Multiple Equilibria in DNA Rings

Patrick Furrer, Robert Manning, and John Maddocks

Département de Mathématiques
École Polytechnique Fédérale de Lausanne
CH-1015 Lausanne SWITZERLAND

http://lcvwww.epfl.ch/
pfurrer@masg1.epfl.ch, rmanning@masg1.epfl.ch, maddocks@dma.epfl.ch

December 3, 1998

In preparation for Biophysical Journal
1 Introduction

Much effort has been directed recently toward using an elastic rod model to calculate the equilibrium configurations of a DNA molecule under various externally imposed constraints. One type of constraint commonly studied is the constraint of circularity: the requirement that the two ends of the DNA close to form a looped molecule. If both strands of the sugar-phosphate backbone are required to close, circular DNA is commonly referred to as cyclized, while if only one strand is closed, the DNA is referred to as nicked. The prevailing computational strategy for finding equilibria of circular DNA has been to directly minimize the DNA strain energy functional derived from the elastic rod model, e.g. by a Newton-Raphson-type algorithm [16]. An alternate approach, see e.g. [2, 11, 21, 14, 13], is to compute solutions of the first-order equilibrium equations for this strain energy functional, as determined through the standard procedures of the calculus of variations. In this paper, we demonstrate how, via an analysis of these equilibrium equations, one may obtain qualitative understanding of the set of DNA equilibria not readily available from the direct minimization of the strain energy functional.

In particular, we analyze these equilibrium equations in the transition from the symmetric case of an intrinsically straight isotropic rod to the nonsymmetric case incorporating intrinsic curvature. This transition was the focus of the recent Metropolis Monte Carlo study by Katritch and Vologodskii [8], and we compare our theories directly against their findings. In their Monte Carlo simulations, Katritch and Vologodskii discovered that the distribution $P(Lk)$ of linking number in nicked DNA is dramatically affected by the presence of intrinsic DNA curvature. They observed that while for intrinsically straight DNA, this distribution always looks Gaussian, the addition of intrinsic curvature led in certain cases to distributions that could only be fit by a sum of two (or more) Gaussians. They noted that this bimodal behavior occurred especially when the intrinsic shape of the DNA is S-like. They concluded from these findings that the addition of intrinsic curvature induces a second stable nicked DNA equilibrium not found in the intrinsically straight case.

Of course, Metropolis Monte Carlo simulations do not directly compute equilibria, but rather sample fluctuations about one or several existing equilibria. Thus, the presence of multiple equilibria, while having an important effect on Monte Carlo results, can only be detected indirectly by such methods. On the other hand, we will show that a direct analysis of the equilibrium rod equations yields a test for the presence of multiple nicked (or cyclized) equilibria in good agreement with the Monte Carlo findings of Katritch and Vologodskii. Specifically, the existence of multiple nicked minima results from the perturbation of a family of degenerate equilibria for the symmetric problem of the circularization of an intrinsically straight rod. The test for such multiple equilibria involves the determination of simple integrals of intrinsic shape parameters, computable in a matter of seconds from the base-pair sequence, and thus offers a valuable tool in guiding the selection of sequences for the more intensive Monte Carlo computations.

In Sec. 2, we describe the rod equilibrium equations and the perturbation expansion of [13] that is at the heart of our classification of multiple equilibria. In Sec. 3, we describe the equilibrium and Monte Carlo computations used to illustrate our results. In Secs. 4.1–4.2, we verify the efficiency of the perturbation expansion as a classifier of multiple equilibria in a study
of circular DNA equilibria (both nicked and cyclized) for several thousand DNA sequences of 200 and 900 base-pairs. We also investigate how the reliability of this test for multiple equilibria depends on wedge-angle set and DNA length. Finally, in Sec. 4.4, we investigate the connections between the continuum rod equilibria and the results of the Katritch and Vologodskii Metropolis Monte Carlo simulations. Beyond even the count of equilibria, we find remarkable correlations between Monte Carlo peak locations and peak heights, and continuum equilibrium links and energies, respectively. Given the complexity of the strain-energy surface on which the Monte Carlo simulation wanders, it is surprising that a relatively simple equilibrium computation can yield such good predictions of some Monte Carlo results. This thus offers hope that continuum methods can be used in the future as valuable pre-computations to guide Monte Carlo or other similarly complicated computations.

2 Theory

Section 2.1 presents the basic equations of rod mechanics, and Sec. 2.2 describes our procedure for incorporating DNA parameters into this rod model. Section 2.3 describes the static equilibrium configurations of an elastic ring for an intrinsically straight isotropic rod. Then in Sec. 2.4 we present the central result to be applied in this paper: a perturbation computation that determines the number of ring equilibria that result when an infinitesimal amount of intrinsic curvature is added to the rod. Finally, in Sec. 2.5, we discuss the implications of this perturbation computation for intrinsically curved DNA.

2.1 The elastic rod equations

We begin by summarizing the formulation presented in [4, 12] of the special Cosserat theory (see, e.g. [1]), commonly used in continuum mechanics to model an inextensible and unshearable rod. For each value of arclength $s$ along the rod ($0 \leq s \leq 1$), the centerline is denoted by $\mathbf{r}(s)$ and the orientation of the rod cross-section is given by an orthonormal frame of directors $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$. Under the assumption of inextensibility and unshearability, the normal vector $\mathbf{d}_3$ to the cross-section coincides with the unit tangent vector to the centerline $\mathbf{r}'$ (differentiation by $s$ being denoted throughout by a prime). Thus, the rod is completely described by the directors $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$, with the centerline determined from the directors via an integration of $\mathbf{d}_3$.

From a rod configuration $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$, we define strains $u_i(s)$ as follows:

$$ u_1(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \equiv -\mathbf{d}_2^T \mathbf{d}_3' $$

$$ u_2(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \equiv \mathbf{d}_1^T \mathbf{d}_3' $$

$$ u_3(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \equiv \mathbf{d}_2^T \mathbf{d}_1'. $$

In the absence of external forces, we assume the rod has a unique minimal-energy intrinsic shape $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$. The strains corresponding to the intrinsic shape are denoted by $\hat{u}_i(s) \equiv u_i(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)$.
Under the assumption of hyperelasticity, the strains \( u_i \) determine a strain-energy via a density function \( W(u_1, u_2, u_3, s) \):

\[
E = \int_0^1 W(u_1(s), u_2(s), u_3(s), s) ds.
\]

In this paper, we further assume the particular form

\[
W(u_1, u_2, u_3, s) = \frac{1}{2} \sum_{i=1}^{3} K_i (u_i - \bar{u}_i)^2, \quad K_1 = K_2,
\]

although the procedures we describe could similarly be applied to other strain-energy functions. Further, in all computed examples, we take \( K_3 = 0.8K_1 \), although the theoretical results presented hold for arbitrary values of \( K_3 \) and the computations could readily be repeated for any other value.

With the above assumptions and notation, the equilibrium configurations of a rod are critical, or stationary, points of the energy \( \int_0^1 W ds \) subject to appropriate boundary conditions. As shown in [4, 12], these critical points satisfy a Hamiltonian system of first-order ODEs:

\[
\begin{align*}
\mathbf{r}' &= \frac{\partial H}{\partial \mathbf{n}} = d_3(q), \\
\mathbf{q}' &= \frac{\partial H}{\partial \mu} = \frac{1}{2} \sum_{i=1}^{3} u_i(\mu, \mathbf{q}) B_i \mathbf{q}, \\
\mathbf{n}' &= -\frac{\partial H}{\partial \mathbf{r}} = 0, \\
\mu' &= -\frac{\partial H}{\partial \mathbf{q}} = \frac{1}{2} \sum_{i=1}^{3} u_i(\mu, \mathbf{q}) B_i \mu - \frac{\partial d_3}{\partial \mathbf{q}}^T \mathbf{n},
\end{align*}
\]

where

\[
u_i(\mu, \mathbf{q}) = \bar{u}_i + \frac{\mu^T B_i \mathbf{q}}{2K_i},
\]

and

\[
B_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.
\]

Here, \( q \in \mathbb{R}^4 \) are Euler parameters (or quaternions) used to parameterize the directors \( (d_1, d_2, d_3) \in SO(3) \):

\[
R(q) = \begin{bmatrix} d_1(q) & d_2(q) & d_3(q) \end{bmatrix} = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2q_1 q_2 - 2q_3 q_4 & 2q_1 q_3 + 2q_2 q_4 \\ 2q_1 q_2 + 2q_3 q_4 & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2q_2 q_3 - 2q_1 q_4 \\ 2q_1 q_3 - 2q_2 q_4 & -2q_2 q_3 + 2q_1 q_4 & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix}.
\]
The variables \( n \in \mathbb{R}^3 \) and \( \mu \in \mathbb{R}^4 \) are the Hamiltonian conjugate variables to \( r \) and \( q \), respectively.

The ODEs (3) must be augmented with appropriate boundary conditions. In this article, we consider the elastic ring shown in Fig. 1, corresponding to boundary conditions (cf. (5)):

\[
\begin{align*}
\mathbf{r}(0) &= \langle 0, 0, 0 \rangle, \quad \mathbf{r}(1) = \langle 0, 0, 0 \rangle, \quad q_1(0) = q_2(0) = q_3(0) = 0, \\
\mathbf{q}(1) &= \langle 0, 0, -\sin(\alpha/2), -\cos(\alpha/2) \rangle, \quad \mu_4(0) = 0. \tag{6}
\end{align*}
\]

As discussed in [12], the non-intuitive condition \( \mu_4(0) = 0 \) replaces \( q_4(0) = 1 \) to remove a gauge freedom that is an artifact of the use of \( q \in \mathbb{R}^4 \) to parameterize the locally 3-dimensional group \((d_1, d_2, d_3) \in SO(3)\).

![Figure 1: Ring boundary conditions: the centerline \( r(s) \) is depicted as a tube, and is constrained to form a loop. The director \( d_i(s) \) is depicted as a ribbon, and a twist angle \( \alpha \) is imposed between \( d_i(0) \) and \( d_i(1) \).](image)

Throughout this article, a ring equilibrium denotes a solution \((r, q, n, \mu)\) to the differential equations (3) subject to boundary conditions (6). Since the approach taken here is to solve equilibrium equations (3) rather than numerically minimize the strain energy functional \( W \), these ring equilibria may be (unstable) saddle points of the strain energy in addition to the (stable) local minima of primary interest. Stability can be determined by a straightforward computation based on the conjugate point test described in [15].

Consistent with the Monte Carlo study in [8], to which we will make comparisons, we shift our focus from the angle \( \alpha \) in Fig. 1 to the related quantity link:

\[
Lk = m + \frac{\alpha}{2\pi}, \tag{7}
\]
where $m$ is the number of full turns $d_1(s)$ makes around $r(s)$ for $0 < s < 1$. More precisely, $L_k$ is computed via the Calugareanu-White formula\(^1\):

$$L_k = Tw + W_r,$$

where

$$Tw \equiv \frac{1}{2\pi} \int_0^1 w_3(s) ds,$$

$$W_r \equiv \frac{1}{4\pi} \int_0^1 \frac{(r(s) - r(\sigma))^T (r'(s) \times r'(\sigma))}{|r(s) - r(\sigma)|^3} \, ds \, d\sigma.$$

### 2.2 Applying the continuum rod model to DNA

In order to use continuum rod theory to model DNA molecules, one must determine from existing experimental DNA data appropriate values for the continuum parameters $K_i$ and $\hat{u}_i(s)$. The bending stiffness $K_1 = K_2$ (often denoted $A$) is generally derived from experimental determinations of persistence length [6]:

$$K_1 = K_2 = \frac{PRT}{N\ell},$$

where $T$ is the temperature, $R = 8.314$ J/mol-Kelvin, $P \approx 460$ Angstroms is the persistence length, $\ell \approx 3.4$ Angstroms is the helix-rise per base-pair, and $N$ is the number of base-pairs (which appears here due to the scaling in the continuum model that the rod has length one). We note that at the base-pair level, DNA almost certainly has a preferential direction for bending, but that this local anisotropy is effectively averaged by the rapid DNA intrinsic twist to yield an effective isotropic rod over the length scales of concern here (hence the widely used assumption $K_1 = K_2$); see [9] for some rigorous results on this topic. The twist stiffness $K_3$ (often denoted $C$) is less widely agreed upon. We have here chosen $K_3 = 0.8K_1$, near the middle of the accepted range.

The continuum minimum-energy centerline is derived by a smoothing of the minimum-energy centerline of a base-pair-level wedge-angle model (see, e.g., [3]). Details of this smoothing procedure can be found in [14]. One noteworthy smoothing parameter is the window-width $w$ for a tapered averaging filter that is applied to the wedge-angle centerline. The choice of $w$ corresponds to a modeling decision of the length-scale of interest in the problem. We therefore vary $w$ according to the number of base-pairs $N$. For example, when $N = 200$, we choose $w = 20$, the value used in our earlier study of 150–160 base-pair DNA [14]. On the other hand, when $N = 900$, we choose $w = 50$, since in that case, oscillations on the order of ten or fewer base-pairs are to be suppressed as the behavior of interest is curvature on the order of several tens of base-pairs.

\(^1\)The validity of this formula for $\alpha = 0$ is well-known in topology and DNA modeling. The extension to general $\alpha$ is consistently upheld numerically, and has appeared occasionally in the literature (including in [8]), but we know of no existing proof.
Having thus determined the intrinsic centerline $\tilde{r}$, the director $\hat{d}_3$ is determined via the inextensibility-unshearability assumption $d_3 = r'$. It is then straightforward to generate a continuum $(\hat{d}_1, \hat{d}_2)$ using the intrinsic twist of the wedge-angle model. Unfortunately, the rapid DNA intrinsic twist causes rapid variations in $s$ of the intrinsic curvatures $\hat{\omega}_1$ and $\hat{\omega}_2$, hindering the coarse rod discretizations that are a goal of the continuum model. Fortunately, the equilibrium configurations of an intrinsically twisted isotropic rod can be exactly recovered from the equilibrium configurations of the corresponding rod with zero intrinsic twist [14], simply by adding in the intrinsic twist after the equilibrium configurations have been determined. That is, given $(\hat{d}_1(s), \hat{d}_2(s), \hat{d}_3(s))$, compute $(\tilde{D}_1(s), \tilde{D}_2(s))$ by the conditions

\[
\tilde{D}_1(0) = \hat{d}_1(0), \\
\lambda_3(\tilde{D}_1(s), \tilde{D}_2(s), \hat{d}_3(s)) = 0, \quad \forall s \in (0, 1).
\]

Next let $\Omega(s)$ denote the angle between $\hat{d}_1(s)$ and $\tilde{D}_1(s)$. Then, given an equilibrium shape $(\tilde{D}_1(s), \tilde{D}_2(s), \hat{d}_3(s))$ for the zero-intrinsic-twist rod, an equilibrium shape of the intrinsically twisted rod is recovered by rotating $(\tilde{D}_1(s), \tilde{D}_2(s))$ about $\hat{d}_3(s)$ by $\Omega(s)$. For convenience, we write $\Omega(s)$ as an increasing continuous function with $\Omega(0) = 0$, by choosing $\Omega(s)$ in the interval $[2\pi n, 2(n+1)\pi]$ if $\hat{d}_1$ has undergone $n$ full rotations with respect to $\tilde{D}_1$ for $0 \leq \sigma \leq s$. For an $N$-base-pair DNA whose double-helix makes a full turn approximately every $\gamma_0$ base-pairs,

\[
\Omega(s) \approx \frac{2\pi N s}{\gamma_0}.
\]

For the purposes of this paper, the important implication of this twist transformation is the relation between the link $L_{k_c}$ in the continuum computation and the corresponding link $L_k$ of the real DNA molecule:

\[
L_k = L_{k_c} + L_{k_0},
\]

where $L_{k_0}$ is defined by\(^2\)

\[
L_{k_0} = \frac{\Omega(1)}{2\pi}.
\]

Within the continuum rod model, nicked and cyclized DNA are represented by differing conditions. A cyclized DNA is modeled as a stable ring equilibrium for which $L_k$ is an integer. A nicked DNA, following the model adopted by Katritch and Vologodskii [8], is modeled as a stable solution to the equilibrium equations (3) subject to closure of the centerline $r$ and tangent vectors $d_3$, but with no restrictions placed on the twist vectors $d_1$. Equivalently, if we evaluate the energy along the family of ring equilibria as $L_k$ varies, a nicked DNA is a stable ring equilibrium for which $E(L_k)$ is a local minimum. (If $E$ were not a local minimum, then the DNA could lower its energy by rotating $d_1(1)$ about $d_3(1)$, and hence would not be a minimum-energy nicked configuration).

\(^2\)Note that the symbol $L_{k_0}$ is defined in various ways in the DNA literature (e.g., as $N/\gamma_0$, or as the experimentally estimated peak position in gel filtration linking-number distribution). The values of $L_{k_0}$ under these various definitions, including the definition taken in this paper, are approximately but not precisely equal.
2.3 The perfect diagram

The principal goal of this article is to understand the multiplicities of ring equilibria in the transition from the symmetric perfect problem \((\bar{u}_e = 0)\) to various imperfect problems \((\bar{u}_e \neq 0)\). Physically, the perfect problem involves a uniform isotropic rod that is intrinsically straight, while the imperfections considered here involve the introduction of intrinsic curvature. The perfect problem has been the subject of study by many authors; see [10, 12, 20, 17] for a discussion of the literature. Here we follow the notation and formulation that are described in detail in [4, 12].

2.3.1 The set of perfect equilibria

Figure 2 shows a portion of the set of ring equilibria for the perfect problem with \(K_3 = 0.8K_1\) for various continuum links \(L_{k_e}\). For other values of \(K_3\), the diagram is qualitatively the same, although the exact positions of points \(A\)–\(D\) and the shapes of the branches change. There exists an infinite family of branches of perfect equilibria other than those shown in Fig. 2, but since they have higher energy and/or are unstable, they will not be of concern here.

This perfect bifurcation diagram is described in detail in [12]. The trivial branch containing points \(A\), \(B\), and \(C\) corresponds to configurations with circular centerlines and constant twist rates. Connected to the trivial branch is a nontrivial branch \(ADC\) corresponding to ring equilibria with nonplanar centerlines (with the single exception of the planar figure-eight located at \(D\)). Note that the link jumps discontinuously by two when the rod passes through itself, as occurs at \(D\).

Solid lines in Fig. 2 denote stable ring equilibria, while dashed lines denote unstable equilibria. Rogers (1997) [19] showed that for arbitrary \(K_3\), the segment \(ABC\) of the trivial branch contains stable solutions, and all other trivial solutions are unstable. Further, for \(K_3/K_1 < 1.375\), all nontrivial solutions are unstable, while for \(K_3/K_1 > 1.375\), some or all of the solutions on the nontrivial branch \(ADC\) are stable.

2.3.2 The register symmetry

Due to material symmetries of the perfect rod, every point on the perfect diagram actually represents an entire manifold of ring equilibria, all with the same energy. Since the equilibria on this manifold are inter-related by the action of a symmetry transformation, we call the manifold an orbit, following the algebraic terminology for the image of a group action. Most crucial for this paper is the symmetry transformation we call register following Lavery. This register symmetry involves spinning the rod at every point by some angle \(\theta\) about its centerline. In terms of the fundamental rod description in Sec. 2.1, this amounts to rotating \((d_1(s), d_2(s))\) by \(\theta\) about \(d_3(s)\) at each \(s\).\(^3\) In the language of DNA, the register determines whether the major groove at a particular base-pair of a circular DNA will face toward the center of the circle, away from this center, or somewhere in between.

\(^3\)Note that the register transformation thus described does not respect the q boundary conditions in (6), but this is easily remedied by applying a compensating rigid-body rotation.
Figure 2: Portion of the bifurcation diagram for the perfect problem for $K_3 = 0.8K_1$. As the link $L_k$ is varied, the energy $E$ and link $L_k$ of the ring equilibria are plotted. Solid curves denote stable equilibria, while dashed curves denote unstable equilibria. The trivial branch $ABC$ corresponds to planar equilibria with circular centerlines, while the nontrivial branch $ADC$ corresponds to nonplanar equilibria.

Using this register transformation, we find that every point on the trivial branch $ABC$ actually represents a circle of ring equilibria.\textsuperscript{4} Our focus will be to determine the fate of this circular orbit of degenerate solutions when intrinsic curvature is added.

2.4 Perturbation of the trivial branch

As described in the previous section, when $\dot{u}_i = 0$, the two-point boundary value problem (3+6) at any fixed $\alpha_c$, or corresponding $L_k$ determined by (7), has a circle of degenerate trivial solutions, parametrized by the register $\theta$. For $\dot{u}_i \neq 0$, this circle of solutions yields in general only a finite set of solutions. By a perturbation expansion (as described in [13]) of this

\textsuperscript{4}Similarly, the register transformation plus a second transformation based on translation along the arclength of the rod imply that each point on the nontrivial branch of Fig. 2 represents a \textit{torus} of ring equilibria [13]. However, since for $K_3/K_1 = 0.8$ this nontrivial branch is unstable, we will not have need in this paper to consider the nature of the splitting of the nontrivial branch (although the images after perturbation of the nontrivial branch will certainly appear in all computed diagrams shown).
two-point boundary value problem, we find that if at least one of the following two integrals is nonzero:

\[
I_1(Lk_c) = \int_0^1 \left[ \dot{u}_1(s) \sin(2\pi Lk_c s) + \dot{u}_2(s) \cos(2\pi Lk_c s) \right] ds,
\]

\[
I_2(Lk_c) = \int_0^1 \left[ \dot{u}_1(s) \cos(2\pi Lk_c s) - \dot{u}_2(s) \sin(2\pi Lk_c s) \right] ds,
\]

then the orbit of trivial solutions at link \( Lk_c \) will yield exactly two imperfect ring equilibria for infinitesimally small \(|\dot{u}_i|\).\(^5\)

Further, of these two equilibria, only one is expected to be stable, as seen in Fig. 3. Actually

![Bifurcation Diagram](image)

**Figure 3**: Portion of the bifurcation diagram for an imperfect rod. Stability is indicated by line style as in Fig. 2. Approximate images of points A, B, C, and D are marked. Each circle labelled A or C represents two images of the respective point in the perfect diagram. The stable trivial branch ABC from Fig. 2 yields two branches in the imperfect diagram, one stable and one unstable, while the nontrivial branch ADC yields four branches.

proving this assertion about stability involves computations beyond the perturbation expansion, but the reason for it is clear from a finite-dimensional example. Take a surface of revolution as shown in Fig. 4 containing a circular “valley” of neutrally-stable critical points. (This valley is

\(^5\)Although this result involves the typical assumptions and cautions of any formal perturbation expansion, its predictions have been consistently upheld by our numerical computations, see e.g., Fig. 3.
exactly analogous to the circular orbit of solutions existing at each point on the stable trivial branch in the perfect diagram). Each critical point is a local minimum, but contains a single flat direction along the path tangent to the orbit. When we add a small perturbation to the surface of revolution as in Fig. 5 (analogous to adding intrinsic curvature to the perfect problem), the circle of critical points is perturbed to only two critical points in the tilted surface. Further, of these, only one is a local minimum, while the other is a saddle point. Of course, we can engineer perturbations to the surface of revolution so that the perturbed surface has more than two critical points, or more than one local minimum; such special perturbations are exactly analogous to intrinsic shapes for which $I_1(Lk_e) = I_2(Lk_e) = 0$. 
Figure 4: A finite-dimensional example of a circular orbit of critical points. The surface $z = r^4/4 - r^2/2$ ($r = \sqrt{x^2 + y^2}$) has a circle of critical points $x^2 + y^2 = 1$. The surface and its contour plot are shown, with the set of critical points drawn with a thick line.

### 2.5 Implications of perturbation results for curved DNA

By the very nature of perturbation expansions, the results described in Sec. 2.4 are only directly applicable to small intrinsic curvatures. However, for small-to-moderate curvatures, the perturbation results should still correlate with observed behavior. The perturbation expansion predicts that for infinitesimal curvatures, the only rod shapes that could possibly yield multiple stable ring equilibria at link $Lk_c$ are those which satisfy the special conditions $I_1(Lk_c) = I_2(Lk_c) = 0$. For small curvatures, we can thus conjecture that there should be a good correlation between shapes yielding multiple stable ring equilibria and shapes with small $|I_1|, |I_2|$. This is precisely the conjecture we verify in Sec. 4.

Further, this correlation should in principle be better for smaller intrinsic curvatures than for larger ones. This is again verified in Sec. 4 by comparing results for different lengths of DNA (longer DNA have more overall curvature, and hence larger local curvatures in the continuum model in which total length is scaled to one).
3 Methods

3.1 Design of random sequences

Random DNA sequences with equal probability of A, C, G, or T at each base were generated, using a random number algorithm based on the ran1 method [18]. For this study, 5017 random 200-base-pair and 2176 random 900-base-pair sequences were used.

3.2 Intrinsic shape parameters

For a given base-pair sequence, a preliminary non-continuum intrinsic shape was determined using the standard wedge-angle model, see e.g. [3]. We confined this study to dinucleotide wedge angles based on those in [3]. Our first angle set (which we call BT1) uses the exact intrinsic tilts and rolls derived in [3], while our second angle set (BT2) scales these tilts and rolls by approximately 0.6, thereby producing less intrinsic curvature (or in other words increasing the static persistence length by approximately $1/((0.6)^2$ from 168 to 467 nm). This scaling is used here for purely theoretical reasons, to demonstrate the effects of smaller intrinsic curvature, but it is inspired by the actual scaling of the wedge angles of [3] used by Kahn, Crothers et. al. [7]. In contrast to the base-pair-dependent intrinsic twists found in [3], our wedge-angle sets BT1 and BT2 use a common intrinsic twist of $34.45^\circ$ per base-pair for all dinucleotide steps. This change is purely a matter of convenience so that the total intrinsic twist $L_{k0}$ (see (11)) will be
3.3 Continuum computations

Each preliminary wedge-angle intrinsic shape was smoothed via the procedure summarized in Sec. 2.2 to yield a continuum intrinsic shape \( \mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s) \), plus the auxiliary function \( \Omega(s) \) needed to recover the true intrinsic shape from this intrinsically untwisted one. From this, the unstressed strains \( \dot{u}_i(s) \) appearing in the equilibrium equations (4) are determined directly from the definition (1).

The boundary value problem (BVP) (3.6) is solved via parameter continuation using the software package AUTO [5]. To this end, we insert a parameter \( \epsilon \) in front of \( \dot{u}_i \) in (4), and initially set \( \epsilon = 0.01 \). We then choose a solution (known in closed-form) \( \mathbf{z}^0(s) \equiv (\mathbf{r}^0(s), \mathbf{q}^0(s), \mathbf{n}^0(s), \mu^0(s)) \) on the stable trivial branch of the perfect problem (see Fig. 2), say at link \( L_{h0}^0 \) (with corresponding angle \( \alpha_0 \)). The function \( \mathbf{z}^0(s) \) is thus a solution of the BVP when \( \epsilon = 0 \), and we then derive from it an approximate solution \( \mathbf{z}^0(s) + 0.01\mathbf{z}^1(s) \) to the BVP when \( \epsilon = 0.01 \) using the perturbation expansion described in [13]. This function \( \mathbf{z}^0(s) + 0.01\mathbf{z}^1(s) \) serves as the starting point for the parameter continuation computation in AUTO. (Alternatively, one may use the simpler starting point \( \epsilon = 0, \mathbf{z} = \mathbf{z}_0 \), at the cost of a somewhat more expensive computation; see the discussion in [13].)

Using AUTO, we compute Branch 1: solutions to the BVP as \( \epsilon \) is increased slowly from 0.01 to 1 (\( \alpha = \alpha_0^0 \) fixed). Then, holding \( \epsilon = 1 \) fixed, AUTO computes solutions to the BVP as \( \alpha \) varies, thus computing Branch 2, the imperfect diagram. For the first step of Branch 2, \( \alpha \) is required to increase, but as the branch progresses it may change directions. Computation is stopped when either (a) the branch closes up on itself, (b) a point of index 4 is reached, or (c) the number of computed points reaches a user-defined maximum. If either (b) or (c) occurs, the procedure returns to the end of Branch 1 and repeats the computation of Branch 2, now with \( \alpha \) initially decreasing, with the same stopping conditions (a–c). In this way, we are quite confident of finding all stable BVP solutions lying on the same component as the end of Branch 1.

In about 90% of the molecules, the above computation determines all stable BVP solutions. However, there are exceptional intrinsic shapes for which stable solutions exist on two or more components in the imperfect diagram, and the above procedure will only locate one of these components. To remedy this difficulty, we repeat the above procedure for several values of \( L_{h0}^0 \). At the end of Branch 1 in each case, we check if the solution already exists on a previously computed Branch 2, and if not, we compute a new Branch 2, and append it to the previous one. By the very nature of parameter continuation computations, one is never guaranteed to have found all BVP solutions, but by this multiple-starting-point procedure, we believe that all stable solutions are located in more than 99% of the molecules.

For each solution on the imperfect diagram, various integrals of interest such as link and energy are computed numerically. In addition, stability is determined by solving a 54-dimensional initial value problem of ODEs (cf. [15]). Finally, stable nicked and stable cyclized minima are located numerically. Cyclized equilibria are easily detected, since it suffices to search the stable branches in the diagram for points where \( \alpha \) crosses a multiple of \( 2\pi \). Nicked equilibria are
slightly more tricky to locate because a section of the diagram where \( E \) is nearly flat (inflection points or local maxima) may be mistaken for a local minimum due to small numerical errors (e.g. in the fourth or fifth decimal place) due to the tolerances set in AUTO. To avoid such problem points, we look for quintuplets of adjacent solutions so that (a) all five solutions are stable, (b) the energy profile along this quintuplet is down-down-up-up, and (c) the twist moment (computable from the unknowns \( z \) crosses zero between the first and fifth points. Criterion (c) is based on the fact that within the continuum theory one may prove that \( E \) being a local minimum for this problem implies that the twist moment vanishes.

The computation of the intrinsic shape and the associated integrals \( I_1, I_2 \) requires less than a minute on a single Sparc CPU. The time required to compute a bifurcation diagram varies with the complexity of the diagram. Roughly 50\% of the time, the stable solutions lie on a simple closed loop which can be computed in approximately 5–10 minutes. For more complicated diagrams, especially when stopping conditions (b) or (c) are invoked, runs can take as long as 30–45 minutes. This CPU time is not dependent on the number of base-pairs, as opposed to Monte Carlo simulations described below. Once the bifurcation diagram is known, it takes just a few seconds to extract the nicked and cyclized equilibria.

3.4 Monte Carlo simulations

We used the Metropolis Monte Carlo procedure described in [8] without modification of the source code. This program is designed to simulate the thermodynamic equilibrium distribution for nicked conformations of intrinsically curved DNA, and as a result we have confined the comparison with continuum computations to the nicked case.

Intrinsic curvature parameters were determined using the dinucleotide angle sets BT1 and BT2 (see Sec. 3.2). Bending and torsional rigidity constants between the Monte Carlo and continuum computations were related via the equations described in Sec. 2.2. The bending rigidity constant used corresponded to a Kuhn statistical length of 100 nm. The torsional rigidity constant was set to \( C = 1.61 \times 10^{-19} \text{ erg-cm} \) to ensure compatibility with continuum computations (run at \( K_3/K_1 = 0.8 \)). The DNA effective diameter was set to 2 nm, although self-contact does not appear to play a role in the effects investigated here (since we are looking at relatively relaxed nicked conformations). We chose elemental Monte Carlo segments of 10 bp each, so there were 20 total segments the 200 bp simulations and 90 total segments for the 900 bp simulations. The temperature was set to 293.15 K for all runs.

The segmented chain is subject to three types of moves; after every move the writhe and the twist of the new conformation are calculated, and from them the linking number \( L_k \) according to the Calugareanu-White equation (8). The cumulative \( L_k \) distribution was updated every 100 moves.

As the minimal number of moves necessary to reach thermodynamic equilibrium may vary from sequence to sequence, we chose to run all Monte-Carlo simulations with \( 50 \times 10^6 \) steps, a number apparently sufficient to guarantee that the parameters we measure on the \( L_k \) distributions (peak positions and areas) have equilibrated. A 20-segment (200 bp) simulation on a single Sparc processor requires about 210 minutes, and a 90-segment simulation about 71 hours (the CPU time scales as the square of the number of segments).
We used the Matlab `leastsq` function to fit the $L_k$ distribution with one or more Gaussians, whose positions and areas are extracted for comparison with continuum results.

4 Results

4.1 Bifurcation diagrams exhibiting multiple equilibria

The prevailing wisdom about nicked or cyclized equilibria in minicircles has been that only one equilibrium will occur. However, as emphasized by Katritch and Vologodskii [8], this prevailing wisdom is biased by the fact that most early studies treated what we have called here the perfect problem (the case of an intrinsically straight rod), and that in fact the story is more complicated for rods with intrinsic curvature.

Indeed, for an intrinsically straight rod, the bifurcation diagram of DNA equilibria is simply the perfect diagram in Fig. 2 shifted horizontally by the constant $Lk_0$ (to convert $Lk_0$ to $Lk$). Thus, there is exactly one stable nicked equilibrium (the image after horizontal shift of point $B$ in Fig. 2), a single stable cyclized equilibrium for each integer $Lk$ that falls between the links of points $A$ and $C$ after shift by $Lk_0$, and no other stable cyclized configurations of other links.

However, when we add intrinsic curvature, the resulting imperfect diagram need not look as simple as this shifted perfect diagram, as shown in Figs. 6–8. In these figures, we have selected three examples from our database of random 200-base-pair molecules, determined their continuum intrinsic shapes using the wedge-angle sets BT1 and BT2, and computed the branches of the resulting bifurcation diagrams containing stable equilibria.

In the first case in Fig. 6, although there are clearly some changes introduced by the intrinsic curvature (e.g., energies are shifted downward, the range of links covered by the stable solutions is slightly reduced), the diagram retains the qualitative shape of the loop ABCD from the perfect diagram. Indeed, there is once again a single stable nicked equilibrium (marked with a black triangle), and at most a single stable cyclized equilibrium (marked with a gray circle) at each link. Quite naturally, the diagram for the smaller angle set BT2 resembles the perfect diagram more closely than the diagram for BT1.

On the other hand, in Fig. 7, the intrinsic curvature induces a qualitative change in the diagram near the bottom, with the result that for both angle sets BT1 and BT2, there are two stable nicked equilibria. The diagram for the angles BT2 straightens out this kink (pulling it closer to the perfect diagram) so that the second nicked equilibrium is on the verge of disappearing.

Similarly, in Fig. 8, we see an induced kink in the diagram in the neighborhood of link 18, with the result that for angles BT1 two stable link 18 cyclized equilibria exist. Here again, the kink is removed for angles BT2, but in other cases, it can remain.

Figures 6–8 illustrates the (unsurprising) fact that larger wedge-angles create imperfect diagrams further from the perfect diagram and hence more likely to exhibit multiple equilibria. The same effect can be seen by considering longer DNA, for which the effective local curvatures for the length-one continuum rod are larger. Both of these effects are illustrated by the statistical study in Sec. 4.2.
Figure 6: Qualitatively simple imperfect diagrams for a 200-bp DNA molecule modeled by angle sets BT1 and BT2. The perfect diagram is superimposed. Stability is indicated as before by line style (stable equilibria on solid lines, unstable equilibria on dashed lines). Stable nicked DNA (for the imperfect problem) are denoted by black triangles, and stable cyclized DNA by gray circles. As in the perfect diagram, there is a single stable nicked DNA, and at most one stable cyclized DNA at each link.

The only way to detect with certainty the existence of multiple equilibria is to directly compute the set of equilibria as in Figs. 6–8. However, armed with the perturbation analysis of Sec. 2.4, we have a simple calculation that should be correlated with the existence of multiple equilibria, namely the smallness of the integrals $I_1(Lk_c)$ and $I_2(Lk_c)$. If multiple nicked equilibria are sought, then one should seek intrinsic shapes yielding small $I_1(0)$ and $I_2(0)$, since the nicked equilibrium of the perfect problem occurs at $Lk_c = 0$. For example,

$$I_1(0), I_2(0) = 0.36, 1.25 \quad \text{in Fig. 6a,}$$

$$I_1(0), I_2(0) = -0.04, 0.83 \quad \text{in Fig. 7a.}$$

---

We note in the Discussion that the existence of multiple peaks in the Monte Carlo simulations of nicked DNA reported in [8] are certainly correlated to the existence of multiple nicked equilibria, but the correlation is not perfect, since the Monte Carlo simulations sample the entire energy surface and do not directly compute equilibria. A similar statement could be made for Monte Carlo simulations of cyclized DNA.
Figure 7: Imperfect diagrams showing a kink near the stable nicked equilibrium of the perfect problem. The intrinsic shape is for a 200-bp DNA modeled by angle sets BT1 and BT2. The perfect diagram is superimposed, stability is indicated by linestyle, and stable nicked and cyclized equilibria are labeled as in Fig. 6. For each angle set, the kink introduces a second stable nicked equilibrium for the imperfect problem, although for angles BT2, the kink is straightened out so that this second nicked equilibrium is on the verge of disappearing.

On the other hand, multiple cyclized equilibria of link $Lk$ should correspond to small values of $I_1(Lk - Lk_0), I_2(Lk - Lk_0)$ by (10). For example,

$$I_1(18 - Lk_0), I_2(18 - Lk_0) = -1.29, -1.27 \quad \text{in Fig. 6a,}$$

$$I_1(18 - Lk_0), I_2(18 - Lk_0) = 0.59, -0.70 \quad \text{in Fig. 8a.}$$

Note that since $Lk_0$ will vary slightly with intrinsic shape, the values of $Lk_0$ differ slightly between Figs. 6a and 8a.

We stress that the three molecules used in Figs. 6–8 were chosen to illustrate most clearly the way in which multiple nicked or cyclized equilibria manifest themselves in bifurcation diagrams. Many of the bifurcation diagrams we computed for random DNA sequences are more complicated than these Figures, but the general pattern of introduction of multiple equilibria via branch kinking is persistent across the range of computed examples.

In the same vein, the discussion of the “typical” connection between multiple equilibria and smallness of $(I_1, I_2)$ in this Section has thus far been purely anecdotal. The relation
Figure 8: Imperfect diagrams showing a kink near link 18. The intrinsic shape is for a 200-bp DNA modeled by angle sets BT1 and BT2. The perfect diagram is superimposed, stability is indicated by linestyle, and stable nicked and cyclized equilibria are labeled, as in Fig. 6. The kink introduces a second stable cyclized equilibrium of link 18 for angle set BT1, but with angle set BT2 the kink is straightened and only one stable link-18 equilibrium occurs.

between \((I_1, I_2)\) and equilibrium multiplicity for non-infiniteesimal intrinsic curvatures is only a correlation and not a certainty, and as such can only be justified by a statistical study, the results of which are presented in the next section.

### 4.2 Statistical study of multiple equilibria

We generated bifurcation diagrams for 5017 random 200-base-pair sequences and 2176 random 900-base-pair sequences with the wedge angle set BT1 and analyzed them as described in Sec. 3. In addition, we selected 1000 of the 200-base-pair sequences for computation with the wedge angle set BT2. The number of stable nicked equilibria for these three datasets is reported in Table 1.

This Table demonstrates that the probability to finding multiple equilibrium increases with intrinsic curvature as well as with the length of the DNA. For instance, angle set BT2 produces smaller intrinsic curvatures than BT1, and accordingly, we see in Table 1 that multiple stable
<table>
<thead>
<tr>
<th></th>
<th>number of molecules</th>
<th>number of equilibria (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 bp (BT1)</td>
<td>5017</td>
<td>56.3</td>
</tr>
<tr>
<td>200 bp (BT2)</td>
<td>1000</td>
<td>89.2</td>
</tr>
<tr>
<td>900 bp (BT1)</td>
<td>2176</td>
<td>35.9</td>
</tr>
</tbody>
</table>

Table 1: The effect of length and intrinsic curvature on the likelihood of multiple stable nicked equilibria

nicked equilibria occur only half as often with BT2 as compared to BT1. Similarly, within a fixed angle set BT1, the increase in DNA length from 200 to 900 bp promotes the likelihood of finding multiple stable nicked equilibria from 23 in the length-one continuum rod in the 900-bp case.

We performed a statistical analysis of the correlation of multiple stable nicked or cyclized equilibria with the size of $(I_1(Lk_0), I_2(Lk_0))$, for the appropriate values of $Lk_0$. In each study, the distribution of $I_1(Lk_0)^2 + I_2(Lk_0)^2$ was first separated into deciles, so that the first decile contains the 10% of the molecules with lowest values of $I_1(Lk_0)^2 + I_2(Lk_0)^2$, the second decile the 10% with next lowest values, etc. Within each decile, the fraction of molecules with multiple stable equilibria was then determined. These correlations are shown in Figs. 9 and 10 for nicked equilibria and cyclized equilibria respectively. The cyclized equilibria considered in Fig. 10 are those with link 20 for 200 base-pair DNA and link 87 for 900 base-pair DNA (these are the links one greater than the nearest integer to $Lk_0$). The correlations for other links are similar.

The correlation in the case of cyclized equilibria is quite striking. Even for 900-base-pair molecules, the likelihood of finding multiple equilibria with link $Lk$ is highly coupled to the size of $I_1^2 + I_2^2$ at $Lk_0 = Lk - Lk_0$. For the smallest angle set 200-BT2 especially, the correlation is nearly perfect, with all 10 (out of 1000) molecules exhibiting multiple cyclized equilibria occurring in the first decile. Indeed, the correlation for 200-BT2 is even better than shown in Fig. 10, as can be seen if the first decile into percentiles (lowest 1% of $I_1^2 + I_2^2$, next lowest 1% of $I_1^2 + I_2^2$, etc.) (data not shown).

For nicked equilibria, the story is a little more complicated, because kinks in the bifurcation diagram over a wide range of $Lk$ can eventually, for sufficiently high intrinsic curvature, create a second nicked equilibrium. Indeed, for 900-base-pair molecules, over 60% of all DNA exhibited two or more stable nicked configurations. Nevertheless, the correlation to the size of $I_1(0)^2 + I_2(0)^2$ is still clear from Fig. 9, and particularly for 200-base-pair DNA.

### 4.3 Geometric Interpretation of $I_1, I_2$

The quantities $I_1, I_2$ from Eq. (12) that arise naturally from an analysis of the continuum equilibrium equations are global averages of the local intrinsic curvatures $\theta_1$ and $\theta_2$. As such, these averages do not lend themselves to immediate geometric intuition, but with some investigation, we can find relations to more intuitive geometric properties.

Recall first that the continuum intrinsic shape was derived by starting with a sequence
Figure 9: Fraction of DNA with multiple stable nicked equilibria as a function of the size of $I_1(0)^2 + I_2(0)^2$. The values of $I_1(0)^2 + I_2(0)^2$ have been grouped into deciles, with the first decile holding the 10% of samples with lowest values, the second decile the next lowest 10%, etc. The absolute ranges of $I_1(0)^2 + I_2(0)^2$ vary with dataset. For 200-BT1, the minimum value was 0.0004, maximum value 26.3, median 2.25. For 200-BT2, the minimum value was 0.002, maximum value 7.8, median 0.79. For 900-BT1, the minimum value was 0.007, maximum value 110, median 10.8.

of $N$ director frames $(d_1^{(i)}, d_2^{(i)}, d_3^{(i)})$, whose relative orientations were determined from base-pair-dependent wedge-angle parameters. The resulting wedge-angle intrinsic shape was then smoothed to give a continuum intrinsic shape $(D_1(s), D_2(s), D_3(s))$ (that also removed the rapid intrinsic twist). We may now think of discretizing this continuum shape to give $N$ director frames $(D_1^{(i)}, D_2^{(i)}, D_3^{(i)})$, at each $s = i/N$ for $i = 0, \ldots, N-1$. These new frames trace out a similar centerline as the original frames but without rapid local bending fluctuations and without rapid intrinsic twist.

Consider the rotation between $(D_1^{(i)}, D_2^{(i)}, D_3^{(i)})$ and $(D_1^{(i+1)}, D_2^{(i+1)}, D_3^{(i+1)})$ and write it in the standard way as a product of fundamental rotations by three Euler angles $(\theta^{(i)}, \phi^{(i)}, \tau^{(i)})$ about $D_1^{(i)}, D_2^{(i)}$, and $D_3^{(i)}$. In this smoothed shape, these three angles will be small, and so one may easily show that $\theta_1^{(i)}(\text{deg}) \approx N\theta^{(i)}$ and $\theta_2^{(i)}(\text{deg}) \approx N\phi^{(i)}$. Thus, if we discretize the integrals $I_1(0)$ and $I_2(0)$ as sums over the $N$ new frames, we find

$$I_1(0) \approx \sum_{i=1}^{N} \theta^{(i)}, \quad I_2(0) \approx \sum_{i=1}^{N} \phi^{(i)}.$$

If the intrinsic shape is roughly planar and not too bent, one may approximate yet again to say that $I_1(0), I_2(0)$ are approximately equal to the Euler angles with respect to $D_1^{(1)}, D_2^{(1)}$ respectively of the overall rotation between the first and last base-pairs. By basic analysis of Euler angles, we would then have that the end-to-end cosine, i.e., the cosine of the angle between the initial and final tangent vectors, is roughly equal to $\cos(I_1(0)) \cos(I_2(0))$, or approximately

21
Figure 10: Fraction of DNA with multiple stable cyclized equilibria as a function of the size of $I_1(L_{k_c})^2 + I_2(L_{k_c})^2$. Here we count only link-20 equilibria for 200-base-pair DNA or link-87 equilibria for 900-base-pair DNA, with corresponding continuum links $L_{k_c} = 20 - L_0, 87 - L_0$. The values of $I_1(L_{k_c})^2 + I_1(L_{k_c})^2$ have been grouped into deciles, with the first decile holding the 10% of samples with lowest values, the second decile the next lowest 10%, etc. The absolute ranges of $I_1(L_{k_c})^2 + I_2(L_{k_c})^2$ vary with dataset. For 200-BT1, the minimum value was 0.0003, maximum value 34.4, median 2.23. For 200-BT2, the minimum value was 0.0003, maximum value 8.5, median 0.83. For 900-BT1, the minimum value was 0.0005, maximum value 111, median 10.1.

equal to $1 - (I_1(0)^2 + I_2(0)^2)/2$. Clearly many approximations are involved in this analysis, but nevertheless, a decent correlation exists between $1 - (I_1(0)^2 + I_2(0)^2)/2$ and the end-to-end cosine over our entire database of molecules (data not shown here).

This connection is similarly seen if we look at the intrinsic shapes of molecules with particularly high or low values of $I_1(0)^2 + I_2(0)^2$, as in Fig. 11. For 200-base-pair molecules, the twenty molecules with the largest values of $I_1(0)^2 + I_2(0)^2$ (Fig. 11A) are intrinsically C-shaped (with relatively small end-to-end cosines), while the twenty with the smallest values (Fig. 11B) are intrinsically S-shaped (with end-to-end cosines near one). We have seen in Sec. 4.2 that small values of $I_1(0)^2 + I_2(0)^2$ promote the existence of multiple stable nicked equilibria, so Figs. 11AB reinforce the theme noted by Katritch and Vologodskii that S-shaped DNA are more likely than C-shaped DNA to yield multiple nicked equilibria. The quantity $I_1(0)^2 + I_2(0)^2$ provides a precise categorization of this geometric phenomenon and allows analysis of cases falling between the “C” and “S” extremes.

When we look at the more irregular 900-base-pair intrinsic shapes in Figs. 11CD, the S-versus-C distinction is less apt, but one can still clearly see a qualitative difference between molecules with large and small $I_1(0)^2 + I_2(0)^2$, with large values (Fig. 11C) yielding highly globally bent DNA (with negative end-to-end cosines) and small values (Fig. 11D) yielding relatively straight DNA.
Figure 11: Intrinsic shapes for molecules with the highest (Figs. A and C) and lowest (Figs. B and D) values of $I_1(0)^2 + I_2(0)^2$. Figs. A and B are for 200 bp DNA; here, the molecules with large $I_1(0)^2 + I_2(0)^2$ (Fig. A) are C-shaped, while those with small $I_1(0)^2 + I_2(0)^2$ (Fig. B) are roughly S-shaped. Figs. C and D are for 900 bp DNA; the molecules with large $I_1(0)^2 + I_2(0)^2$ (Fig. C) have global bends of at least 180 degrees, while those with small $I_1(0)^2 + I_2(0)^2$ (Fig. D) are globally much straighter.

4.4 Metropolis Monte-Carlo

Is there any correlation between the ring equilibria computed by solving the DNA rod equations (3) and the thermodynamic equilibrium distributions of conformations generated by the Metropolis Monte-Carlo (MMC) approach? In other words, how do the energy $E$ and the linking number $L_k$ of the stable ring equilibria compare with their Monte-Carlo counterparts? For instance, can one predict how many peaks the $L_k$ distribution computed by the MMC algorithm should exhibit, as well as their positions and relative intensities, based on the knowledge of the elastic energies and $L_k$ values of the static equilibria?

At least qualitatively, there does seem to be a clear correlation between the results of those two approaches. For DNA with 200 bp (see Figure 12) and 900 bp (see Figure 13), using the
angle set BT1, the simplest bifurcation diagrams (those with one single nicked equilibrium) generally correspond to single-peaked $L_k$ distributions, and the more intricate diagrams (those with more swirls and loops) tend to give $P(L_k)$ distributions with more than one peak. This correlation is demonstrated on a more quantitative basis in the next Section.

4.4.1 Number of peaks versus number of equilibria

*Single equilibria.* As a control, we first selected at random twenty-five 200-bp sequences for which the continuum approach established a single nicked equilibrium. The Metropolis Monte-Carlo $L_k$ distributions were obtained for these sequences and exhibited a single peak without exception (see, e.g., Fig. 12A). The same conclusion was reached for twenty-five 900-bp sequences (see, e.g., Fig. 13A). The following table summarizes the peak locations $\mu$ and widths $\sigma$ of these $L_k$ distributions, when fit to Gaussians:

<table>
<thead>
<tr>
<th></th>
<th>$\mu$ (rms)</th>
<th>$\sigma$ (rms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 bp</td>
<td>19.15 (0.11)</td>
<td>0.18 (0.02)</td>
</tr>
<tr>
<td>900 bp</td>
<td>86.17 (0.79)</td>
<td>0.54 (0.05)</td>
</tr>
</tbody>
</table>

*Double equilibria.* Next, we selected at random two sets of 50 molecules (one set of 200 bp and one set of 900 bp) from among those whose continuum bifurcation diagrams exhibited two stable nicked equilibria. The $L_k$ distributions of these molecules were computed via Metropolis Monte Carlo. These distributions come out either with two nicely separated smooth peaks (two maxima surrounding a minimum, as in Figs. 12D and 13B), as a single smooth peak (as in Fig. 12B), or as an intermediate case, called a “shoulder” for convenience (as in Fig. 12C). The number of occurrences of each of these situations is reported in the following table:

<table>
<thead>
<tr>
<th></th>
<th>2 peaks</th>
<th>shoulder</th>
<th>1 peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 bp</td>
<td>15</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>900 bp</td>
<td>11</td>
<td>36</td>
<td>3</td>
</tr>
</tbody>
</table>

An analysis of this data (results not shown, but see http://lcvwww.epfl.ch/~pfurier/Public/conditional_twopeaks.html) suggests two primary reasons that a molecule with two equilibria will fail to produce two distinct peaks in the Monte Carlo $P(L_k)$ distribution:

1. The $L_k$ values of the two equilibria lie too close to each other (compared to the values of $\sigma$ of each roughly Gaussian component) for the Monte-Carlo protocol to be able to resolve them (see, e.g., the two central equilibria in Fig. 13D).

2. The relative difference in elastic energy between the two equilibria is so large that the highest energy equilibrium has almost no probability of occurring in the Monte-Carlo runs (see, e.g., Figs. 12B and Figs. 13A).

Finally, we selected the ten molecules from the 200-BT2 database exhibiting the largest splits in $L_k$ values of two continuum equilibria. Of the Monte Carlo $L_k$ distributions of these 10 molecules, seven exhibited shoulders, while three gave two nicely separated smooth peaks. This
Figure 12: Comparison of continuum bifurcation diagrams and Monte Carlo linking distributions for four 200-bp DNA (angle set BT1).

This test thus demonstrates the efficiency of the continuum computations as a probe for interesting Monte Carlo behavior, because at the range of low intrinsic curvature seen in 200-BT2, Kratich
Figure 13: Comparison of continuum bifurcation diagrams and Monte Carlo linking distributions for four 900-bp DNA (angle set BT1).

and Vologodskii reported they were unable to find evidence of multiple equilibria in their Monte Carlo simulations, but here, with the aid of rapid continuum screening of thousands of
candidates, we have located ten examples with nontrivial Monte Carlo Lk profiles.

Triple and Quadruple equilibria. Although our database of 200 bp sequences contains some (24) examples giving three stable continuum nicked minima, none exhibits three equilibria sufficiently well-separated in $Lk$ and close in $E$ to lead to triple peaks in Metropolis Monte Carlo simulations. However, when we search our 900 bp database, we were able to find several examples that do in fact yield triple-peaked Monte Carlo profiles (see Figs. 13C and D). The continuum computations also located 26 900-bp molecules (out of 1276) with 4 stable nicked equilibria. However, for the same reasons cited above for triple-equilibria in 200 bp sequences, the number of peaks in the Monte Carlo Lk distributions for these molecules is only two or three (see, e.g., Fig. 13E).

4.4.2 Lk correlation

Do the $Lk$ values of individual equilibria match the peak positions in their corresponding $P(Lk)$ distributions? To investigate this question, we studied molecules with single equilibria, as well as molecules with two equilibria that also exhibit two peaks in their Monte Carlo $P(Lk)$ distributions. Molecules exhibiting shoulders were discarded due to the large uncertainty of their peak positions $\mu$. In Fig. 14, for both the 200 bp and 900 bp cases, we compare the equilibrium values of $Lk$ (on the $x$ axis) to the centers of the Gaussians fit to the $P(Lk)$ distribution (on the $y$ axis). The global correlation coefficients are 0.992 ($n = 55$) and 0.962 ($n = 47$) for the 200 bp and the 900 bp cases, respectively. These correlation coefficients indicate a very good agreement between the $Lk$ values calculated by the two methods, within the accuracy of the $Lk$ determination (see Sec. 4.4.4).

4.4.3 Energy correlation for double equilibria

Continuum ring equilibria (stable minima) are characterised by well defined elastic energies $E$. The Monte-Carlo approach is using an equivalent energy function to sample, via the Metropolis
algorithm, the configurational space of those nicked minicircles, so that in theory, the probability of accessing a configuration with energy $E$ is proportional to $\exp(-E/RT)$. For molecules with two stable equilibria with elastic energies $E_1$ and $E_2$, where the MMC algorithm produces a $P(Lk)$ distribution with two peaks (or a main peak plus an obvious shoulder to fit), we might then hope that the ratio of intensities of the $P(Lk)$ peaks will be related to the relative elastic energies of the stable ring equilibria by this same expression: i.e., the ratio of intensities will be $\exp(-(E_2 - E_1)/RT)$. This correlation is reported in Fig. 15.

![Correlation between the differences in elastic energy of the stable equilibria and the intensities (areas) of the corresponding Lk peaks](image)

**Figure 15:** Correlation between the differences in elastic energy of the stable equilibria and the intensities (areas) of the corresponding Lk peaks

On the horizontal axis, we plot the elastic energy differences between the two equilibria, and on the vertical axis, the natural logarithm of the ratio of the areas under the two fit Gaussians $\ln(A_1/A_2)$. The correlation coefficients for these comparisons are 0.95 ($n = 38$) and 0.89 ($n = 35$) for 200 and 900 bp respectively. Thus, despite the sampling of the Monte Carlo trajectory of the entire strain energy surface, the relative peak intensities are quite well approximately by a simple expression of the relative energy levels of the two energy minima. As expected, this correlation is weaker for the 900 bp case, in which excursions from the energy minima are more frequent and distant, and yet still, the correlation is quite good even for 900 base pairs.

### 4.4.4 Sensitivity of the results to discretization and smoothing

**Influence of the Monte-Carlo segment length:** Whereas the continuum computation delivers a very precise $Lk$ value for each equilibrium, the corresponding central value computed from the $Lk$ distribution suffers from inaccuracy. First the segmentation of the chain affects the $Lk$
distribution and may shift it by a small but non-negligible amount. To evaluate this influence, we simulated ten single-peaked 200bp molecules with 4 different segment lengths, namely 5, 10, 20 and 25 bp per segment; the standard deviations of the centers of the Gaussian fits to the \( Lk \) distribution give an estimate of \( \pm 0.04 \) for the \( Lk \) position uncertainty due to segmentation. The uncertainty is about twice as large for 900bp sequences. Note that this uncertainty has been estimated in a situation of well isolated peaks of the \( P(Lk) \) distributions, and therefore significantly larger can be expected in estimating peak positions from shoulders. Energy correlations are less affected by this discretization effect since the logarithm of a ratio of areas is considered.

**Influence of the smoothing of the continuum computations.** As explained in Sec. 2.2, the continuum minimum-energy centerline is obtained by a smoothing of the discrete base-pair level wedge model. We checked on a single 900 bp sequence by how much the energy and link of its unique nicked equilibrium were affected by the choice of the smoothing window \( w \):

<table>
<thead>
<tr>
<th>( w )</th>
<th>( Lk )</th>
<th>( E/RT )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>86.42</td>
<td>0.995</td>
</tr>
<tr>
<td>50</td>
<td>86.42</td>
<td>0.980</td>
</tr>
<tr>
<td>75</td>
<td>86.37</td>
<td>1.045</td>
</tr>
<tr>
<td>100</td>
<td>86.36</td>
<td>1.093</td>
</tr>
</tbody>
</table>

Thus, for \( w = 25 \) or \( w = 50 \) as used here, the associated error in \( Lk \) appears to be on the order of 0.05, with a similarly small error if \( E \).

5 Conclusion

5.1 Multiple equilibria for DNA minicircles are no surprise

In this paper, we analysed how in the continuum rod equations, the incorporation of intrinsic curvature to capture the equilibrium DNA wedge angles breaks the symmetry of the perfect diagram. By studying this symmetry-breaking in detail, one can even formulate conditions for the likelihood of the formation of multiple DNA equilibria in terms of the simple integrals of intrinsic curvature, \( (I_1, I_2) \). This symmetry breaking naturally produces stable multiple minima for a large portion of the 200 and 900 bp random sequences studied here. We performed a statistical survey based on more than 8000 such random sequences, which helped us verify that the smallness of \( (I_1, I_2) \) is a good qualitative predictor of the probability of a given sequence to show multiple equilibria.

5.2 Monte-Carlo agrees with continuum computations

We then performed systematic Metropolis Monte-Carlo runs on about 200 randomly selected sequences, and found a good correlation between the number of equilibria predicted by the continuum computations and the number of peaks seen in the Monte-Carlo \( Lk \) distribution. Moreover, not only do the \( Lk \) values of those equilibria correspond to the positions of the fitted Gaussians peaks, but the relative areas of these peaks is related to the elastic energies of the equilibria according to the Boltzmann distribution.
Katritch and Vologodskii ([8]) studied the multiple equilibria for nicked DNA minicircles of 900 bp, because they were "looking for a conformational property that would be the most sensitive to the intrinsic curvature and is easily measurable". From our statistical analysis based on continuum computations, we can estimate that effectively their chance to see multiple peaks at 900 bp was about twice as large than for 200 bp with the same wedge angles BT1 (see paragraph 4.2). They reported that multiple peaks did occur in their Lk distributions mostly for lengths between 300 and 1000 bp, and only for $a_{\text{stat}} \leq 250 \text{nm}$. The comparison we carried over the two methods allows us to widen these thresholds, as we were able to predict and observe several instances of multiple peaks in the Lk distributions for 200 bp minicircles. Also, even for such short DNA 200 bp minicircles, we found sequences which showed two peaked Lk distributions for the BT2 angleset (corresponding to $a_{\text{stat}} \geq 400 \text{nm}$).

5.3 Future developments

A natural extension of this work will be to study the correlations between equilibrium computations and Monte Carlo simulations for cyclized equilibria (as opposed to nicked ones), for which we already have all the continuum data, but currently lack a "cyclized" Monte-Carlo algorithm (incorporating intrinsic curvature). Although the particular Monte-Carlo code of Katritch and Vologodskii has been designed to be very CPU-efficient, it still is on average at least 20 times slower than a continuum computation for a given sequence, and at least a hundred times slower than a simple computation of $(I_1, I_2)$. Therefore the selection of interesting DNA sequences on which to test theoretical predictions of multiple equilibria is most efficiently done in the continuum realm.

Also under development is an algorithm, based on the same rod equations with intrinsic curvature, for determining a solution of the Langevin equations of motion for DNA minicircles. With this algorithm, one could conceivably confirm not only the Monte-Carlo P(Lk) distributions, but also discover the typical timescales of exchange between multiple equilibria.

We thank Seva Katritch for putting the Monte-Carlo program at our disposal.

References


