DNA Rings with Multiple Energy Minima

Patrick B. Furrer[1], Robert S. Manning[2], and John H. Maddocks[3]

to appear in Biophysical Journal

Abstract

Within the context of DNA rings, we analyze the relationship between intrinsic shape and the existence of multiple stable equilibria, either nicked, or cyclized with the same link. A simple test, based on a perturbation expansion of symmetry-breaking within a continuum elastic rod model, provides good predictions of the occurrence of such multiple equilibria. The reliability of these predictions is verified by direct computation of nicked and cyclized equilibria for several thousand DNA minicircles with lengths of 200 and 900 basepairs. Further, our computations of equilibria for nicked rings predict properties of the equilibrium distribution of link, as calculated by much more computationally intensive Monte Carlo simulations.

Keywords:

Symmetry breaking, Elastic rod model, DNA minicircles, DNA intrinsic curvature, Topoisomer distribution, Metropolis Monte Carlo

Running Title:

Multiple Equilibria of DNA rings.

[1]Département de Mathématiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne Switzerland, `Patrickb.Furrer@epfl.ch`

[2]Mathematics Department, Haverford College, Haverford, PA 19041 USA, `rmanning@haverford.edu`

[3]Département de Mathématiques, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne Switzerland, `maddocks@dma.epfl.ch`

1 Introduction

Various elastic models have been used to calculate DNA configurations subject to externally imposed constraints. One of these constraints, studied experimentally by (Pulleyblank et al. 1975; Depew and Wang 1975; Shore and Baldwin 1983; Horowitz and Wang 1984), is circularity: the requirement that the two ends of the DNA close to form a ring. If both strands of the sugar-phosphate backbone are required to close, the DNA ring is referred to as *cyclized*, while if only one strand is closed, the DNA is said to be *nicked*.

Models of such cyclization experiments often assume an intrinsically straight reference state for the unconstrained double-helical DNA, as for instance in (Levene and Crothers 1986; Shimada and Yamakawa 1988; Klenin et al. 1991; Tobias 1998). Although this hypothesis may seem a fair first approximation, experimental evidence for a mild but significant intrinsic curvature has accumulated during the last twenty years (for a review, see for instance Olson and Zhurkin 1996). These observations have motivated a refinement of DNA models by the incorporation of intrinsically curved reference states, either as isolated bent sites as in (Bauer et al. 1993; Tobias and Olson 1993; Westcott et al. 1995; Yang et al. 1995; Klenin et al. 1995; Rippe et al. 1995), or in a sequence-dependent manner as in (De Santis et al. 1992; Katritch and Vologodskii 1997; Manning et al. 1996; Kahn and Crothers 1998).

DNA molecules can cyclize into several local minima of different links, forming the so-called topoisomer distribution (as reviewed for instance in Bates and Maxwell 1993; Levene 1994; Stasiak 1996). In this paper we focus on a different phenomenon that arises when intrinsic curvature of the DNA is taken into account: the existence of more than one cyclized local minimum of the *same* link, which we will refer to as the existence of *multiple cyclized minima*. In addition, the presence of intrinsic curvature can cause the DNA to have more than one nicked local minimum, a phenomenon we refer to as *multiple nicked minima*. We will describe the two phenomena collectively as the existence of *multiple minima*.

As reviewed in (Schlick 1995; Olson 1996), there are several numerical strategies for determining local energy minima within elastic DNA models, including direct minimization, solution of equilibrium equations, or some form of simulated annealing, e.g. using Monte Carlo. This paper will primarily be concerned with solution of equilibrium equations, in particular those associated with a continuum rod model (Benham 1979; LeBret 1979; Yang et al. 1993; Manning et al. 1996; Manning and Maddocks 1999), as can be derived through standard procedures of the calculus of variations. More precisely, we present an analysis of the equilibrium equations that provides a qualitative understanding of the existence of multiple minima not readily available from numerical minimization or simulated annealing[1].

In particular, we analyze the important consequences of symmetry-breaking in the equilibrium conditions in the transition from the symmetric case of an intrinsically straight, isotropic model to the nonsymmetric case incorporating intrinsic curvature. The same transition was the focus of the recent Metropolis Monte Carlo study (Katritch and Vologodskii 1997) (hereafter abbreviated as [KV]), and we compare our results directly against their findings. For cyclized DNA, there is an integer-valued link Lk, which is related to total twist Tw and writhe Wr through the well-known formula $Lk = Tw + Wr$. Both the twist and writhe remain well-

[1]Since our results are derived from the equilibrium equations, they apply not only to multiple minima, but also to multiple equilibria of other types (such as saddle points), but we will focus here on the implications for local minima.

defined for nicked minicircles, and [KV] therefore *define* the sum $Tw + Wr$ to be a generalized, usually non-integer, link Lk. In their Monte Carlo simulations, [KV] discovered that the distribution $P(Lk)$ of this non-integer link in nicked DNA is dramatically affected by the presence of intrinsic DNA curvature. They observed that while $P(Lk)$ for intrinsically straight DNA always looks Gaussian, the addition of intrinsic curvature led in certain cases to a $P(Lk)$ that could only be fit by a sum of two (or more) Gaussians. They noted that this bimodal behavior occurs especially when the intrinsic shape of the DNA is S-like. They concluded from these findings that the addition of intrinsic curvature can induce a second nicked minimum not found in the intrinsically straight case.

Of course, Metropolis Monte Carlo simulations do not directly compute equilibria, but rather sample an assumed equilibrium distribution of shapes about one or more local minima. Thus, the presence of multiple minima can only be detected indirectly by Monte Carlo methods, and then perhaps computed via the addition of simulated annealing. In contrast, we will show that a direct analysis of the equilibrium equations yields a computationally simple test for the presence of multiple minima. This test is in good agreement with the Monte Carlo findings of [KV], but is computationally much less intensive. Specifically, we show that the existence of multiple nicked minima arises from the perturbation of the *family* of degenerate equilibria that exists for the symmetric case of an intrinsically straight rod. The existence of multiple minima can be predicted via the evaluation of certain simple integrals of intrinsic shape parameters that are computable from the basepair sequence in a few seconds. Our predictions are in agreement with the distinction between C and S shapes proposed by [KV], but they also provide a quantitative, more refined, selection criterion in other cases where the presence of multiple minima is less intuitive. The new criterion accordingly offers a valuable tool in guiding the selection of interesting sequences for further investigation.

In Sec. 2, we describe the continuum rod equilibrium equations, along with the perturbation expansion developed in (Manning and Maddocks 1999) that is at the heart of our classification of multiple minima. In Sec. 3, we describe both the equilibrium and Monte Carlo computations used to illustrate our results. In Secs. 4.1–4.3, we present a study of circular DNA minima (both nicked and cyclized) for several thousand different DNA sequences of lengths 200 and 900 basepairs (bp) that verifies the efficiency of the perturbation expansion as an indicator of multiple minima. Finally, in Sec. 4.4, we investigate in detail the connections between the *equilibria* computed within our static elastic model and the *equilibrium distribution* simulated by the [KV] Metropolis Monte Carlo. We find strong correlations between the number, locations, and areas of peaks in the distribution $P(Lk)$ with the number, links, and energies of the equilibria. Given the complexity of the strain-energy surface on which the Monte Carlo simulation wanders, it is perhaps surprising that a relatively simple and fast equilibrium computation can yield such good predictions of equilibrium distributions. This observation thus offers hope that equilibrium computations can be used in the future as valuable pre-computations to guide selection of interesting sequences for Monte Carlo, or indeed, experiment.

2 Theory

Section 2.1 presents the basic assumptions of rod mechanics, and Sec. 2.2 describes our procedure for incorporating DNA parameters into this continuum 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 intrinsic curvature is added to the model. Finally, in Sec. 2.5, we discuss what this infinitesimal result implies for the typical curvatures appearing in real DNA.

2.1 Elastic rod equilibria

We begin by summarizing the formulation presented in (Dichmann et al. 1996) of the special Cosserat theory (see, e.g. (Antman 1995)), commonly used in continuum mechanics to model an inextensible and unshearable elastic 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 assumptions 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).

Rod equilibria are, by definition, critical or stationary points of a strain energy E over a specified space of rod configurations, generally described via a set of constraints. First we describe the particular form that we will assume for E, and then present the constraints of interest for application to DNA rings.

The strain energy E is defined in terms of the *strains* $u_i(s)$:

$$\begin{aligned} u_1(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) &\equiv -\mathbf{d}_2^T \mathbf{d}_3', \\ u_2(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) &\equiv \mathbf{d}_1^T \mathbf{d}_3', \\ u_3(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) &\equiv \mathbf{d}_2^T \mathbf{d}_1'. \end{aligned} \tag{1}$$

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

$$E = \int_0^1 W(u_1(s), u_2(s), u_3(s), s) ds. \tag{2}$$

In this paper, we further assume the particular form

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

although the procedures we describe could similarly be applied to other more general strain-energy functions. In all computed examples, we take $K_3 = 0.8 K_1$, a value near the bottom of the typically reported range for the ratio of twist to bending stiffnesses $0.7 \leq K_3/K_1 \leq 2.0$ (Schlick 1995). For this choice, we have previously found good agreement between rod model energies and cyclization J factors (Manning, Maddocks, and Kahn 1996). However, all our analytic considerations (such as formulas 9) hold for arbitrary values of $K_3 > 0$, and the numerical computations could readily be repeated for any other value.

We first consider the constraints for a ring, as shown in Fig. 1, which require that the rod centerline be closed, the tangent vectors match at the closure point, and a twist angle α be imposed:

$$\begin{aligned} \mathbf{r}(1) &= \mathbf{r}(0), \\ \mathbf{d}_3(1) &= \mathbf{d}_3(0), \\ \mathbf{d}_1(1) &= \cos\alpha\ \mathbf{d}_1(0) + \sin\alpha\ \mathbf{d}_2(0), \\ \mathbf{d}_2(1) &= -\sin\alpha\ \mathbf{d}_1(0) + \cos\alpha\ \mathbf{d}_2(0). \end{aligned} \tag{3}$$

If α is a multiple of 2π, then these conditions model cyclization. In contrast, and following [KV], our physical model of a nicked equilibrium is that the DNA double-helix on either side of the nick site has a common tangent direction, but is free to rotate about it. Thus for a nick site at $s = 1$ we impose only the constraints that require the centerline to be closed and the tangent vectors to match at the closure point:

$$\begin{aligned} \mathbf{r}(1) &= \mathbf{r}(0), \\ \mathbf{d}_3(1) &= \mathbf{d}_3(0). \end{aligned} \tag{4}$$

In the case of cyclization, the ribbon is closed, and the two edges of the ribbon are closed curves. We can thus define an integer linking number (or simply *link*) *Lk* of the two edges, which is a topological invariant of the ribbon. The Călugăreanu-White-Fuller[2] theorem (Călugăreanu

[2]See (Moffatt and Ricca 1992) for a discussion of the history of this theorem.

1959; Călugăreanu 1961; White 1969; Fuller 1971) states that the integer Lk can be written as

$$Lk = Tw + Wr, \tag{5}$$

where the twist Tw and writhe Wr are defined as

$$\begin{aligned} Tw &\equiv \frac{1}{2\pi}\int_0^1 u_3(s)ds, \\ Wr &\equiv \frac{1}{4\pi}\int_0^1\int_0^1 \frac{(\mathbf{r}(s)-\mathbf{r}(\sigma))^T(\mathbf{r}'(s)\times\mathbf{r}'(\sigma))}{|\mathbf{r}(s)-\mathbf{r}(\sigma)|^3}\, ds\, d\sigma. \end{aligned}$$

Because Tw and Wr are well-defined for non-closed ribbons, it is reasonable to generalize the notion of Lk to a real-valued quantity *defined* by Eq. 5. This generalization is consistent with the Monte Carlo study in [KV], to which we will make comparisons. We will also use the fact, reported by [KV], that the fractional parts of Lk and of $\frac{\alpha}{2\pi}$ are equal. See (Hoffman, Manning, and Maddocks, in preparation) for a proof of this result and further discussion.

By standard techniques in the calculus of variations, critical points of E are found by solving a system of first-order ordinary differential equations (ODEs); see Sec. 3.3 for a precise description of these equations. Throughout this article, a *ring equilibrium* denotes a rod configuration satisfying these ODEs subject to the twisted ring boundary conditions in Eq. 3. A cyclized equilibrium is a ring equilibrium with α a multiple of 2π, i.e. Lk an integer. For nicked molecules the absence of the twist angle α in the imposed conditions in Eq. 4 leads to a "natural boundary condition" (Gelfand and Fomin 1963, p. 26) that must be satisfied by any equilibrium. The specific form of this natural boundary condition is $m_3(1) = 0$, where $m_3(s) = K_3\, u_3(s)$ is the twist moment about the tangent vector $\mathbf{d}_3(s)$. Thus, a nicked equilibrium is any ring equilibrium that also satisfies $m_3(1) = 0$.

Since the approach taken here is to solve equilibrium equations rather than numerically minimize E, the ring equilibria that we find may be (*unstable*) saddle points of the strain energy in addition to the (*stable*) local minima of primary interest. Minima and saddle points can be distinguished by the computation of a certain *instability index* that gives the number of negative eigenvalues of an appropriate second derivative (roughly, the number of independent downward directions of the energy surface); in particular, local minima have index zero, while equilibria with positive index can be anticipated to be unstable. Index can be determined by a straightforward computation based on the conjugate point test described in (Manning et al. 1998).

The solutions of primary interest in this paper are stable cyclized equilibria and stable nicked equilibria, which we refer to as *cyclized minima* and *nicked minima.* We have seen above that a nicked equilibrium is a ring equilibrium for some value of α at which $m_3(1) = 0$. It is furthermore easy to show that a nicked *minimum* must satisfy two additional conditions. It must be a stable ring equilibrium at fixed link, i.e. a minimum among nearby configurations with the same value of α, and its energy must also be a local minimum among ring equilibria at nearby Lk (since if E were not a local minimum, the rod could lower its energy by rotating $\mathbf{d}_1(1)$ about $\mathbf{d}_3(1)$, and hence would not be at a nicked minimum).

2.2 Applying the continuum rod model to DNA

We summarize the procedure from (Manning et al. 1996) for determining values of the continuum parameters K_i and $\hat{u}_i(s)$ appropriate for our DNA sequences. For the stiffnesses K_i, we take

$$K_1 = K_2 = \frac{PRT}{N\ell}, \quad K_3 = 0.8K_1,$$

where T is the temperature, $R = 8.314$ J/mol-Kelvin, $P \approx 500$ Angstroms is the persistence length, $\ell \approx 3.4$ Angstroms is the helix-rise-per-basepair, and N is the number of basepairs (which appears here due to the fact that the continuum rod has non-dimensionalized length one). We note that at the basepair level, DNA almost certainly has a preferential direction for bending, i.e. $K_1 \neq K_2$, 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 $K_1 = K_2$); cf. (Kehrbaum 1997).

The centerline $\hat{\mathbf{r}}$ of the continuum intrinsic shape is derived by smoothing a piecewise linear centerline constructed from a basepair-level wedge-angle model (see e.g., Bolshoy et al. 1991). First, a tapered averaging filter of width w basepairs is applied to the wedge-angle centerline, and then the filtered centerline is fit via least squares to a piecewise polynomial with continuous third derivative throughout the rod[3]. 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

[3]See (Manning et al. 1996) for further details on the smoothing algorithm, including the padding of the centerline to minimize end effects of the averaging filter and the use of a double filtration to reduce the tendency of an averaging filter to straighten out intrinsic curvature

basepairs N. When $N = 200$, we choose $w = 20$, the value used in (Manning et al. 1996) for a study of 150–160 bp DNA. When $N = 900$, we choose $w = 50$, since in that case, we are interested in curvature on the order of several tens of basepairs.

Having determined the intrinsic centerline $\hat{\mathbf{r}}$, the director $\hat{\mathbf{d}}_3$ is found via the inextensibility-unshearability assumption $\hat{\mathbf{d}}_3 = \hat{\mathbf{r}}'$. It is then straightforward to generate a continuum $(\hat{\mathbf{d}}_1, \hat{\mathbf{d}}_2)$ using the intrinsic twist of the wedge-angle model. Physically, $\hat{\mathbf{d}}_1(s)$ will track the orientation of the major groove in the DNA intrinsic shape. Similarly, $\mathbf{d}_1(s)$ tracks the major groove in any *deformed* shape. Thus, we will call $(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)$ the *DNA-framing* of the rod.

Unfortunately, the high twist in the DNA frame induces rapid variations in $\hat{u}_1(s)$ and $\hat{u}_2(s)$, hindering the coarse (but highly accurate) numerical rod discretizations that are one goal of the continuum model. This difficulty can be overcome by re-expressing the equilibrium equations in terms of a new *natural framing* of the rod that we denote by $(\mathbf{D}_1(s), \mathbf{D}_2(s), \mathbf{D}_3(s) = \mathbf{d}_3(s))$. A crucial fact is that given an equilibrium shape $(\mathbf{D}_1(s), \mathbf{D}_2(s), \mathbf{d}_3(s))$ for the naturally-framed rod, an equilibrium shape $(\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s))$ of the DNA-framed rod is recovered by rotating $(\mathbf{D}_1(s), \mathbf{D}_2(s))$ about $\mathbf{d}_3(s)$ by an angle $\Omega(s)$ (Manning and Maddocks 1999), as shown in Fig. 2. Both the DNA and natural framings depend on the configuration of the rod, but the angle Ω between the two framings is independent of the configuration; the angle Ω is defined once and for all by the following procedure. Given the DNA frame $(\hat{\mathbf{d}}_1(s), \hat{\mathbf{d}}_2(s), \hat{\mathbf{d}}_3(s))$ of the intrinsic shape, define the intrinsic natural frame $(\hat{\mathbf{D}}_1(s), \hat{\mathbf{D}}_2(s))$ by the conditions[4]:

$$\begin{aligned} \hat{\mathbf{D}}_1(0) &= \hat{\mathbf{d}}_1(0), \\ u_3(\hat{\mathbf{D}}_1(s), \hat{\mathbf{D}}_2(s), \hat{\mathbf{d}}_3(s)) &= 0, \ \ \forall s \in (0,1). \end{aligned}$$

Then let $\Omega(s)$ denote the angle between $\hat{\mathbf{d}}_1(s)$ and $\hat{\mathbf{D}}_1(s)$[5]. We emphasize that no approximation is involved in the use of the natural framing for computations. The equilibria of the naturally-framed rod with the twist function $\Omega(s)$ added back in are exactly the same as the equilibria of the DNA-framed rod; the natural frame is merely a computational convenience that makes the coefficients of the equilibrium equations much more slowly varying. Similarly, there is no assumption about local twist variation implied by the use of the natural frame. Any basepair-to-basepair twist variation is encapsulated in the function $\Omega(s)$ and thus fully accounted for by the continuum computations.

Setting $s = 1$ in the definition of $\Omega(s)$, the twist angle α for a DNA-framed ring is related to the twist angle α_n for the corresponding naturally-framed ring by $\alpha = \alpha_n + \Omega(1)$. Further, by Eq. 5, the link Lk of any DNA-framed configuration is related to the link Lk_n of the corresponding naturally-framed configuration by

$$Lk = Lk_n + \widehat{Tw}, \tag{6}$$

[4]Note that $u_3(\mathbf{D}_1(s), \mathbf{D}_2(s), \mathbf{D}_3(s))$ is not in general zero for deformed shapes, so that the natural framing is only known to have zero twist for the intrinsic shape.

[5]It is important to ensure that $\Omega(s)$ is an increasing continuous function, by choosing $\Omega(s) \in [2n\pi, 2(n+1)\pi)$ if $\hat{d}_1(\sigma)$ has undergone n full rotations with respect to $\hat{D}_1(\sigma)$ for $0 \leq \sigma \leq s$.

where $\widehat{Tw}$ is the twist of the intrinsic DNA frame

$$\widehat{Tw} = \frac{1}{2\pi}\int_0^1 u_3(\hat{\mathbf{d}}_1(s), \hat{\mathbf{d}}_2(s), \hat{\mathbf{d}}_3(s))ds = \frac{\Omega(1)}{2\pi}. \tag{7}$$

Some discussion of notation is in order. $\widehat{Tw}$ is the appropriate continuum version of the quantities "relaxed twist" or "relaxed link", denoted Tw_0 or Lk_0 in the literature (Bates and Maxwell 1993). There seems to be no standard definition for Tw_0 or Lk_0, however; e.g., they have been used to denote the center of a Gaussian fit to the topoisomer distribution, or the ratio of number of basepairs to the average number of basepairs per full turn. The absence of an exact definition for Tw_0 or Lk_0 may be no accident; this article demonstrates that for some sequences there are multiple nicked minima so that there may not be a unique "relaxed" ring.

In the literature, one often uses the symbol ΔLk to denote $Lk - Lk_0$, i.e. the difference of a given link from its relaxed value. Thus $Lk_n = Lk - \widehat{Tw}$ is the exact analogue of this concept for the continuum rod, so we shall henceforth write ΔLk instead of Lk_n, so that Eq. 6 can be rewritten as

$$Lk = \Delta Lk + \widehat{Tw}. \tag{8}$$

Most of our results in Sec. 4 refer to the true DNA link Lk, but elements of our computations and theory are expressed in terms of the natural frame link ΔLk. We caution that our ΔLk is not generally an integer for cyclized configurations, as this notation sometimes implies. Instead, we will call a cyclized configuration a 0-*topoisomer* if its ΔLk is such that Lk is the nearest integer to $\widehat{Tw}$, a +1-*topoisomer* if its link is one more than the nearest integer to $\widehat{Tw}$, etc.

2.3 The perfect diagram

The principal goal of this article is to understand the multiplicities of ring equilibria in the transition from an idealized symmetric case which we call the *perfect problem* ($\hat{u}_i \equiv 0$) to various *imperfect problems* ($\hat{u}_1, \hat{u}_2 \not\equiv 0$). The reason for studying the perfect problem is that its solution set acts as an organizing center for the solution set of more realistic problems with nonzero intrinsic curvature. Physically, the perfect problem involves a uniform isotropic rod that is intrinsically straight, while the imperfections considered here involve the introduction of intrinsic curvature. Since all computations involve the naturally-framed rod, we need only consider zero intrinsic twist $\hat{u}_3 \equiv 0$. Arbitrary intrinsic twist parameters, whether constant or sequence-dependent, are accounted for via the angle $\Omega(s)$ relating the natural and DNA framings of the rod, as described in Sec. 2.2.

The perfect problem has been studied by several authors; see, e.g., (Schlick 1995; Olson 1996; Kehrbaum and Maddocks 1997) for a discussion of the literature. Here we follow the notation and formulation that are described in detail in (Dichmann et al. 1996).

2.3.1 The set of perfect equilibria

Fig. 3 shows a portion of the set of ring equilibria for the perfect problem, with the energy E plotted as a function of the link ΔLk. There are many branches of higher-energy equilibria not shown, but they are not of concern here. This *perfect bifurcation diagram* is described in detail in (Li and Maddocks 1999). 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 a planar figure-eight embedded at D. Note that the link jumps by two at points of self-intersection such as D, i.e. the whole segment between the two points marked D actually represents the same figure-eight solution. Solid lines denote stable equilibria (local minima), while dashed lines denote unstable equilibria.

2.3.2 The register symmetry

Due to material symmetries of the perfect rod, every point on the perfect diagram is not a single equilibrium, but instead represents an entire manifold of ring equilibria, all with the same energy. The equilibria on this manifold are related by the action of a symmetry transformation that we will call rotational register, or simply *register*, following (Sanghani et al. 1996). This register symmetry involves keeping the centerline fixed, but spinning the framing $(\mathbf{d}_1(s), \mathbf{d}_2(s))$ by a constant angle θ about the tangent $\mathbf{d}_3(s)$. In the language of DNA, the register determines whether the major groove at a particular basepair of a DNA ring faces toward the center of the ring, away from the center, or somewhere in between. See (Manning and Maddocks 1999) for further discussion of this symmetry. Using this register transformation, we find that every point on the trivial branch ABC actually represents a circle of ring equilibria[6]. Our focus will be to determine the fate of this circle of degenerate solutions when intrinsic curvature is added.

[6]Similarly, the register transformation plus a second transformation based on translation in arclength imply that each point on the nontrivial branch of Fig. 3 represents a *torus* of ring equilibria (Manning and Maddocks 1999). However, since for $K_3/K_1 = 0.8$ all solutions on this nontrivial branch are 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).

2.4 Perturbation of the trivial branch

As described in the previous section, when $\hat{u}_i \equiv 0$, the rod equilibrium conditions at any fixed ΔLk have a circle of twisted-circle solutions, parametrized by the register θ. For $\hat{u}_i \not\equiv 0$, this circle of solutions typically breaks up into only a finite set of solutions. For infinitesimal intrinsic curvatures, i.e. intrinsic curvatures of the form $\epsilon \hat{u}_i$ with ϵ sufficiently small, we can draw some general conclusions about the symmetry-breaking from the perturbation expansion described in (Manning and Maddocks 1999). Of course, we are ultimately interested in the case $\epsilon = 1$ (see Sec. 2.5).

It has been proven (Manning and Maddocks 1999) that for infinitesimal intrinsic curvatures, whenever one of the following two integrals:

$$\begin{aligned} I_1(\Delta Lk) &= \int_0^1 \left[\hat{u}_1(s)\sin(2\pi\ \Delta Lk\ s) + \hat{u}_2(s)\cos(2\pi\ \Delta Lk\ s)\right]\ ds, \\ I_2(\Delta Lk) &= \int_0^1 \left[\hat{u}_1(s)\cos(2\pi\ \Delta Lk\ s) - \hat{u}_2(s)\sin(2\pi\ \Delta Lk\ s)\right]\ ds, \end{aligned} \tag{9}$$

is nonzero, then the circle of trivial solutions at link ΔLk will yield exactly two solutions; see Sec. 4.3 for one geometric interpretation of the integrals I_1, I_2.

Of these two equilibria, only one is expected to be stable, as shown in Fig. 4. Proving this assertion about stability involves computations beyond the perturbation expansion presented here, but can be made plausible by a finite-dimensional example. Consider a surface of revolution as shown in Fig. 5 containing a circular "valley" of local minima (analogous to the circle 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 circle of critical points. When we add a small perturbation to the surface of revolution as in Fig. 6 (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, one local minimum and one saddle point. Of course, we can engineer infinitesimal 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 $I_1(\Delta Lk) = I_2(\Delta Lk) = 0$.

2.5 Implications of perturbation results for curved DNA

The perturbation expansion from Sec. 2.4 predicts that for infinitesimal curvatures $\epsilon\hat{u}_i$, if there are to be two or more stable ring equilibria of link ΔLk, we must have $I_1(\Delta Lk) = I_2(\Delta Lk) = 0$. By the nature of perturbation expansions, this result is only directly applicable to sufficiently small ϵ. However, for small-to-moderate ϵ, the perturbation results should still correlate with the observed behavior. Thus, for realistic DNA curvatures we can conjecture that the existence of multiple stable ring equilibria is well correlated with $|I_1|, |I_2|$ being *small*. We verify this conjecture in Secs. 4.1–4.2.

3 Methods

3.1 Random sequences

Random DNA sequences, 5,017 of 200 bp and 2176 of 900 bp, were generated using the random number generator `ran1` (Press et al. 1992), with equal probability of A, C, G, or T at each base. Data collection was stopped at the above sample sizes as sufficient numbers of examples exhibiting multiple minima occurred to give qualitatively stable statistics.

3.2 Intrinsic shape parameters

For each sequence, a preliminary piecewise-linear intrinsic shape was determined using a standard wedge-angle model. We confined our study to dinucleotide wedge angles based on those presented in (Bolshoy et al. 1991), but our computations could easily be repeated for any other set of wedge angles, or other basepair level models for intrinsic shape. Our first angle set (AS1) uses the intrinsic tilts and rolls derived in (Bolshoy et al. 1991), while our second angle set (AS2) scales these tilts and rolls by 0.61, thereby producing less intrinsic curvature (and increasing the static persistence length (Trifonov et al. 1987) by $1/(0.61)^2$ from 168 to 467 nm). This scaling, inspired by (Kahn and Crothers 1992), is used here to demonstrate the effects of smaller intrinsic curvature on the quality of predictions based on the perturbation expansion. In contrast to the sequence-dependent intrinsic twists found in (Bolshoy et al. 1991), our angle sets AS1 and AS2 both use a common intrinsic twist of 34.45° per basepair for all dinucleotide steps. This change is purely a matter of convenience for the statistical survey of many molecules in Sec. 4.2, so that the total intrinsic twist $\widehat{Tw}$ will be approximately constant across the molecules in the sample. However, the general theory presented here allows basepair-dependent intrinsic twists if desired. Each preliminary wedge-angle intrinsic shape was smoothed via the procedure summarized in Sec. 2.2 to yield continuum intrinsic shape parameters $\hat{u}_i(s)$ and the auxiliary function $\Omega(s)$.

3.3 Equilibrium equations

For completeness we here present the specific equations we solved numerically in our determination of ring equilibria, namely the differential equations (10) subject to the ring boundary conditions (12). However a detailed understanding of the particular form of these equations is not necessary to understand the results presented later.

We chose to parametrize the space $SO(3)$ of directors $(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)$ using *Euler parameters* (or *quaternions*) $\mathbf{q} \in \mathbb{R}^4$:

$$\begin{bmatrix} \mathbf{d}_1(\mathbf{q}) & \mathbf{d}_2(\mathbf{q}) & \mathbf{d}_3(\mathbf{q}) \end{bmatrix} = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2q_1q_2 - 2q_3q_4 & 2q_1q_3 + 2q_2q_4 \\ 2q_1q_2 + 2q_3q_4 & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2q_2q_3 - 2q_1q_4 \\ 2q_1q_3 - 2q_2q_4 & 2q_2q_3 + 2q_1q_4 & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix}.$$

In this case, the equilibrium equations can be cast in the Hamiltonian form (see, e.g. Dichmann et al. 1996):

$$\begin{aligned} \mathbf{r}' &= \frac{\partial H}{\partial \mathbf{n}} = \mathbf{d}_3(\mathbf{q}), \\ \mathbf{q}' &= \frac{\partial H}{\partial \boldsymbol{\mu}} = \frac{1}{2}\sum_{i=1}^{3} u_i(\boldsymbol{\mu}, \mathbf{q})\mathbf{B}_i\mathbf{q}, \\ \mathbf{n}' &= -\frac{\partial H}{\partial \mathbf{r}} = \mathbf{0}, \\ \boldsymbol{\mu}' &= -\frac{\partial H}{\partial \mathbf{q}} = \frac{1}{2}\sum_{i=1}^{3} u_i(\boldsymbol{\mu}, \mathbf{q})\mathbf{B}_i\boldsymbol{\mu} - \frac{\partial \mathbf{d}_3}{\partial \mathbf{q}}^T \mathbf{n}, \end{aligned} \tag{10}$$

where:

$$u_i(\boldsymbol{\mu}, \mathbf{q}) = \epsilon \hat{u}_i + \frac{\boldsymbol{\mu}^T \mathbf{B}_i \mathbf{q}}{2K_i}, \tag{11}$$

$$\mathbf{B}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \ \mathbf{B}_2 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \ \mathbf{B}_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix},$$

$$H = \sum_{i=1}^{3} \left[\frac{(\boldsymbol{\mu}^T \mathbf{B}_i \mathbf{q})^2}{8K_i} + \frac{\hat{u}_i(\boldsymbol{\mu}^T \mathbf{B}_i \mathbf{q})}{2} \right] + \mathbf{n}^T \mathbf{d}_3(\mathbf{q}),$$

and the variables $\mathbf{n} \in \mathbb{R}^3$ and $\boldsymbol{\mu} \in \mathbb{R}^4$ are the conjugate variables to $\mathbf{r}$ and $\mathbf{q}$. The ring boundary conditions from Eq. 3 take the form:

$$\begin{aligned} &\mathbf{r}(0) = \langle 0, 0, 0 \rangle, \ \ \mathbf{r}(1) = \langle 0, 0, 0 \rangle, \ \ q_1(0) = q_2(0) = q_3(0) = 0, \\ &\mathbf{q}(1) = \langle 0, 0, -\sin(\alpha/2), -\cos(\alpha/2) \rangle, \ \ \mu_4(0) = 0. \end{aligned} \tag{12}$$

3.4 Equilibrium computations

Here we describe our procedure for solving the ring equilibrium equations from Sec. 3.3 using the software package AUTO, and for locating nicked and cyclized minima from among these ring equilibria.

The boundary value problem (BVP) defined by Eqs. 10,12 has two parameters: the twist angle α is the main physical parameter of interest, and the parameter ϵ in Eq. 11 is the symmetry-breaking parameter discussed in Secs. 2.4–2.5. This BVP was solved via parameter continuation using the software package AUTO (Doedel et al. 1991). We choose a closed-form solution $\mathbf{z}^0(s) \equiv (\mathbf{r}^0(s), \mathbf{q}^0(s), \mathbf{n}^0(s), \boldsymbol{\mu}^0(s))$ on the stable trivial branch of the perfect ($\epsilon = 0$) problem (see Fig. 3), say at link ΔLk^* (with corresponding angle α_n^*). We then derive an approximate solution $\mathbf{z}^0(s) + \epsilon\mathbf{z}^1(s)$ to the BVP for $\epsilon = 0.01$ using the perturbation expansion described in (Manning and Maddocks 1999). This function $\mathbf{z}^0(s) + \epsilon\mathbf{z}^1(s)$ serves as the initial approximate solution for the numerical parameter continuation computations carried out in AUTO.

Using AUTO, we first compute solutions to the BVP through continuation in ϵ from 0.01 to 1, holding $\alpha_n = \alpha_n^*$ fixed. Then, holding $\epsilon = 1$ fixed, we switch to continuation in α_n to compute a branch in the imperfect diagram. In this first step of α-continuation, α_n 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 instability 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 start of the α-continuation branch, but recommences with α_n decreasing, with the same stopping conditions (a–c). In this way, we can be confident of finding all stable BVP solutions lying on the same component as the $\alpha_n = \alpha_n^*$, $\epsilon = 1$ solution.

In about 90% of all the molecules considered, the above procedure appears to compute all stable BVP solutions. However, there are exceptional intrinsic shapes for which stable solutions exist on two or more distinct components in the imperfect diagram, and the above procedure will only compute one of these components. To remedy this difficulty, we repeat the above procedure for several values of ΔLk^*. In each case, at the end of the ϵ-continuation, we check if the solution already exists on a previously computed branch of the imperfect diagram, and if not, we compute a new branch using α-continuation, and append it to the previous one. By the nature of parameter continuation computations, one is never guaranteed to have found all BVP solutions, but by this multiple-starting-point procedure, we maximize our chances of locating all stable solutions of the BVP problem.

For each solution on the imperfect diagram, various integrals of interest such as writhe, twist, and energy are computed numerically. In addition, the instability index is determined by solving a 54-dimensional initial value problem of ODEs (cf. Manning et al. 1998).

Finally, the particular solutions in the imperfect diagram that correspond to nicked and cyclized minima are located numerically. Cyclized minima (where Lk is an integer) are easily detected by searching the stable branches in the diagram for points where $\alpha = \alpha_n + \Omega(1)$ crosses a multiple of 2π. Nicked minima (which correspond to local minima of E along a stable branch of the diagram) are slightly harder 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 the AUTO computation. To avoid such problem points, we look for quintuplets of adjacent solutions so that (a) all five solutions are stable, (b) the changes in energy along this quintuplet are down-down-up-up, and (c) the twist moment $m_3(1)$ (computed from the unknowns $\mathbf{z}$) crosses zero between the first and fifth points (see Sec. 2.1).

For a given sequence, the computation of the intrinsic shape and the associated integrals I_1, I_2 requires less than a minute on a single SparcII CPU. The time required to compute a bifurcation diagram varies with the complexity of the diagram. In roughly half of the cases, 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 strongly dependent on the number of basepairs, as opposed to the Monte Carlo simulations described below. Once the bifurcation diagram has been computed, it takes a few seconds to extract the nicked and cyclized minima.

3.5 Monte Carlo simulations

To provide a basis for comparison, we used the Metropolis Monte Carlo procedure described in [KV] without modification of their source code. See (Katritch and Vologodskii 1997) for details of this procedure. Their program is designed to simulate the thermodynamic equilibrium distribution for nicked conformations of intrinsically curved DNA, and as a result we have confined our comparison with equilibrium computations to the nicked case.

Intrinsic curvature parameters were determined using the dinucleotide angle sets AS1 and AS2 (see Sec. 3.2). The Kuhn statistical length was set to 100 nm, corresponding to a bending rigidity constant $A = 2.0 \times 10^{-19}$ erg-cm, and the torsional rigidity constant was set to $C = 1.6 \times 10^{-19}$ erg-cm in order to allow direct comparison to our equilibrium computations at $K_3/K_1 = C/A = 0.8$. The DNA effective diameter was set to 2 nm, although self-contact does not appear to play a significant role in the effects investigated here (since we are looking at relatively relaxed, nicked conformations of short DNA rings). We chose linear segments of 10 bp each, so that there were 20 total segments for 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. The simulation was sampled every 100th move, the writhe and twist of the new conformation were calculated, and from them the linking number Lk was computed using the definition $Lk = Tw + Wr$.

Each simulation involved 50×10^6 steps, a number apparently sufficient for the peak positions and areas of $P(Lk)$ to equilibrate. A single 20-segment (200 bp) simulation on a single Sparc processor requires about 200 minutes, and a single 90-segment simulation about 70 hours (in this Monte-Carlo implementation, the CPU time scales as the square of the number of segments).

We used the Matlab (The MathWorks, Inc., Natick, MA) `leastsq` function to fit $P(Lk)$ with one or more Gaussians, whose centers and areas (measured by integration) were computed for comparison with continuum results, and the function `corrcoef` to evaluate correlation coefficients reported in Sections 4.4.2 and 4.4.3.

4 Results

4.1 Bifurcation diagrams exhibiting multiple minima

The prevailing belief is that only one nicked minimum will arise for any DNA minicircle. However, as emphasized by [KV], this belief is biased by the fact that most early studies treated the perfect problem, i.e. the case of an intrinsically straight rod. In fact, the story is more complicated for rods with intrinsic curvature.

4.1.1 The origin of multiple minima

For an intrinsically straight rod, the bifurcation diagram for DNA equilibria is simply the perfect diagram in Fig. 3 shifted horizontally by the constant $\widehat{Tw}$ (which converts the natural-frame link ΔLk to the DNA-frame link Lk). Recall from Sec. 2.3.2 that each point on this diagram represents an infinite family of symmetry-related equilibria with the same energy, the same centerline, but different register. Thus, in the perfect problem, and modulo this register symmetry, there is indeed exactly one nicked minimum (the image after horizontal shift of point B in Fig. 3). Similarly, there is a single cyclized minimum for each integer Lk that falls between the links of points A and C (after shift by $\widehat{Tw}$), and no cyclized minima of other links (in the absence of self-contact).

However, when we add intrinsic curvature, which breaks the register symmetry, the resulting imperfect diagram need not look as simple as the shifted perfect diagram, as shown in Figs. 7–9. The three molecules used in these figures were chosen to illustrate most cleanly the way in which multiple minima can manifest themselves in bifurcation diagrams. Many of the diagrams we computed are more complicated than these Figures, but the general pattern of introduction of multiple minima via branch kinking is persistent across the range of computed examples.

In Fig. 7, the diagram retains the qualitative shape of the loop ABCD from the perfect diagram, although there are some clear changes introduced by the intrinsic curvature (e.g., energies are shifted downward, the range of links covered by the stable solutions is slightly reduced). Indeed, there is once again a single nicked minimum (marked with a black triangle), and at most one cyclized minimum at each integer link (marked with gray circles). Quite naturally, the diagram for the smaller angle set AS2 resembles the perfect diagram more closely than the diagram for AS1.

On the other hand, in Fig. 8, the intrinsic curvature induces a kink near the bottom of the diagram, with the result that for both angle sets AS1 and AS2, there are two nicked minima. The smaller angle set AS2 straightens out this kink (pulling it closer to the perfect diagram) so that the second nicked minimum is on the verge of disappearing.

Similarly, in Fig. 9, we see an induced kink in the diagram in the neighborhood of link 18, with the result that for AS1, two link-18 cyclized minima exist. Here, the kink is removed for AS2, but in other cases, it can remain.

4.1.2 Experimentally observable multiple minima

To our knowledge, no experimental evidence for multiple minima (either nicked or cyclized) has been reported. Might this lack of evidence be due to the fact that the differences among these multiple minima are too small to be observed experimentally? For instance, in gel migration experiments, where the mobility of DNA minicircles has been observed to correlate strongly with writhe, it seems reasonable to conjecture that multiple equilibria would generate distinct bands only if their differences in writhe are sufficiently large.

Motivated by this question, we pursued a statistical study of the differences in writhe between multiple equilibria. In particular, we selected from our 200 bp database all molecules having either two nicked minima or two cyclized minima of the same link (using the case of angle set AS1). For each such molecule, we computed the difference in writhes for its pair(s) of multiple minima. We fitted the distributions of these differences to a Gaussian and found the best fits to have (approximately) mean zero, with standard deviation $\sigma = 0.018$ for nicked, and $\sigma = 0.017$

for cyclized. Presumably these differences in writhe would be too small for the multiple minima to be distinguished on a gel.

However, when we repeated this procedure on our 900 bp database, we found significantly larger writhe differences; for pairs of nicked minima $\sigma = 0.152$, and for pairs of cyclized minima $\sigma = 0.138$. In particular, we found several examples of pairs of minima in 900 bp molecules with substantial differences in writhe (see Fig. 10). As a further assessment of the possible differences in writhe, we remark that among the five 900 bp molecules in Fig. 16, the writhes of the thirteen nicked minima lie between -0.26 and 0.18; in particular some centerlines, especially for the extreme values of Lk, are quite nonplanar.

It may be interesting to seek to detect such differences in writhes in gel mobility experiments. However due to the long duration of the gel migration experiment (several minutes) it is not clear that two distinct writhes will be observed at all, but rather a single band corresponding to an average mobility might arise. Other experimental techniques could perhaps overcome this limitation: for instance, electron cryomicroscopy, with its flash-freezing protocol, may allow one to immobilize configurations close to the different minima (Dubochet et al. 1992), and therefore such experimental methods offer the possibility of revealing the existence of the multiple minima predicted here.

4.1.3 An efficient predictor of multiple minima

Larger intrinsic curvatures create imperfect diagrams further from the perfect diagram, which can therefore more easily exhibit multiple minima. Such curvature enhancement arises, for example, if one assumes larger angle sets for the same sequence, as exhibited by Figures 7–9.

Alternatively, larger curvatures arise when treating longer DNA for a fixed angle set, since the non-dimensionalized local curvatures for the (length-one) continuum rod are larger. These effects are illustrated by the statistical study presented in Sec. 4.2.

For a given molecular length and angle set, can we predict which sequences are more likely to exhibit multiple minima? The only way to know for certain if a given molecule has multiple minima is to directly compute the set of equilibria as in Figs. 7–9. However, armed with the perturbation analysis of Sec. 2.4, we have a simple predictor for the existence of multiple minima, namely the smallness of the integrals $|I_1(\Delta Lk)|$ and $|I_2(\Delta Lk)|$. If multiple nicked minima are sought, then one should consider intrinsic shapes with $|I_1(\Delta Lk)|$ and $|I_2(\Delta Lk)|$ small at $\Delta Lk = 0$, since the nicked minimum of the perfect problem occurs at $\Delta Lk = 0$. For example, $I_1(0)^2 + I_2(0)^2$ is smaller in Fig. 8a than in Fig. 7a (0.7 versus 1.7). On the other hand, by Eq. 8, multiple cyclized minima of link Lk should arise for sequences with small values of $|I_1(Lk - \widehat{Tw})|$ and $|I_2(Lk - \widehat{Tw})|$. For example, $I_1(18 - \widehat{Tw})^2 + I_2(18 - \widehat{Tw})^2$ is smaller in Fig. 9a than in Fig. 7a (0.8 versus 3.3).

Our discussion of the "typical" connection between multiple minima, length of molecule, angle set, and smallness of (I_1, I_2) has thus far been purely anecdotal, but is reinforced by the statistical study described in the next section.

4.2 Statistical study of multiple minima

We generated bifurcation diagrams for 5,017 random 200-bp sequences and 2,176 random 900-bp sequences with the angle set AS1 and analyzed each one as described in Sec. 3.4. In addition, we then selected the first thousand of the 200-bp sequences for further computation with the angle set AS2 (corresponding to smaller intrinsic curvature than AS1).

4.2.1 Nicked minima

The distribution of numbers of nicked minima for these three data sets is reported in Fig. 11. This figure demonstrates that the probability of finding multiple minima is increased by using larger wedge angles, or by lengthening the DNA. For instance, angle set AS1 produces larger intrinsic curvatures than AS2, and accordingly, we see in Fig. 11 that multiple nicked minima occur twice as often with AS1 as compared to AS2. Similarly, within a fixed angle set AS1, the increase in DNA length from 200 to 900 bp promotes the likelihood of finding multiple nicked minima by nearly threefold, from 23% to 64%.

Fig. 12 shows the results from a statistical analysis of the correlation of multiple nicked minima with the size of $(I_1(0), I_2(0))$ among sets of the same length and angle set. The distribution of $I_1(0)^2 + I_2(0)^2$ was first separated into deciles, so that the first decile contains the 10% of the molecules with lowest values of $I_1(0)^2 + I_2(0)^2$, the second decile the 10% with next lowest values, etc. Within each decile, the fraction of molecules with multiple nicked minima was then determined. The absolute ranges of $I_1(0)^2 + I_2(0)^2$ vary with dataset. The minimum, maximum, and median were (0.0004,26.3,2.25) for 200-AS1, (0.002,7.8,0.79) for 200-AS2, and (0.007,110,10.8) for 900-AS1.

This Figure demonstrates that the molecules most likely to exhibit multiple nicked minima are those with the lowest values of $I_1(0)^2 + I_2(0)^2$. This correlation is better for 200-bp DNA than for 900-bp DNA. There are at least two possible explanations for this. As remarked in Sec. 4.1, the continuum intrinsic curvatures for 900-bp DNA are generally larger than those for 200-bp. For such large perturbations, kinks in the bifurcation diagram emanating from a wide range of ΔLk can create a second nicked minimum, so that a perturbation expansion at $\Delta Lk = 0$ can not capture all behaviors. Indeed, over 60% of all 900-bp DNA exhibited two or more nicked minima. A second possible explanation is that the perturbation expansion we have used addresses transitions from the perfect diagram at fixed link, and as such is more directly related to predicting multiple cyclized minima. It is possible that a perturbation expansion directly tailored to the nicked problem would give a better correlation than that seen in Fig. 12. This conjecture is supported by the results in the next Section showing that for cyclized minima, the predictor $I_1^2 + I_2^2$ is highly accurate even for 900-bp.

4.2.2 Cyclized minima

Fig. 13 shows the correlation between the size of $I_1^2 + I_2^2$ and the appearance of multiple cyclized +1-topoisomer minima (link 20 for 200 bp DNA and link 87 for 900 bp DNA). The corresponding continuum links are $\Delta Lk = 20 - \widehat{Tw}$ and $87 - \widehat{Tw}$. The correlations for other links are similar (data not shown). The absolute ranges of $I_1(\Delta Lk)^2 + I_2(\Delta Lk)^2$ vary with dataset: the minimum, maximum, and median were (0.0003,34.4,2.23) for 200-AS1, (0.0003,8.5,0.83) for 200-AS2, and (0.005,111,10.1) for 900-AS1.

Even for 900-bp DNA, the likelihood of finding multiple cyclized minima with link Lk is highly coupled to the size of $I_1^2 + I_2^2$ at $\Delta Lk = Lk - \widehat{Tw}$. For the smallest intrinsic curvatures 200-AS2 especially (angle set AS2 and 200 bp), the correlation is nearly perfect, with all 10 (out of 1000) molecules exhibiting multiple cyclized minima occurring in the first decile. Indeed, the correlation for 200-AS2 is even better than shown in Fig. 13, with 4 out of the 10 multiple cyclized minimum examples occurring in the first $I_1^2 + I_2^2$ *percentile.*

4.3 Geometric Interpretation of I_1, I_2

The quantities I_1 and I_2 in Eq. 9 are global averages of the local intrinsic curvatures $\hat{u}_1$ and $\hat{u}_2$ that arise naturally from an analysis of symmetry-breaking in the continuum equilibrium equations. As such, these averages provide precise information regarding the emergence of multiple minima, but they do not lend themselves to an immediate geometric intuition. In this section, we present a heuristic discussion of their relation to simple geometric properties.

Recall first that the continuum intrinsic shape was derived by starting with a sequence of N director frames $(\hat{\mathbf{d}}_1^{(i)}, \hat{\mathbf{d}}_2^{(i)}, \hat{\mathbf{d}}_3^{(i)})$, whose relative orientations were determined from basepair-dependent wedge-angle parameters. The resulting intrinsic shape was then smoothed, and the rapid intrinsic twist removed, to give a continuum intrinsic shape $(\hat{\mathbf{D}}_1(s), \hat{\mathbf{D}}_2(s), \hat{\mathbf{D}}_3(s))$. We may now think of rediscretizing this continuum shape to give N director frames $(\hat{\mathbf{D}}_1^{(i)}, \hat{\mathbf{D}}_2^{(i)}, \hat{\mathbf{D}}_3^{(i)})$, 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.

Write the rotation between $(\hat{\mathbf{D}}_1^{(i)}, \hat{\mathbf{D}}_2^{(i)}, \hat{\mathbf{D}}_3^{(i)})$ and $(\hat{\mathbf{D}}_1^{(i+1)}, \hat{\mathbf{D}}_2^{(i+1)}, \hat{\mathbf{D}}_3^{(i+1)})$ as a product of fundamental rotations by three Euler angles $(\theta^{(i)}, \phi^{(i)}, \tau^{(i)})$ about $\hat{\mathbf{D}}_1^{(i)}$, $\hat{\mathbf{D}}_2^{(i)}$, and $\hat{\mathbf{D}}_3^{(i)}$. In the smoothed shape, these three angles will be small, and one may easily show that $\hat{u}_1(\frac{i-1}{N}) \approx N\theta^{(i)}$ and $\hat{u}_2(\frac{i-1}{N}) \approx N\phi^{(i)}$. Thus, if we discretize the integrals $I_1(0)$ and $I_2(0)$ as sums over the N new frames, we find

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

If the intrinsic shape is roughly planar and not too bent, then $I_1(0), I_2(0)$ are approximately equal to the Euler angles with respect to $\hat{\mathbf{D}}_2^{(1)}, \hat{\mathbf{D}}_1^{(1)}$ of the *overall* rotation between the first and last basepairs. We then find that the end-to-end cosine, i.e. the cosine of the angle between the initial and final tangent vectors, is approximately equal to $\cos(I_1(0))\cos(I_2(0))$, or to $1-(I_1(0)^2+I_2(0)^2)/2$. Clearly many approximations are involved in this analysis, but nevertheless, a good 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 also 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. 14. The intrinsic shape of the twenty 200-bp DNA with the largest values of $I_1(0)^2+I_2(0)^2$ (Fig. 14A) are essentially C-shaped (with relatively small end-to-end cosines), while the twenty with the smallest values (Fig. 14B) appear more or less 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 nicked minima, so Figs. 14A and B reinforce the theme noted by [KV] that S-shaped DNA are more likely than C-shaped DNA to yield multiple nicked minima. The quantity $I_1(0)^2+I_2(0)^2$ provides a more precise categorization of this geometric classification, and allows analysis of cases falling between the "C" and "S" extremes.

When we look at the more irregular 900-bp intrinsic shapes in Figs. 14C and D, the S-versus-C distinction is less apt, but one can still see a clear qualitative difference between molecules with large (14C) and small (14D) values of $I_1(0)^2+I_2(0)^2$. Large values correspond to highly bent DNA (with negative end-to-end cosines) and small values to relatively straight DNA.

4.4 Metropolis Monte Carlo

Stable ring equilibria are local minima of the elastic energy E in Eq. 2. The Metropolis Monte Carlo (MMC) simulations of [KV] simulate an equilibrium distribution of a nearly equivalent energy function, i.e. they sample the configuration space of nicked rings with the probability density function $\exp(-E/RT)$ (suitably normalized). How much information can a study of the set of energy minima give about the thermodynamic equilibrium distribution of conformations generated by MMC? For instance, how well can one predict the number, positions, and relative intensities of peaks in the Lk distribution $P(Lk)$ computed by MMC, based purely on the number, links, and elastic energies of the static energy nicked minima?

The data presented in Figs. 15 and 16 suggests, at least qualitatively, a clear connection. In Fig. 15, the rows represent four different molecules of 200 bp. In the left column, we plot the E vs. Lk projection of the bifurcation diagram as in Sec. 4.1. In the right column, we show the MMC $P(Lk)$. In order to compare these two quantities, we insert in the center column a plot we refer to as the Boltzman transform: at each Lk, we compute the sum of $e^{-E/RT}$ over all stable equilibria (i.e., points on the solid curves in the bifurcation diagram) with that link, and then normalize so that the entire curve becomes a probability density function. A similar approach was recently followed by (Swigon et al. 1998). This transform, although it involves only a small subset of the configuration space sampled by MMC, namely the local minima at prescribed Lk, fits $P(Lk)$ for nicked rings remarkably well. In Fig. 16, we repeat this analysis for a subset of five molecules of 900 bp, with similar results.

The quality of the fit for 900 bp is worse than for 200 bp, which is not surprising, since the Monte Carlo simulation would be expected to sample a wider range of configurations for the longer, effectively more flexible, molecules. However, even at 900 bp, the qualitative fit of the Boltzman transform to the Monte Carlo results is good enough to suggest that our fast equilibrium computations can be used as an effective guide for the design of molecules that can then be further studied with Monte Carlo simulations or actual experiments. For longer DNA molecules, we presume that entropic contributions would increasingly dominate the influence of the equilibria on the Monte Carlo distribution.

The molecules in Figs. 15 and 16 were selected from our database to illustrate the appearance of multiple peaks in $P(Lk)$ when there are multiple nicked minima. Heuristically, the number of nicked minima equals the number of peaks in $P(Lk)$ unless:

1. The Lk values of two nicked minima lie too close to each other (compared to the width of a typical peak in $P(Lk)$) for the Monte Carlo computation to be able to resolve them (see, e.g., the two central equilibria in Fig. 16E).

2. The relative difference in elastic energy between two nicked minima is so large that the higher energy nicked minimum has negligible probability of occurring in the Monte Carlo runs (see, e.g., Fig. 15B).

Whenever either of these conditions holds, one may see fewer peaks in $P(Lk)$ than the number of local minima, as seen in Fig. 15B, although one may also see a peak with an attached "shoulder" as in Fig. 15C. Exactly the same phenomenon arises between the number of minima and number of peaks in the Boltzman transform.

The correlation between minima and $P(Lk)$ is better quantified in the statistics presented in Secs. 4.4.1—4.4.3. Since approximately 10 hours is required for each MMC simulation, it was impractical to test the correlation using all 8,193 random sequences for which we have computed equilibria, so we instead randomly selected small subsets (usually 25-50 molecules) for MMC simulation.

4.4.1 Number of peaks versus number of nicked minima

Single nicked minima. As a control, we selected at random twenty-five 200-bp sequences and twenty-five 900-bp sequences for which the equilibrium approach predicted a single nicked minimum. $P(Lk)$ for these sequences exhibited a single peak without exception (as in Figs. 15A, 16A).

Double nicked minima. Applying criteria (1) and (2) from the previous Section, we can screen large numbers of DNA to identify multiple-minima sequences that are likely to exhibit two peaks or shoulders in $P(Lk)$. For example, applying only criterion (1), we selected the ten molecules from the data set 200-AS2 exhibiting the largest splits in Lk values of two nicked minima, and found that three of the $P(Lk)$ exhibited two separated peaks, while seven gave shoulders. Hence, with the benefit of this preliminary screening technique, we were able to find multiple nicked minima at lower intrinsic curvatures, or equivalently higher static persistence lengths, than was feasible for [KV] using only MMC simulations.

Triple and quadruple nicked minima. Although our database of 200-bp sequences contains twenty-four cases with three nicked minima (and one with four nicked minima), we found none with more than two peaks in $P(Lk)$. However, in our 900-bp database, we were able to find several examples that do in fact yield triple-peaked $P(Lk)$ (see Figs. 16C and D). The equilibrium computations also located twenty-six 900-bp molecules (out of 2,176) with four nicked minima. However, for the reasons cited above, the number of peaks in $P(Lk)$ for these molecules is three at most (see, e.g., Fig. 16E).

4.4.2 Lk correlation

Do the Lk values of individual nicked minima match the peak positions in the corresponding $P(Lk)$? To investigate this question, we studied molecules with single nicked minima, as well as molecules with two nicked minima that also exhibit two peaks in $P(Lk)$. Molecules exhibiting shoulders were discarded due to the large uncertainty in their peak positions. In Fig. 17, we compare the values of Lk for the nicked minima with the centers of the Gaussians fit to the $P(Lk)$ peaks. The global correlation coefficients are 0.992 ($n = 55$) and 0.962 ($n = 47$) for 200 bp and 900 bp, respectively.

While the above correlations are good, there do appear to be some differences between the Lk values of equilibria and the corresponding $P(Lk)$ peak positions in Monte Carlo, which merit further investigation. Some of these differences could be due to various computational errors and differences in the underlying energy models, as discussed in Sec. 4.4.4. In addition, there is no reason for the Monte Carlo peaks to occur *exactly* at the equilibria, as Monte Carlo simulations produce averages over many configurations.

4.4.3 Area/energy correlation for double nicked minima

Consider a molecule with two nicked minima of energies E_1, E_2, for which $P(Lk)$ has two peaks (or one peak plus a prominent shoulder). Based on the Boltzman probability density $e^{-E/RT}$, we might hope that the ratio of the areas A_1, A_2 of the Gaussians fit to the $P(Lk)$ peaks is approximately equal to $\exp(-(E_2 - E_1)/RT)$. This conjecture is supported by Fig. 18. The correlation coefficients are 0.95 ($n = 38$) and 0.89 ($n = 35$) for 200 and 900 bp respectively. The correlation is weaker for the 900 bp case, perhaps because excursions from the nicked minima are more frequent and distant.

4.4.4 Possible sources of error

Continuum computations.

Once a sequence and angle set are chosen, the main source of uncertainty in the continuum computations is the choice of the smoothing parameters in constructing a continuum centerline. We selected ten 200-bp and ten 900-bp sequences each exhibiting a unique nicked minimum, and calculated the average $\bar{SD}$ of the ten standard deviations in the energy and link of the nicked minimum as a function of the smoothing window-width w (see Sec. 2.2). For 200 bp, as w ranges from 10 to 50 bp, we find $\bar{SD} = 0.02$ for Lk and $\bar{SD} = 0.14$ for E/K_1. For 900 bp, as w ranges from 20 to 80 bp, we find $\bar{SD} = 0.04$ for Lk and $\bar{SD} = 0.16$ for E/K_1. These uncertainties should have no significant effect on the results presented.

Monte Carlo simulations

Following [KV], we analyzed the effect on the Monte Carlo distribution $P(Lk)$ of the choice of segment length. We simulated single-peaked $P(Lk)$ for ten 200 bp molecules with segment lengths of 5, 10, 20 and 25 bp per segment. The standard deviation of the $P(Lk)$ peak positions (as determined by Gaussian fits) gives an estimate of ± 0.04 for the 200-bp Lk position uncertainty due to segmentation. This uncertainty is about twice as large for 900-bp Lk fitting, and thus should have no significant effect on the reported Lk correlations.

Differences between continuum and Monte Carlo energies

Our discretization of the continuum energy from Eq. 2 and the energy function in the [KV] Monte Carlo code are close, but not identical. [KV] use a piecewise linear approximation to the DNA centerline and assign a single twist to each linear segment, whereas the rod model uses a continuous centerline and continuous twist function (discretized in AUTO via a piecewise fifth-order polynomial). For the segmentation lengths and discretizations chosen here, neither of these differences appears to be crucial, but a further study of the connection between Monte Carlo simulations and equilibria could benefit from an exact matching of the underlying energy models. One would expect that the fit between equilibria and key Monte Carlo features would only be improved if these differences were removed. Finally, [KV] account for rod self-contact by assigning an effective diameter d to the DNA and disallowing interpenetration during the simulation. However, this should not have a significant effect on our comparisons, as we are comparing nicked minima that are quite far from self-contact; in fact, [KV] assert that their results are not affected significantly by changing d over the range from 0 nm to 8 nm.

5 Conclusion

Since experimental (discrete) topoisomer distributions have not been observed to be bimodal, it has generally been concluded that a DNA molecule has only one nicked (local) minimum of its energy. This conclusion was reinforced by several numerical simulations of the (continuous) link equilibrium distribution $P(Lk)$ for models of intrinsically straight DNA. Recently, however, (Katritch and Vologodskii 1997) reported Monte Carlo simulations yielding bimodal $P(Lk)$ for DNA with sufficiently large intrinsic curvature, and computed multiple nicked minima using a simulated annealing technique. In this paper, we explain the phenomenon of multiple nicked minima. Essentially we reverse the (Katritch and Vologodskii 1997) procedure: we describe ways first to identify sequences with multiple nicked minima, second to compute these equilibria, and third, to predict whether the DNA will exhibit multiple peaked $P(Lk)$. In fact, we show that multiple nicked minima can arise even for arbitrarily small intrinsic curvatures, in particular well below the range where the phenomenon was reported by (Katritch and Vologodskii 1997).

5.1 Multiple minima for DNA rings can be predicted from symmetry-breaking

The emergence of multiple minima for small intrinsic curvature is associated with the fact that an intrinsically straight DNA does not, in fact, have a single nicked minimum, but rather a circle of minima related by the register symmetry. When the register symmetry is broken by the addition of intrinsic curvature, this circle of degenerate minima can yield multiple isolated minima. Our perturbation analysis of this symmetry-breaking provides a simple and accurate predictor for the presence of multiple nicked minima. This predictor refines the heuristic argument given by (Katritch and Vologodskii 1997) that "S"-shaped DNA are more likely to yield multiple nicked minima than "C"-shaped DNA. Essentially the same predictor can be applied to the case of cyclized DNA rings, so that we can identify DNA sequences exhibiting multiple cyclized minima of the same link.

5.2 Computation of equilibria provides good predictions of equilibrium distributions

Armed with this understanding of multiple nicked minima, we quantitatively probed the connection between multiple minima of energy and multiple peaks of $P(Lk)$. It is certainly to be expected that at very low temperatures, the gross features of an equilibrium distribution will be completely dominated by the set of minima of the underlying energy. To what extent is this still true for models of 200–900 bp DNA at room temperature?

Comparing our equilibrium computations with simulations done with the (Katritch and Vologodskii 1997) Monte Carlo code, we find a good correlation between the number of nicked minima and the number of peaks in $P(Lk)$, particularly after effects of separations in peak location and energies are factored in. Moreover, the Lk values of the nicked minima correspond closely to the centers of Gaussian fits to the peaks in $P(Lk)$, and the elastic energies of the nicked minima are related to the areas of these Gaussians according to the Boltzman distribution. In a more speculative vein, we observe that a naive "Boltzman transform" of our bifurcation diagram gives a good fit to the main features of $P(Lk)$.

Having observed this correlation, we propose that equilibrium computations can serve as valuable guides before carrying out more computationally intensive procedures like Monte Carlo simulation on DNA of these scales. For example, we used our equilibrium analysis to screen more than 8,000 DNA sequences to identify a small subset likely to give interesting $P(Lk)$. Then, by running Monte Carlo on these selected sequences, we discovered phenomena not observed previously, such as two-peaked $P(Lk)$ for sequences of only 200 basepairs.

Acknowledgments: We thank Seva Katritch for putting his Monte Carlo program at our disposal.

References

Antman, S. S. (1995). *Nonlinear Problems of Elasticity.* New York: Springer-Verlag.

Bates, A. D. and A. Maxwell (1993). *DNA Topology.* Oxford: Oxford University Press.

Bauer, W., R. Lund, and J. White (1993). Twist and writhe of a DNA loop containing intrinsic bends. *Proc. Natl. Acad. Sci. USA 90*, 833–837.

Benham, C. J. (1979). An elastic model of the large-scale structure of duplex DNA. *Biopolymers 18*, 609–623.

Bolshoy, A., P. McNamara, R. E. Harrington, and E. N. Trifonov (1991). Curved DNA without A-A: Experimental estimation of all 16 wedge angles. *Proc. Natl. Acad. Sci. USA 88*, 2312–2316.

Călugăreanu, G. (1959). L'intégrale de Gauss et l'analyse des nœuds tridimensionnels. *Revue de Mathématiques Pures and Appliquées 4*, 5–20.

Călugăreanu, G. (1961). Sur les classes d'isotopie des nœuds tridimensionnels et leurs invariants. *Czechoslovak Math. J. 1*, 588–625.

De Santis, P., A. Palleschi, M. Savino, and A. Scipioni (1992). Theoretical prediction of the gel electrophoretic retardation changers due to point mutations in a tract of SV40 DNA. *Biophys. Chem. 42*, 147–152.

Depew, R. E. and J. C. Wang (1975). Conformational fluctuations of DNA helix. *Proc. Nathl. Acad. Sci. USA 72*, 4275–4279.

Dichmann, D. J., Y. W. Li, and J. H. Maddocks (1996). Hamiltonian formulations and symmetries in rod mechanics. In J. Mesirov, K. Schulten, and D. Sumners (Eds.), *Mathematical Approaches to Biomolecular Structure and Dynamics*, Volume 82, pp. 71–113. New York: Springer, New York. IMA Volumes in Mathematics and its Applications.

Doedel, E. J., H. B. Keller, and J. P. Kernévez (1991). Numerical analysis and control of bifurcation problems, parts I and II. *Int. J. of Bif. and Chaos 3, 4*, 493–520, 745–772.

Dubochet, J., M. Adrian, I. Dustin, P. Furrer, and A. Stasiak (1992). Cryoelectron microscopy of DNA molecules in solution. *Methods Enzymol. 211*, 507–518.

Fuller, F. B. (1971). The writhing number of a space curve. *Proc. Natl. Acad. Sci. USA 68*(4), 815–819.

Gelfand, I. M. and S. V. Fomin (1963). *Calculus of Variations*. Englewood Cliffs, New Jersey: Prentice-Hall.

Horowitz, D. S. and J. C. Wang (1984). Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J. Mol. Biol. 173*, 75–91.

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

Kahn, J. D. and D. M. Crothers (1998). Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics. *J. Mol. Biol. 276*, 287–309.

Katritch, V. and A. Vologodskii (1997). The effect of intrinsic curvature on conformational properties of circular DNA. *Biophysical Journal 72*, 1070–1079.

Kehrbaum, S. (1997). *Hamiltonian formulations of the equilibrium conditions governing elastic rods: qualitative analysis and effective properties*. Ph. D. thesis, University of Maryland, College Park.

Kehrbaum, S. and J. H. Maddocks (1997). Elastic rods, rigid bodes, quaternions and the last quadrature. *Phil. Trans. R. Soc. Lond. A 355*, 2117–2136.

Klenin, K., A. Vologodskii, V. Anshelevich, A. Dykhne, and M. D. Frank-Kamenetskii (1991). Computer simulation of DNA supercoiling. *J. Mol. Biol. 217*, 413–419.

Klenin, K. V., M. D. Frank-Kamenetskii, and J. Langowski (1995). Modulation of intramolecular interactions in superhelical DNA by curved sequences: a Monte Carlo simulation study. *Biophys. J. 68*, 81–88.

LeBret, M. (1979). Catastrophic variation of twist and writhing of circular DNA with constraint. *Biopolymers 18*, 1709–1725.

Levene, S. (1994). Conformation and energetics of supercoiled DNA: Experimental and theoretical studies. *Nucleic Acids and Molecular Biology 8*, 119–132.

Levene, S. D. and D. M. Crothers (1986). Topological distributions and the torsional rigidity of DNA: A Monte Carlo study of DNA circles. *J. Mol. Biol. 189*, 73–83.

Li, Y. and J. H. Maddocks (1999). On the computation of equilibria of elastic rods, Part I: Integrals, symmetry and a Hamiltonian formulation. *J. Comp. Phys.*. In press.

Manning, R. S. and J. H. Maddocks (1999). Symmetry breaking and the twisted elastic ring. *Computer Methods in Applied Mechanics and Engineering 370*, 313–330.

Manning, R. S., J. H. Maddocks, and J. D. Kahn (1996). A continuum rod model of sequence-dependent DNA structure. *J. Chem. Phys. 105*, 5626–5646.

Manning, R. S., K. A. Rogers, and J. H. Maddocks (1998). Isoperimetric conjugate points with application to the stability of DNA minicircles. *Proc. R. Soc. London A 454*, 3047–3074.

Moffatt, H. K. and R. L. Ricca (1992). Helicity and the Călugăreanu invariant. *Proc. R. Soc. Lond. A 439*, 411–429.

Olson, W. K. (1996). Simulating DNA at low resolution. *Current Opinion in Structural Biology 6*, 242–256.

Olson, W. K. and V. B. Zhurkin (1996). Twenty years of DNA bending. *Biological Structure and Dynamics 2*, 341–370.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1992). *Numerical Recipes in Fortran.* Cambridge [England]: Cambridge University Press.

Pulleyblank, D. E., M. Shure, D. Tang, J. Vinograd, and H. P. Vosberg (1975). Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed ciruclar DNA: formation of a Boltzmann distribution of topological isomers. *Proc. Natl. Acad. Sci. USA 72*, 4280–4284.

Rippe, K., P. H. von Hippel, and J. Langowski (1995). Action at a distance: DNA-looping and initiation of transcription. *Trends in Biological Sciences 20*, 500–505.

Sanghani, S. R., K. Zakrzewska, S. C. Harvey, and R. Lavery (1996). Molecular modelling of $(A_4T_4NN)_n$ and $(T_4A_4NN)_n$: sequence elements responsible for curvature. *Nucl. Acids*

Res. 24, 1632–1637.

Schlick, T. (1995). Modeling superhelical DNA: Recent analytical and dynamical approaches. *Current Opinion in Structural Biology 5*, 245–262.

Shimada, J. and H. Yamakawa (1988). Moments for DNA topoisomers: the helical wormlike chain. *Biopolymers 27*, 657–673.

Shore, D. and R. L. Baldwin (1983). Energetics of DNA twisting. II. Topoisomer analysis. *J. Mol. Biol. 170*, 983–1007.

Stasiak, A. (1996). Circular DNA. In J. A. Semlyen (Ed.), *Large Ring Molecules*, Chapter 2, pp. 43–97. New York: John Wiley & Sons.

Swigon, D., B. Coleman, and I. Tobias (1998). The elastic rod model for DNA and its application to the tertiary structure of DNA minicircles in mononucleosomes. *Biophys. J. 74*, 2515–2530.

Tobias, I. (1998). A theory of thermal fluctuations in DNA miniplasmids. *Biophys. J. 74*, 2545–2553.

Tobias, I. and W. K. Olson (1993). The effect of intrinsic curvature on supercoiling: Predictions of elasticity theory. *Biopolymers 33*, 639–646.

Trifonov, E. N., R. K. Z. Tan, and S. C. Harvey (1987). Static persistence length of DNA. In W. K. Olson, M. H. Sarma, R. H. Sarma, and M. Sundaralingam (Eds.), *DNA Bending and Curvature*, pp. 243–253. New York: Adenine Press.

Westcott, T. P., I. Tobias, and W. K. Olson (1995). Elasticity theory and numerical analysis of DNA supercoiling: An application to DNA looping. *J. Phys. Chem. 99*, 17926–17935.

White, J. H. (1969). Self-linking and the Gauss integral in higher dimensions. *American Journal of Mathematics 91*, 963–728.

Yang, Y., I. Tobias, and W. Olson (1993). Finite element analysis of DNA supercoiling. *J. Chem. Phys. 98*, 1673–1686.

Yang, Y., T. P. Westcott, S. C. Pedersen, I. Tobias, and W. K. Olson (1995). Effects of localized bending on DNA supercoiling. *Trends in Biological Sciences 20*, 313–319.

Figure Captions

Figure 1: Ring boundary conditions: the centerline $\mathbf{r}(s)$ is depicted as a tube, and is constrained to form a smooth ring. The director $\mathbf{d}_1(s)$ is depicted as a ribbon, and a twist angle α is imposed between $\mathbf{d}_1(0)$ and $\mathbf{d}_1(1)$.

Figure 2: Natural and DNA framings of a ring equilibrium. The two framings are related through the known DNA-frame intrinsic twist (see text). Note that a cyclized DNA frame as shown here will not in general correspond to a cyclized natural frame.

Figure 3: Portion of the bifurcation diagram for the perfect problem for $K_3 = 0.8K_1$. As the link ΔLk is varied, the energy E and link ΔLk of the ring equilibria are plotted. Solid curves denote stable equilibria (local minima), while dashed curves denote unstable equilibria.

Figure 4: Portion of the bifurcation diagram for an imperfect rod. Stability is indicated by line style as in Fig.3. 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.3 yields two branches in the imperfect diagram, one stable and one unstable.

Figure 5: A finite-dimensional example of a function $z = r^4/4 - r^2/2$ ($r = \sqrt{x^2 + y^2}$) with a circle of local minima $x^2 + y^2 = 1$ (drawn with a thick line).

Figure 6: The surface from Fig.5 with a small linear perturbation $x/10$ added. The circle of critical points breaks up to yield just two critical points, marked with dots, one a local minimum, and one a saddle point.

Figure 7: Qualitatively simple imperfect diagrams for a 200-bp DNA molecule modeled by

angle sets AS1 and AS2. The perfect diagram is superimposed in the thinner linestyle. Stability is indicated by line style (stable equilibria on solid lines, unstable equilibria on dashed lines). Nicked minima (for the imperfect problem) are denoted by black triangles, and cyclized minima by gray circles. As in the perfect diagram, for this molecule there is a single nicked minimum, and at most one cyclized minimum at each link.

Figure 8: Imperfect diagrams showing a double well near the nicked minimum of the perfect problem. The intrinsic shape is for a 200-bp DNA modeled by angle sets AS1 and AS2. The perfect diagram is superimposed, stability is indicated by linestyle, and nicked and cyclized minima are labeled as in Fig.7. For each angle set, the kink introduces a second nicked minimum for the imperfect problem, although for AS2, the second nicked minimum is on the verge of disappearing.

Figure 9: Imperfect diagrams showing a kink near link 18. The intrinsic shape is for a 200-bp DNA modeled by angle sets AS1 and AS2. The perfect diagram is superimposed, stability is indicated by linestyle, and nicked and cyclized minima are labeled as in Fig.7. The kink introduces a second cyclized minimum of link 18 for AS1, but with AS2 the kink is straightened and only one link-18 cyclized minimum occurs.

Figure 10: Two 900 bp molecules (parts A and B) each exhibiting multiple minima with a significant difference in writhe. The right part of each figure shows a E vs. Lk bifurcation diagram as in Figs. 7–9 (but with larger Lk corresponding to 900 bp). To the left are the physical configurations corresponding to the two minima (two cyclized minima with $Lk = 85$ in part A, and two nicked minima in part B). The physical configurations are drawn with their naturally-framed ribbons.

Figure 11: The influence of intrinsic curvature on the likelihood of multiple nicked minima. Larger effective intrinsic curvature within the continuum model can be induced either by increasing the size of the wedge angles (200-AS2 to 200-AS1) or by increasing the DNA length (200-AS1 to 900-AS1). For each of the three angle sets, the percentage of molecules exhibiting 1, 2, 3, or 4 nicked minima is shown.

Figure 12: Fraction of DNA with multiple nicked minima as a function of the deciles of $I_1(0)^2 + I_2(0)^2$ (see text). These curves support the hypothesis that smaller values of $I_1(0)^2 + I_2(0)^2$ predict multiple nicked minima.

Figure 13: Fraction of DNA with multiple cyclized +1-topoisomer minima as a function of the deciles of $I_1(\Delta Lk)^2 + I_2(\Delta Lk)^2$ (see text). These curves strongly support the hypothesis that smaller values of $I_1(\Delta Lk)^2 + I_2(\Delta Lk)^2$ predict multiple cyclized minima.

Figure 14: Projections of intrinsic shapes for molecules with the highest and lowest values of $I_1(0)^2 + I_2(0)^2$. Figs. A and B each show twenty 200 bp DNA (aligned according to their principal inertial moments); 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 closer to being S-shaped. Figs. C and D each contain twenty 900 bp DNA (aligned according to the first tangent vector); the molecules with large $I_1(0)^2 + I_2(0)^2$ (Fig. C) have global bends of at least 180°, while those with small $I_1(0)^2 + I_2(0)^2$ (Fig. D) are globally straighter.

Figure 15: Comparison of bifurcation diagrams and Monte Carlo $P(Lk)$ for four 200-bp DNA A–D (angle set AS1). The energy in the bifurcation diagram is shifted downward by E_{min}, the energy of the global minimum. For molecule A, a single nicked minimum yields a single Monte Carlo peak. If there are two nicked minima, then, depending on the difference in their energies

(and Lk), the Monte Carlo can either produce no second peak (B), a shoulder or nascent second peak (C) or a well-separated second peak (D). In the center column, the bifurcation diagram is "Boltzman-transformed" (see text) to facilitate comparison with Monte Carlo data.

Figure 16: Comparison of bifurcation diagrams and Monte Carlo $P(Lk)$ for five 900-bp DNA A–E (angle set AS1). As in Fig.14, the number of peaks in $P(Lk)$ is less than or equal to the number of nicked minima. Whereas the molecules B and C show the same number of peaks as nicked minima (2 and 3, respectively), the 4 nicked minima of molecules D and E only give rise to three separated peaks in $P(Lk)$.

Figure 17: Correlation between Lk values of nicked minima vs peak positions in $P(Lk)$ (as determined by Gaussian fits), for single and double nicked equilibria of either 200-bp (A) or 900-bp (B).

Figure 18: Correlation between the difference in elastic energy of a pair of nicked minima and the ratio of areas of the corresponding Lk peaks. The results support the conjecture that $A_2/A_1 \approx \exp(-(E_2 - E_1)/RT)$.

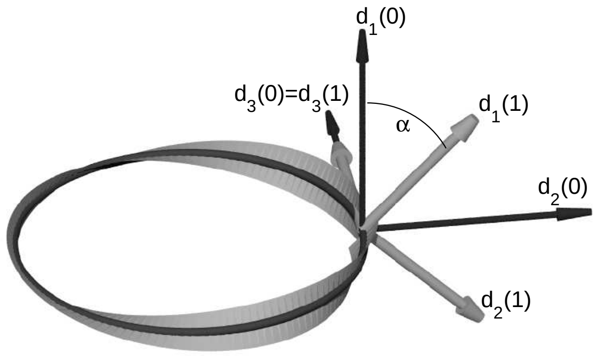


Figure 1: Ring boundary conditions: the centerline $\mathbf{r}(s)$ is depicted as a tube, and is constrained to form a smooth ring. The director $\mathbf{d}_1(s)$ is depicted as a ribbon, and a twist angle α is imposed between $\mathbf{d}_1(0)$ and $\mathbf{d}_1(1)$.

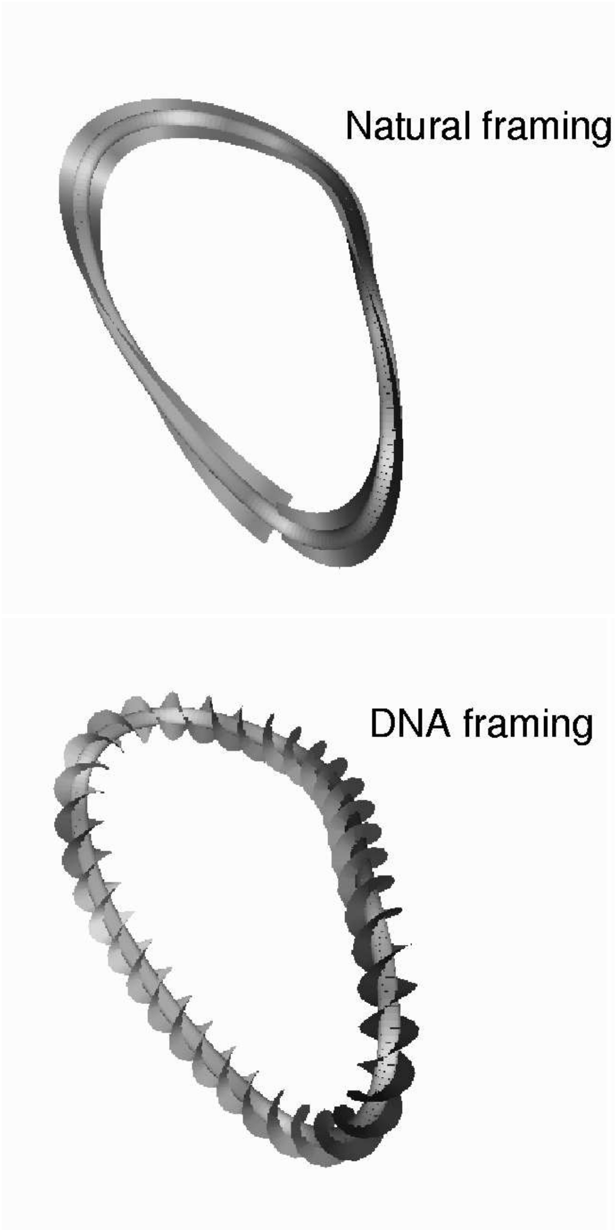


Figure 2: Natural and DNA framings of a ring equilibrium. The two framings are related through the known DNA-frame intrinsic twist (see text). Note that a cyclized DNA frame as shown here will not in general correspond to a cyclized natural frame.

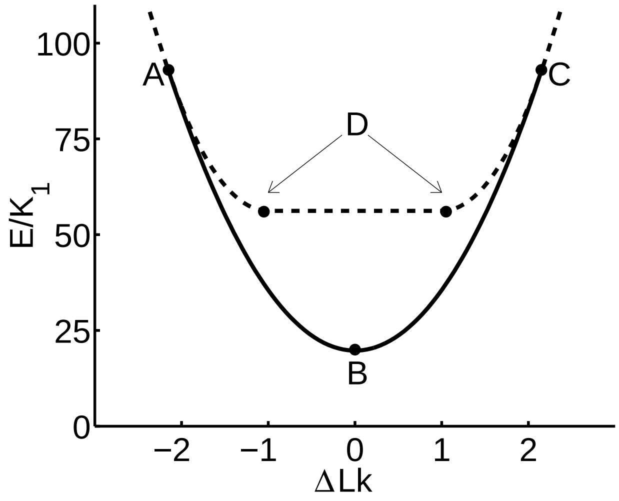


Figure 3: Portion of the bifurcation diagram for the perfect problem for $K_3 = 0.8K_1$. As the link ΔLk is varied, the energy E and link ΔLk of the ring equilibria are plotted. Solid curves denote stable equilibria (local minima), while dashed curves denote unstable equilibria.

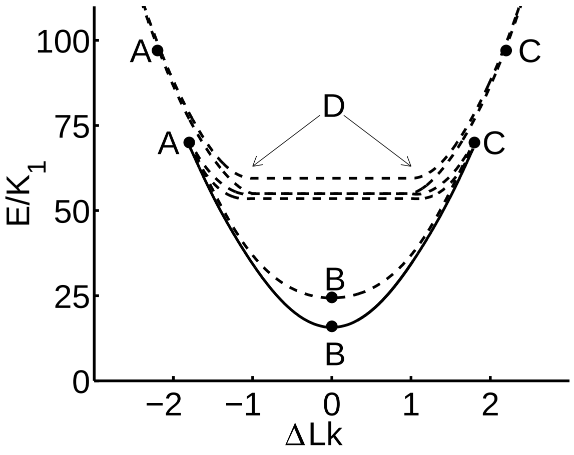


Figure 4: Portion of the bifurcation diagram for an imperfect rod. Stability is indicated by line style as in Fig. 3. Approximate images of points A, B, C, and D are marked. Each circle labelled A or C represents *two* images of the respective point in the perfect diagram. The stable trivial branch ABC from Fig. 3 yields two branches in the imperfect diagram, one stable and one unstable.

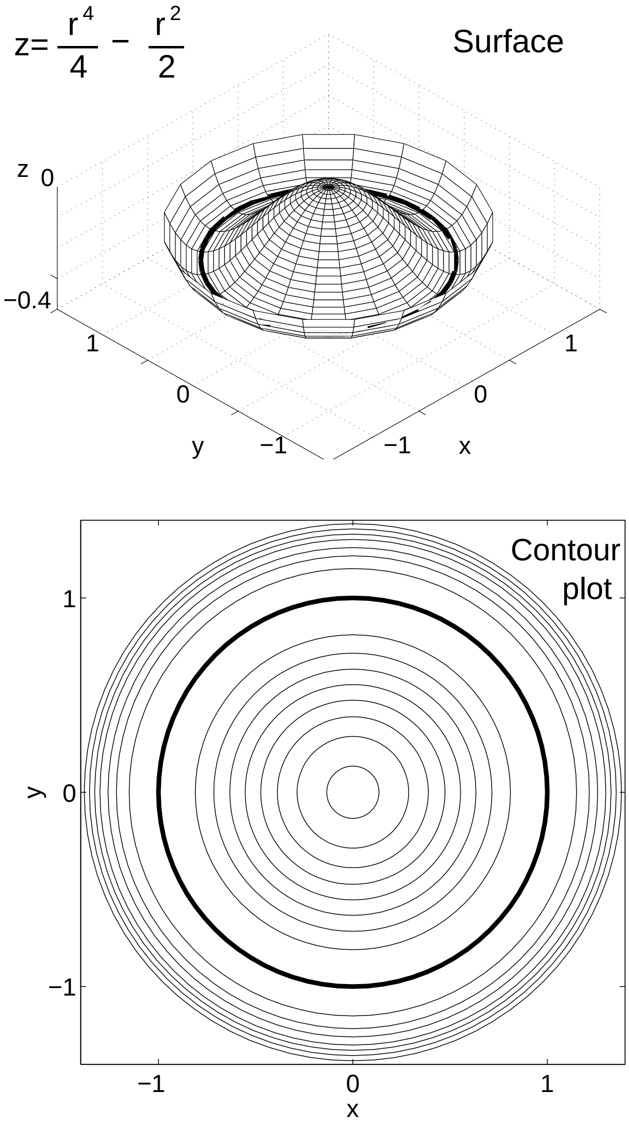


Figure 5: A finite-dimensional example of a function $z = r^4/4 - r^2/2$ ($r = \sqrt{x^2 + y^2}$) with a circle of local minima $x^2 + y^2 = 1$ (drawn with a thick line).

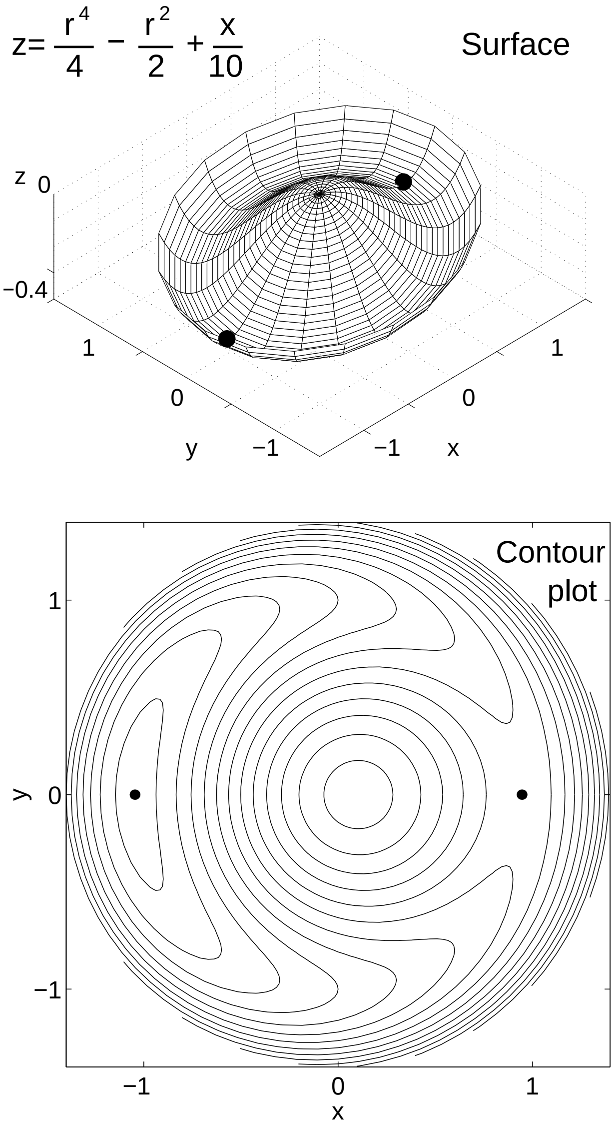


Figure 6: The surface from Fig. 5 with a small linear perturbation $x/10$ added. The circle of critical points breaks up to yield just two critical points, marked with dots, one a local minimum, and one a saddle point.

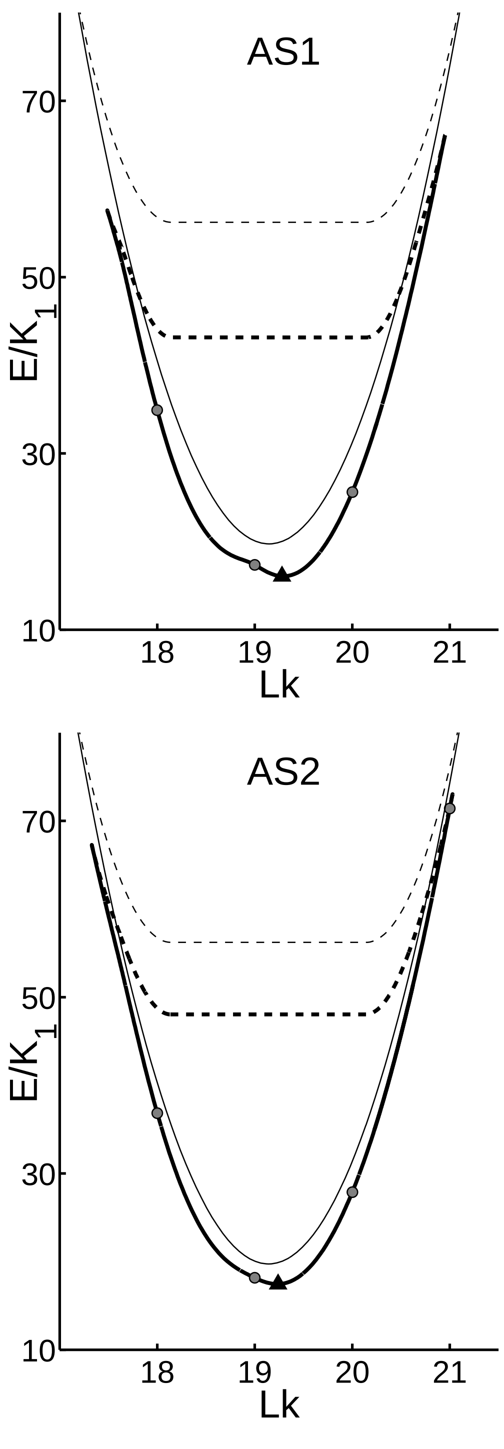


Figure 7: Qualitatively simple imperfect diagrams for a 200-bp DNA molecule modeled by angle sets AS1 and AS2. The perfect diagram is superimposed in the thinner linestyle. Stability is indicated by line style (stable equilibria on solid lines, unstable equilibria on dashed lines). Nicked minima (for the imperfect problem) are denoted by black triangles, and cyclized minima by gray circles. As in the perfect diagram, for this molecule there is a single nicked minimum, and at most one cyclized minimum at each link.

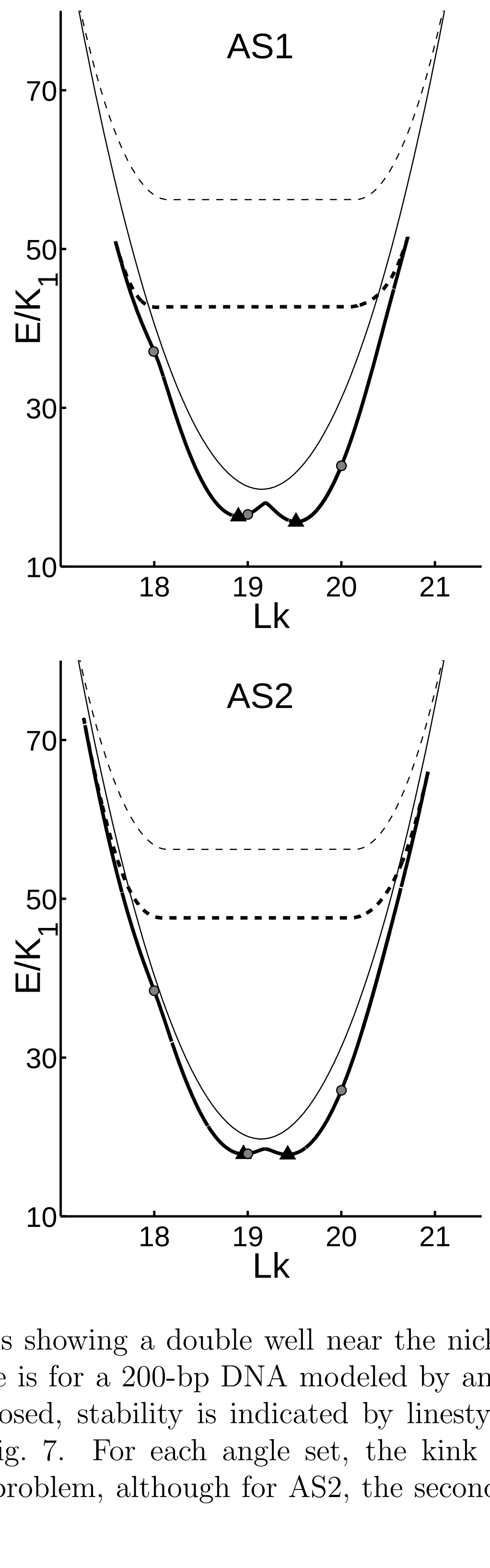


Figure 8: Imperfect diagrams showing a double well near the nicked minimum of the perfect problem. The intrinsic shape is for a 200-bp DNA modeled by angle sets AS1 and AS2. The perfect diagram is superimposed, stability is indicated by linestyle, and nicked and cyclized minima are labeled as in Fig. 7. For each angle set, the kink introduces a second nicked minimum for the imperfect problem, although for AS2, the second nicked minimum is on the verge of disappearing.

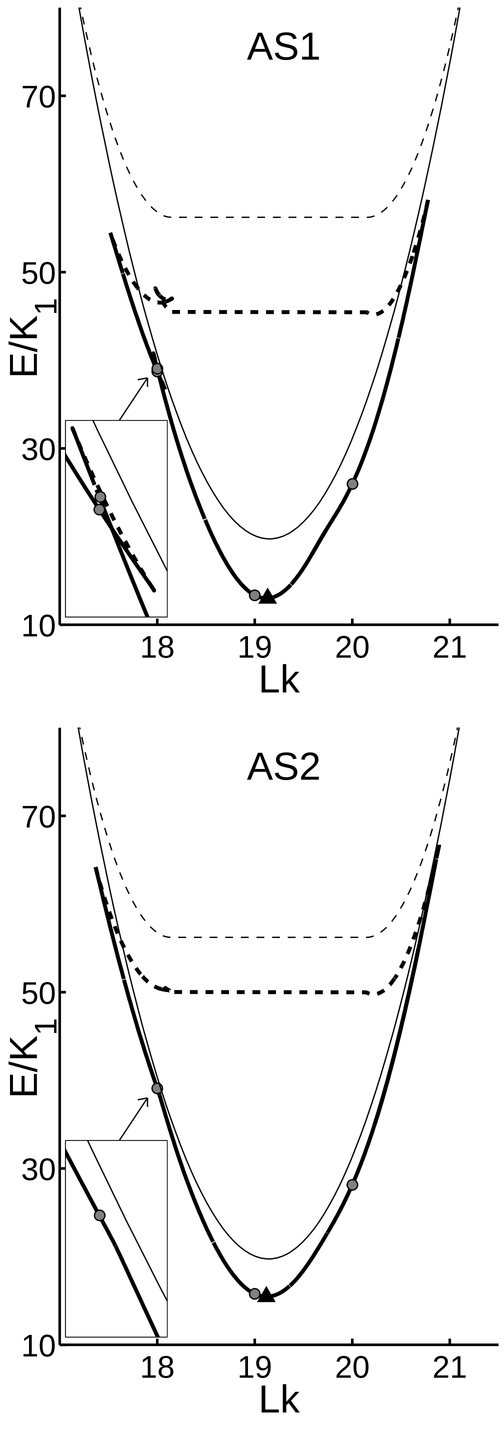


Figure 9: Imperfect diagrams showing a kink near link 18. The intrinsic shape is for a 200-bp DNA modeled by angle sets AS1 and AS2. The perfect diagram is superimposed, stability is indicated by linestyle, and nicked and cyclized minima are labeled as in Fig. 7. The kink introduces a second cyclized minimum of link 18 for AS1, but with AS2 the kink is straightened and only one link-18 cyclized minimum occurs.

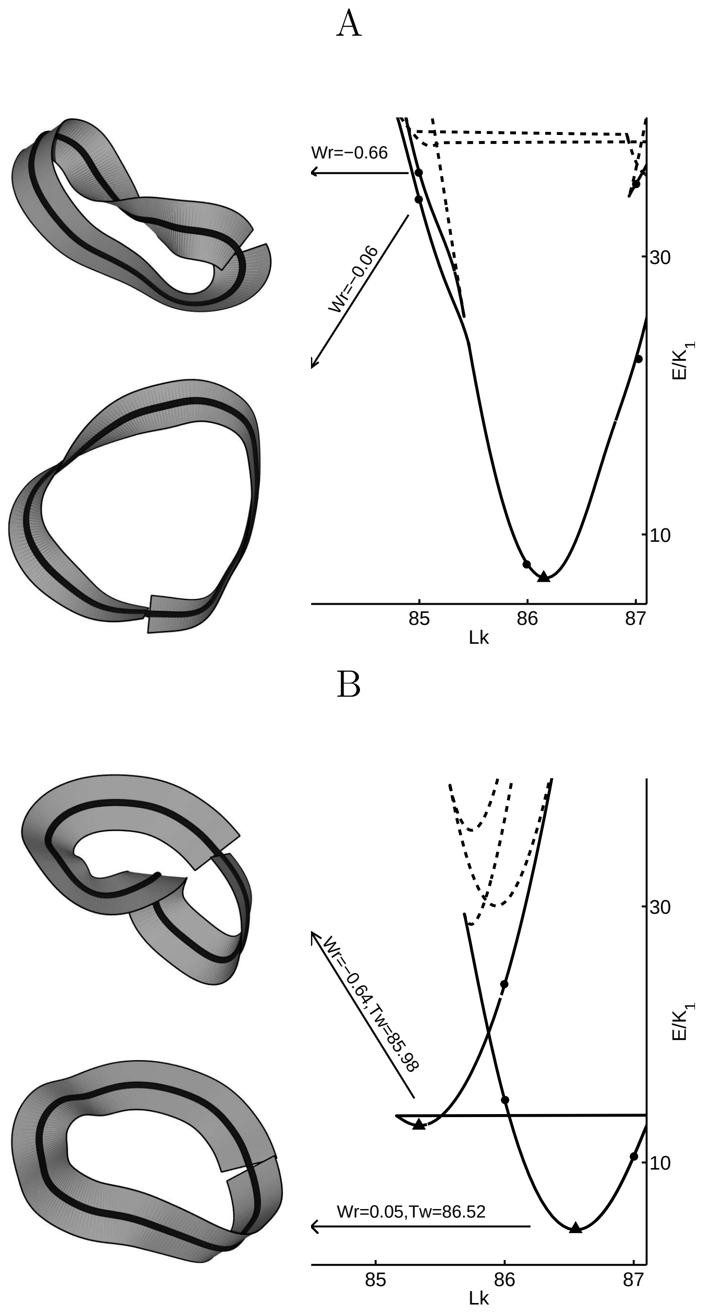


Figure 10: Two 900 bp molecules (parts A and B) each exhibiting multiple minima with a significant difference in writhe. The right part of each figure shows a E vs. Lk bifurcation diagram as in Figs. 7–9 (but with larger Lk corresponding to 900 bp). To the left are the physical configurations corresponding to the two minima (two cyclized minima with $Lk = 85$ in part A, and two nicked minima in part B). The physical configurations are drawn with their naturally-framed ribbons.

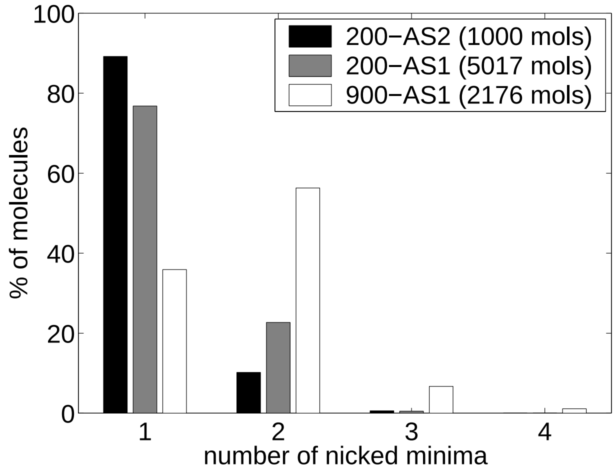


Figure 11: The influence of intrinsic curvature on the likelihood of multiple nicked minima. Larger effective intrinsic curvature within the continuum model can be induced either by increasing the size of the wedge angles (200-AS2 to 200-AS1) or by increasing the DNA length (200-AS1 to 900-AS1). For each of the three angle sets, the percentage of molecules exhibiting 1, 2, 3, or 4 nicked minima is shown.

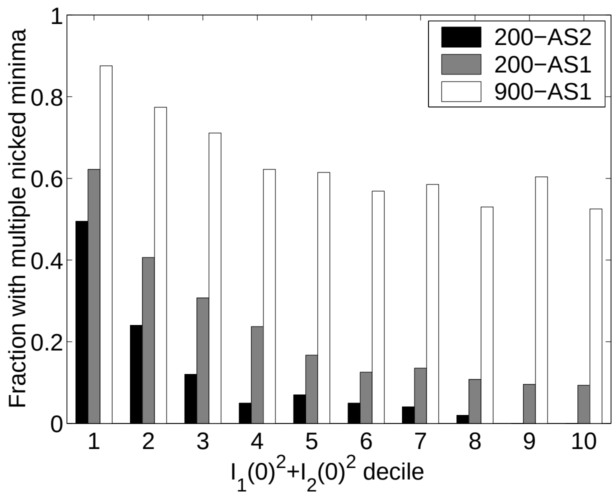


Figure 12: Fraction of DNA with multiple nicked minima as a function of the deciles of $I_1(0)^2 + I_2(0)^2$ (see text). These curves support the hypothesis that smaller values of $I_1(0)^2 + I_2(0)^2$ predict multiple nicked minima.

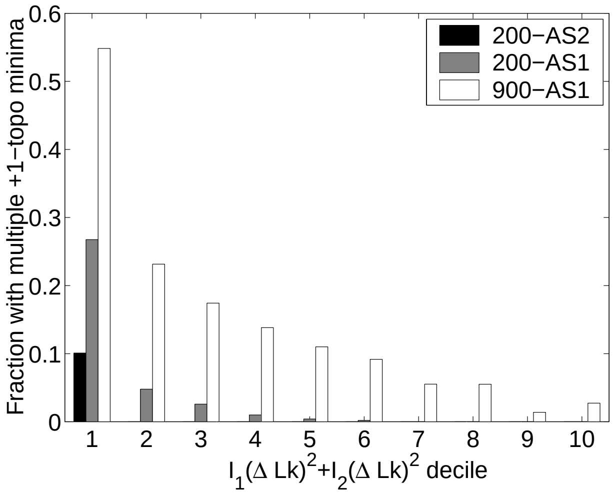


Figure 13: Fraction of DNA with multiple cyclized +1-topoisomer minima as a function of the deciles of $I_1(\Delta Lk)^2 + I_2(\Delta Lk)^2$ (see text). These curves strongly support the hypothesis that smaller values of $I_1(\Delta Lk)^2 + I_2(\Delta Lk)^2$ predict multiple cyclized minima.

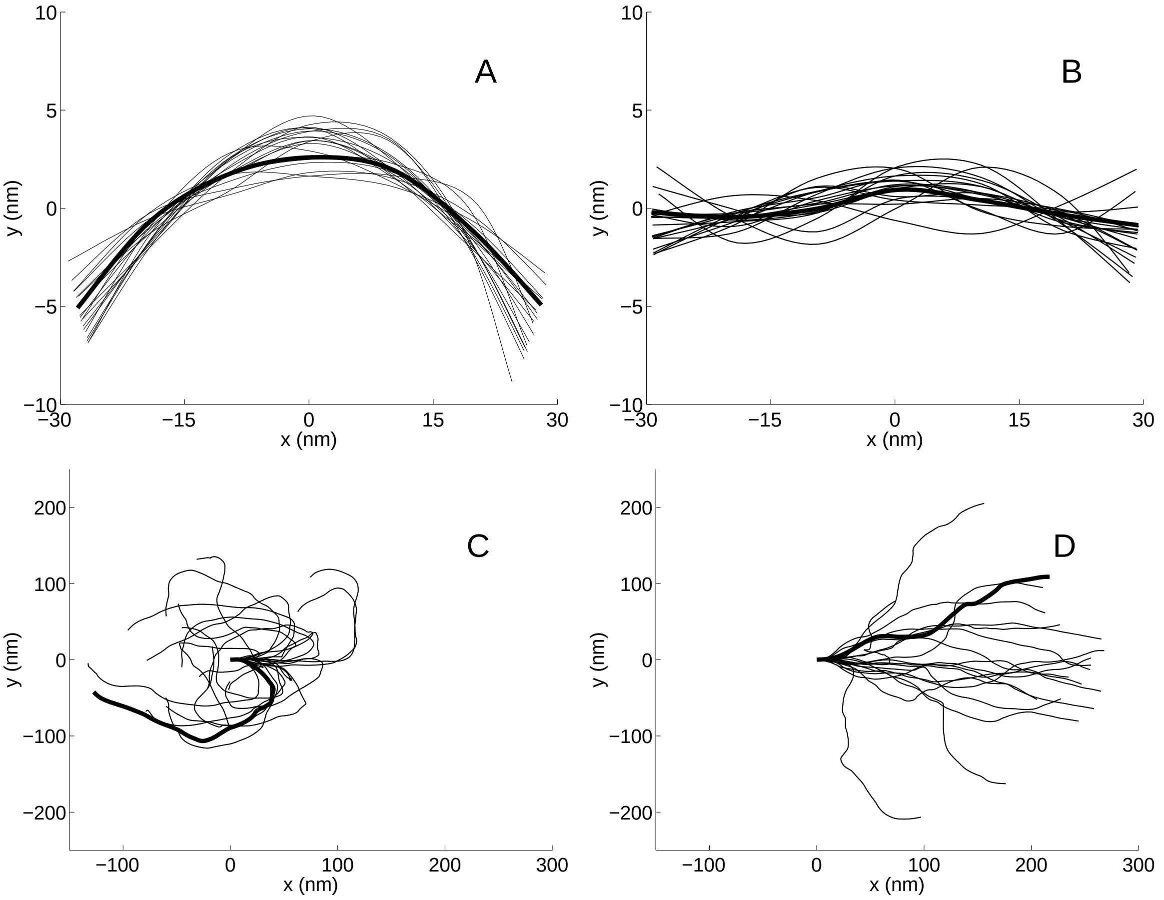


Figure 14: Projections of intrinsic shapes for molecules with the highest and lowest values of $I_1(0)^2 + I_2(0)^2$. Figs. A and B each show twenty 200 bp DNA (aligned according to their principal inertial moments); the molecules with large $I_1(0)^2 + I_2(0)^2$ (Fig. A) are C-shaped, while those with small $I_1(0)^2 + I_2(0)^2$ (Fig. B) are closer to being S-shaped. Figs. C and D each contain twenty 900 bp DNA (aligned according to the first tangent vector); the molecules with large $I_1(0)^2 + I_2(0)^2$ (Fig. C) have global bends of at least 180°, while those with small $I_1(0)^2 + I_2(0)^2$ (Fig. D) are globally straighter.

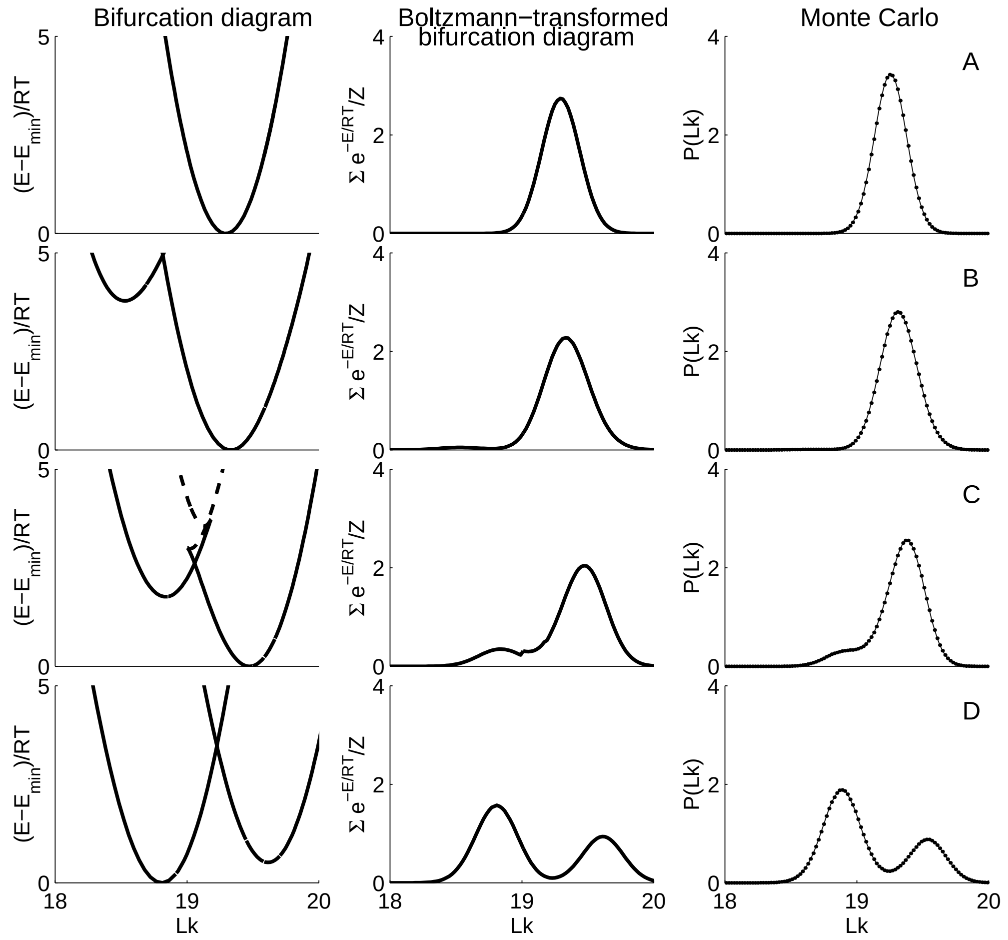


Figure 15: Comparison of bifurcation diagrams and Monte Carlo $P(Lk)$ for four 200-bp DNA A–D (angle set AS1). The energy in the bifurcation diagram is shifted downward by E_{min}, the energy of the global minimum. For molecule A, a single nicked minimum yields a single Monte Carlo peak. If there are two nicked minima, then, depending on the difference in their energies (and Lk), the Monte Carlo can either produce no second peak (B), a shoulder or nascent second peak (C) or a well-separated second peak (D). In the center column, the bifurcation diagram is "Boltzman-transformed" (see text) to facilitate comparison with Monte Carlo data.

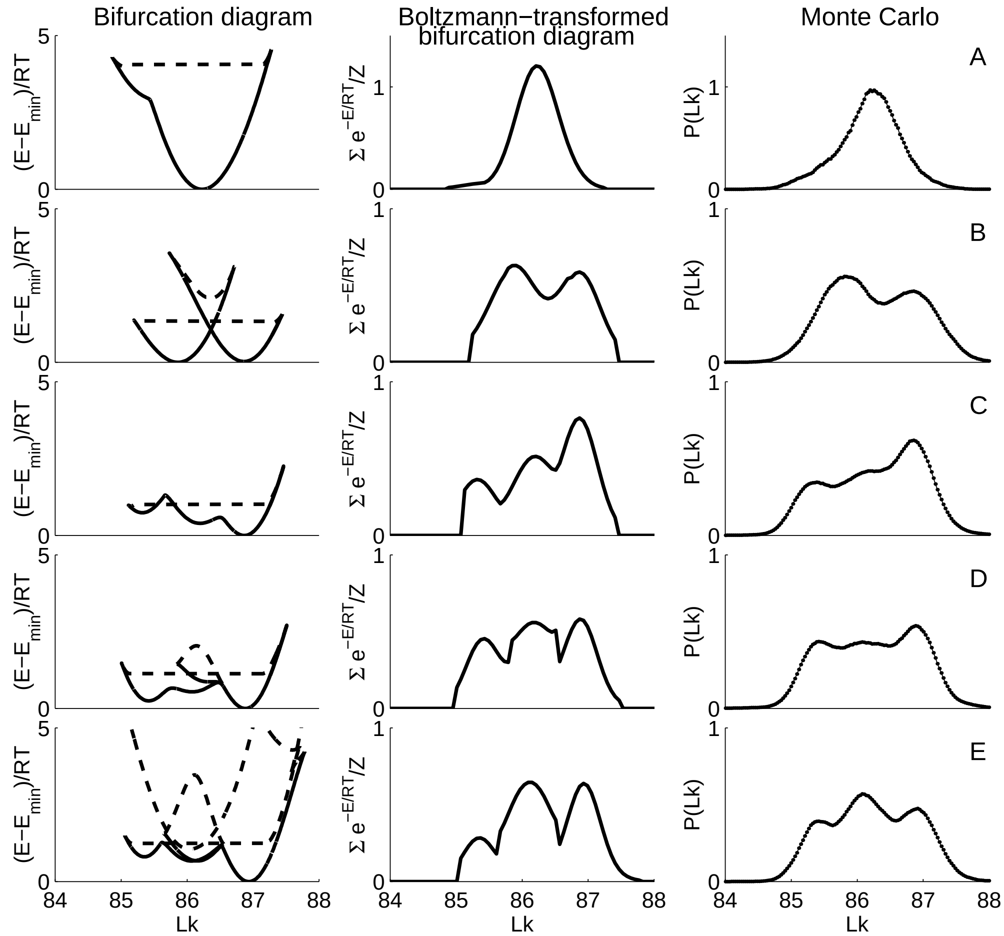
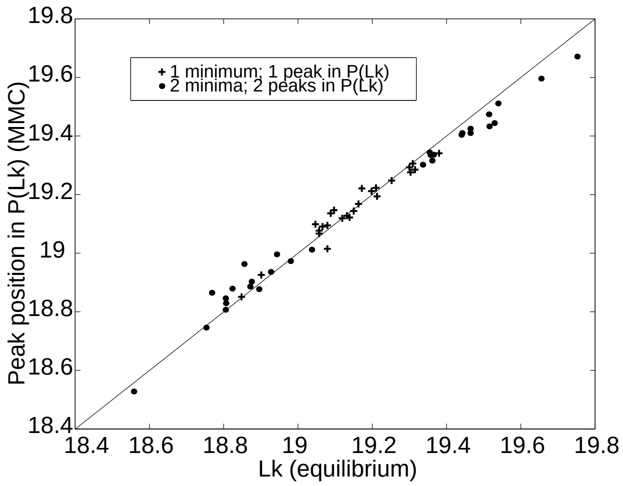


Figure 16: Comparison of bifurcation diagrams and Monte Carlo $P(Lk)$ for five 900-bp DNA A–E (angle set AS1). As in Fig. 15, the number of peaks in $P(Lk)$ is less than or equal to the number of nicked minima. Whereas the molecules B and C show the same number of peaks as nicked minima (2 and 3, respectively), the 4 nicked minima of molecules D and E only give rise to three separated peaks in $P(Lk)$.

A



B

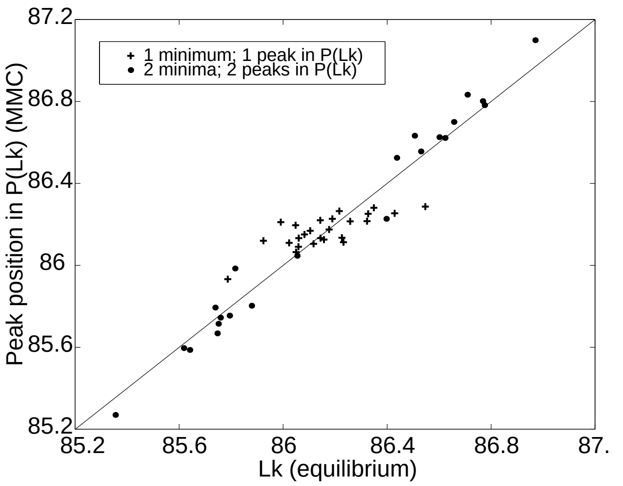


Figure 17: Correlation between Lk values of nicked minima vs peak positions in $P(Lk)$ (as determined by Gaussian fits), for single and double nicked equilibria of either 200-bp (A) or 900-bp (B).

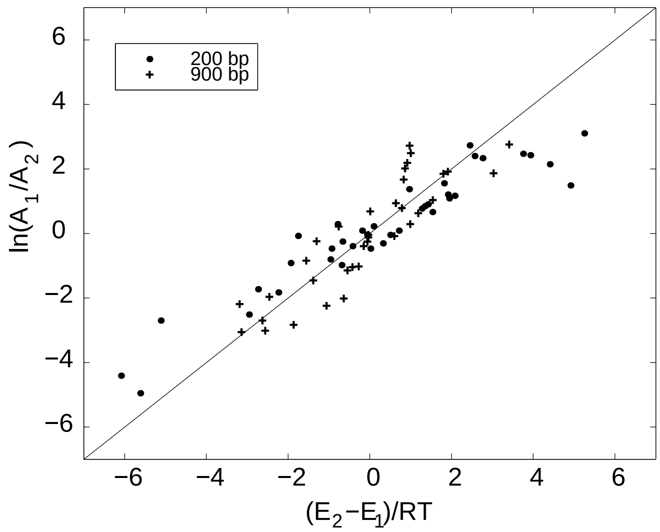


Figure 18: Correlation between the difference in elastic energy of a pair of nicked minima and the ratio of areas of the corresponding Lk peaks. The results support the conjecture that $A_2/A_1 \approx \exp(-(E_2 - E_1)/RT)$.