the Teardrop as seen by the dashed line. From the Teardrop to the Hairpin, $\alpha(\Theta)$ is cubic in $\Theta$.

different entropic coefficients $\Lambda$ in fig. 3. Two of these $J$ factors have constant entropic coefficients $\Lambda \neq \Lambda(\Theta)$, that of the Ring and unconstrained loop, while allowing the normal angular dependence of the enthalpic contributions $E(\Theta)$. Thus fig. 3 demonstrates that the changes to the entropic contributions as a function of $\Theta$ are critical in the $J$ factor calculation. The extrapolation from the orientation averaged loop is better than the ring, except for small angles $\Theta$. For all angles, the Bessel function given in eq. (7) is an excellent fit, with a maximum error of less than 1% for all $\Theta$.

The general DNA-protein complex has spatially separated end points as well as prescribed angles ($\Theta$, $\Phi$, $\Psi$) which can be obtained from DNA-protein co-crystals with LacI protein serving as the canonical example [33]. The extension of this work to include proteins would necessitate a more complete description of the DNA-protein binding regions which would involve modifying our $\delta(\bar{u})$ functions above in eq. (1) as was done in [9,17,21].

Computing the torsionally constrained $J$ factor by including $\delta(\psi)$ in eq. (1) allows a determination of an effective torsional persistence length $\ell^*_\tau$. The coupling of torsion and bending elasticity can be computed as

$$\ell^*_\tau = 2\pi \left( \frac{J(\Theta)}{J(\Theta)} \right)^2,$$

where $J(\Theta)_{\psi=0}$ and $J(\Theta)_{\psi \neq 0}$ are the torsionally constrained and unconstrained $J$ factors, respectively. The effective torsional persistence length represents the conversion between twist and writhe.

The effective torsional persistence length can be written as a torsional and bending spring in series

$$\frac{1}{\ell^*_\tau} = \frac{1}{\ell_\tau} + \alpha(\Theta) \frac{1}{\ell_p},$$

where all of the angular dependence is given by $\alpha$. In fig. 4, it is clear that $\alpha(\Theta)$ has as simple quadratic dependence up until the Teardrop shape and afterwards becomes cubic in $\Theta$

$$\alpha(\Theta) = \frac{1}{2\pi^2} \Theta^2 - \frac{1}{6\pi^3} \Theta, \quad 0 \leq \Theta \leq 0.55\pi,$$

$$\alpha(\Theta) = 0.42\Theta^3 - 2.55\Theta^2 + 5.46\Theta - 3.87, \quad 0.55\pi \leq \Theta \leq \pi.$$  

Conclusion. – We have developed a generalized approach for computing $J$ factors of arbitrary loop shapes, which may include sequence-dependent stiffness and curvature and reproduces the results of Shimada and Yamakawa for the ring and unconstrained loop. We have shown that the $J$ factor varies strongly for near planar loop shapes as a function of loop tangent angle $\Theta$ for intrinsically straight DNA with isotropic bending stiffness. The in-plane $J$ factors can be well fit with analytic functions for all $\Theta$. We have defined an effective torsional persistence length $\ell^*_\tau$ and subsequence torsion-bending coupling $\alpha(\Theta)$ which are shown to vary significantly as a function of loop formation angle, $\Theta$. Finally, our calculation is computationally very quick, taking only a few minutes per $J$ factor for any set of input boundary conditions.

***

The authors would like to thank N. Perkins, S. Goyal, T. Lillian, G. Blab, K. Raghunathan, P. Koehn, Y.-F. Chen and M. Wilson for their many helpful conversations. This work was partially funded through grant GM 65934 from the National Institutes of Health.

REFERENCES

[1] Jacobson H. and Stockmayer W. H., J. Chem. Phys., 18 (1950) 1600.
[2] Yamakawa H. and Stockmayer W. H., J. Chem. Phys., 57 (1972) 2843.

58005-p5
[3] Shimada J. and Yamakawa H., *Macromolecules*, **17** (1984) 689.

[4] Shore D., Langowski J. and Baldwin R. L., *Biochemistry*, **76** (1981) 4833.

[5] Balaeff A., Mahadevan L. and Schulten K., *Phys. Rev. Lett.*, **83** (1999) 4000.

[6] Balaeff A., Koudella C. R., Mahadevan L. and Schulten K., *Philos. Trans. Ser. A, Math. Phys. Eng. Sci.*, **362** (2004) 1355.

[7] Balaeff A., Mahadevan L. and Schulten K., *Phys. Rev. E*, **73** (2006) 1.

[8] Zhang Y. and Crothers D. M., *Proc. Natl. Acad. Sci. U.S.A.*, **100** (2003) 3161.

[9] Zhang Y., McEwen A. E., Crothers D. M. and Levene S. D., *Proc. Natl. Acad. Sci. U.S.A.*, **102** (2005) 5397.

[10] Marky N. L., Olson W. K., Jernigan R. L. and Zhurkin V. B., *J. Mol. Biol.*, **232** (1993) 530.

[11] Olson W. K., *Curr. Opin. Struct. Biol.*, **6** (1996) 242.

[12] Zhurkin V. B., Olson W. K., Gorin A. A., Lu X. J. and Hock L. M., *Proc. Natl. Acad. Sci. U.S.A.*, **95** (1998) 11163.

[13] Olson W. K. and Zhurkin V. B., *Curr. Opin. Struct. Biol.*, **10** (2000) 286.

[14] Matsumoto A. and Olson W. K., *Biophys. J.*, **83** (2002) 22.

[15] Coleman B. D., Olson W. K. and Swigon D., *J. Chem. Phys.*, **118** (2003) 7127.

[16] Olson W. K., Swigon D. and Coleman B. D., *Trans. R. Soc. London, Ser. A*, **362** (2004) 1403.

[17] Swigon D., Coleman B. D. and Olson W. K., *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 9879.

[18] Towles K. B., Beausang J. F., Garcia H. G., Phillips R. and Nelson P. C., *Phys. Biol.*, **6** (2009) 2001.

[19] Panyukov S. and Rabin Y., *Phys. Rev. E*, **64** (2001) 011909.