1996

A Continuum Rod Model of Sequence-Dependent DNA Structure

Robert S. Manning
Haverford College, rmanning@haverford.edu

John H. Maddocks

Jason D. Kahn

A continuum rod model of sequencedependent DNA structure

Robert S. Manning, John H. Maddocks, and Jason D. Kahn

Citation: J. Chem. Phys. **105**, 5626 (1996); doi: 10.1063/1.472373
View online: http://dx.doi.org/10.1063/1.472373
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v105/i13
Published by the American Institute of Physics.

A continuum rod model of sequence-dependent DNA structure

Robert S. Manning[a] and John H. Maddocks[b]
Institute for Physical Science and Technology and Department of Mathematics, University of Maryland, College Park, Maryland 20742

Jason D. Kahn[c]
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

(Received 20 May 1996; accepted 28 June 1996)

Experimentally motivated parameters from a base-pair-level discrete DNA model are averaged to yield parameters for a continuum elastic rod with a curved unstressed shape reflecting the local DNA geometry. The continuum model permits computations with discretization lengths longer than the intrinsic discretization of the base-pair model, and, for this and other reasons, yields an efficient computational formulation. Obtaining continuum stiffnesses is straightforward, but obtaining a continuum unstressed shape is hindered by the ''noisy'' small-scale structure and rapid helix twist of the discrete unstressed shape. Filtering of the discrete data and an analytic transformation from the true normal-vector field to a natural (untwisted) frame allows a stable continuum fit. Equilibrium energies of closed rings predicted by the continuum model are found to match the energies of the underlying discrete model to within 0.5%. The model is applied to a set of 11 short DNA molecules ($\approx$ 150 bp) and properly distinguishes their cyclization probabilities (J factors) when compared both to experimental cyclization rates and to Monte Carlo simulations. The continuum model does not include entropic contributions to the free energy. However, because of its rapid and accurate computation of internal energy, the continuum model should, when combined with further work on entropic effects, be a useful method for computing experimental DNA free energies. © *1996 American Institute of Physics.* [S0021-9606(96)51437-6]

I. INTRODUCTION

Recently, there has been considerable interest in modeling the large-scale deformation of DNA molecules using elastic rod theories (see e.g., recent review articles by Schlick[1] and Olson[2]). However, with some notable exceptions,[3–6] most continuum studies have modeled DNA by a straight uniform rod, which neglects the DNA's intrinsic curvature. Our goal is to develop a model using an intrinsically bent and twisted elastic rod whose curvatures are determined directly from the properties of the underlying base-pair sequence. Sequence-dependent effects have been included in other DNA studies, such as all-atom models[7] and models which treat each base-pair as a rigid unit,[8–10] but these methods can be prohibitively time-consuming for all but the smallest DNA molecules.

A theory of elastic rods encompassing effects of intrinsic curvature is a classic topic of continuum mechanics.[11] Recently, an associated computationally efficient formulation has been developed[12,13] with a particular emphasis on solving the loop boundary-value problem arising in modeling cyclized DNA. The continuum model requires input parameters giving the rod's unstressed shape and stiffnesses, and in this paper, we propose and verify a procedure for determining these parameters from the DNA base-pair sequence.

In Section II we present the discrete wedge-angle model, a widely accepted base-pair level model for DNA, and in Section III we outline the continuum theory of elastic rods. In Section IV, we present the basic formalism for finding equilibria in both the discrete and continuum problems. The determination of continuum stiffnesses is presented in Section V, and the determination of the continuum unstressed shape is presented in Section VI, including a crucial filtration of the discrete centerline and an analytic transformation from the rapidly twisting normal-vector field tracking the sugar–phosphate chains to an untwisted natural frame. In Section VII, we summarize our procedure for determining continuum model parameters from discrete model parameters and subsequently computing equilibrium configurations and energies. In Section VIII we investigate the robustness of these continuum computations. In Section IX, we verify the accuracy of the continuum model by comparing equilibrium energies and configurations of the discrete model and of the corresponding continuum model. Finally, in Section X we apply the continuum rod theory to model the cyclization rates of a family of short DNA molecules and compare cyclization energies to values determined both experimentally and from Monte Carlo simulations. These comparisons demonstrate that for large-scale bending deformations, the continuum model with sequence-dependent structure captures the essential physics of the DNA cyclization.

II. A DISCRETE BASE-PAIR LEVEL MODEL FOR DNA

Each of the two strands of DNA consists of a sugar–phosphate chain and, at regular intervals, side chains called bases, of which there are 4 types: adenine (A), cytosine (C), guanine (G), and thymine (T). The two sugar–phosphate

[a]Electronic mail: rmanning@ipst.umd.edu
[b]Electronic mail: jhm@ipst.umd.edu
[c]Electronic mail: kahn@adnadn.umd.edu

 0021-9606/96/105(13)/5626/21/$10.00

chains wind about each other in a double helix and are bound together by one of two pairings of bases: A with T, or C with G.

We present here one of the commonly used discrete models for DNA, namely the wedge-angle model.[9] We emphasize that the central ideas of this paper do not rely intrinsically on the specific details of this particular choice of discrete model; with slight adaptations, the procedures we propose could be used to obtain continuum versions of other discrete base-pair level models, such as the junction model or trinucleotide models.[14]

The wedge-angle model treats the base-pairs as rigid units, which stack on top of each other according to the atomic geometries and interactions of the base-pairs.[15] Regarded as a rigid body, base-pair i can be completely described by an orthonormal *frame*, i.e., an origin $\mathbf{r}^{(i)}$ and a set of 3 mutually perpendicular unit vectors $(\mathbf{d}_1^{(i)},\mathbf{d}_2^{(i)},\mathbf{d}_3^{(i)})$. The origin $\mathbf{r}^{(i)}$ is located at the center of the base-pair, $\mathbf{d}_3^{(i)}$ points toward the center of the next base-pair, $\mathbf{d}_1^{(i)}$ points to the center of the major groove, and $\mathbf{d}_2^{(i)}$ is determined by mutual perpendicularity $(\mathbf{d}_2^{(i)}=\mathbf{d}_3^{(i)}\times\mathbf{d}_1^{(i)})$; see Figure 1. It will prove convenient to label the first frame by $i=0$ and the last frame by $i=N$.

To determine frame $i+1$ given frame i, one performs a three-dimensional rotation whose representation *in the basis of frame i* is given by an orthogonal 3-by-3 matrix $\mathbf{R}^{(i)}$. For example, if

$$\mathbf{R}^{(i)}=\begin{bmatrix}\cos\tau & -\sin\tau & 0\\ \sin\tau & \cos\tau & 0\\ 0 & 0 & 1\end{bmatrix},$$

then $\mathbf{d}_j^{(i+1)}$, $j=1,2,3$, is obtained by rotating $\mathbf{d}_j^{(i)}$ by an angle τ about $\mathbf{d}_3^{(i)}$ (*not* $\mathbf{e}_3$). The rotation matrix $\mathbf{R}^{(i)}$ is described by 3 Euler angles $(\tau^{(i)},\phi^{(i)},\theta^{(i)})$, which represent successive rotations about $\mathbf{d}_2^{(i)}$, $\mathbf{d}_1^{(i)}$, and $\mathbf{d}_3^{(i)}$. That is, we define

$$\mathbf{R}_\tau(\tau)\equiv\begin{pmatrix}\cos\tau & -\sin\tau & 0\\ \sin\tau & \cos\tau & 0\\ 0 & 0 & 1\end{pmatrix},$$

$$\mathbf{R}_\phi(\phi)\equiv\begin{pmatrix}\cos\phi & 0 & \sin\phi\\ 0 & 1 & 0\\ -\sin\phi & 0 & \cos\phi\end{pmatrix},$$

$$\mathbf{R}_\theta(\theta)\equiv\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{pmatrix},$$

and the ith rotation matrix is given by

$$\mathbf{R}^{(i)}\equiv\mathbf{R}(\theta^{(i)},\phi^{(i)},\tau^{(i)})\equiv\mathbf{R}_\tau(\tau^{(i)})\mathbf{R}_\theta(\theta^{(i)})\mathbf{R}_\phi(\phi^{(i)})$$
$$=\begin{bmatrix}\cos\phi\cos\tau+\sin\theta\sin\phi\sin\tau & -\cos\theta\sin\tau & \sin\phi\cos\tau-\sin\theta\cos\phi\sin\tau\\ \cos\phi\sin\tau-\sin\theta\sin\phi\cos\tau & \cos\theta\cos\tau & \sin\phi\sin\tau+\sin\theta\cos\phi\cos\tau\\ -\cos\theta\sin\phi & -\sin\theta & \cos\theta\cos\phi\end{bmatrix}. \tag{2.1}$$

There are several other definitions of Euler angles in common usage corresponding to different choices and orders of rotation axes.[16] The apparent inconsistency in the definition of $\mathbf{R}_\tau$ as compared to $\mathbf{R}_\theta$ and $\mathbf{R}_\phi$ arises because we want $\tau>0$ to represent a counter-clockwise rotation about $\mathbf{d}_3$ to match the right-handed twist of natural A- and B-form DNA. The change-of-basis formula and a simple recursion show that the coordinates of $(\mathbf{d}_1^{(i)},\mathbf{d}_2^{(i)},\mathbf{d}_3^{(i)})$ in the standard basis (i.e., the laboratory frame) are given by the columns of the matrix $\mathbf{R}^{(0)}\mathbf{R}^{(1)}\ldots\mathbf{R}^{(i-1)}$

$$\begin{aligned}\mathbf{d}_1^{(i)}&=\mathbf{R}^{(0)}\mathbf{R}^{(1)}\ldots\mathbf{R}^{(i-1)}\mathbf{e}_1,\\ \mathbf{d}_2^{(i)}&=\mathbf{R}^{(0)}\mathbf{R}^{(1)}\ldots\mathbf{R}^{(i-1)}\mathbf{e}_2,\\ \mathbf{d}_3^{(i)}&=\mathbf{R}^{(0)}\mathbf{R}^{(1)}\ldots\mathbf{R}^{(i-1)}\mathbf{e}_3.\end{aligned} \tag{2.2}$$

Model parameters. An important set of parameters for this discrete model are the Euler angles $(\hat\theta^{(i)},\hat\phi^{(i)},\hat\tau^{(i)})$ that describe the minimum-energy configuration of the DNA in the absence of external forces. We call this configuration the DNA's *unstressed shape*, and call the corresponding Euler angles the *unstressed angles*. The wedge-angle model assumes that the values of $\hat\theta^{(i)}$, $\hat\phi^{(i)}$, and $\hat\tau^{(i)}$ depend only on the stacking interactions of base-pairs i and $i+1$, and not on any other base-pairs (the nearest-neighbor assumption). There are 4 different base-pairs and hence 16 possible 2-base-pair stacks; thus, the parameters to describe the unstressed shape are 16 triples of angles

$$(\hat\theta,\hat\phi,\hat\tau)_{\mathrm{AT/AT}},\ (\hat\theta,\hat\phi,\hat\tau)_{\mathrm{AT/CG}},\ (\hat\theta,\hat\phi,\hat\tau)_{\mathrm{AT/GC}},\ldots,$$
$$(\hat\theta,\hat\phi,\hat\tau)_{\mathrm{TA/GC}},\ (\hat\theta,\hat\phi,\hat\tau)_{\mathrm{TA/TA}}.$$

However, only 10 of these triples are independent due to symmetry considerations. There are several sets of these unstressed angles used in the literature, determined either by empirical analysis of electrophoretic mobilities for a variety of sequences (e.g., those of Trifonov[9]) or by a similar analysis combined with molecular mechanics (e.g., those of De Santis[10]); see Table I.

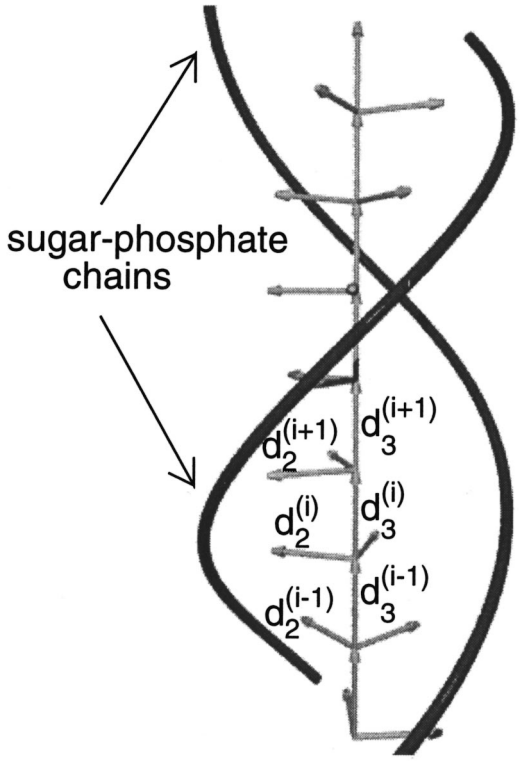

FIG. 1. Discrete model for DNA. Each frame has its origin at the center of a base-pair, its $\mathbf{d}_3$ axis pointing to the next base-pair center, and its $\mathbf{d}_1$ axis pointing to the center of the major groove.

The unstressed angles determine the shape of the minimum-energy configuration of the DNA molecule, but we also need to know the energy penalty incurred by forcing the molecule into another shape. The wedge-angle model assumes that the energy increases quadratically when any of the Euler angles $(\theta^{(i)}, \phi^{(i)}, \tau^{(i)})$ change from their unstressed values $(\hat{\theta}^{(i)}, \hat{\phi}^{(i)}, \hat{\tau}^{(i)})$

$$E = \sum_{i=0}^{N-1} \left[\frac{1}{2} K_\theta (\theta^{(i)} - \hat{\theta}^{(i)})^2 + \frac{1}{2} K_\phi (\phi^{(i)} - \hat{\phi}^{(i)})^2 + \frac{1}{2} K_\tau (\tau^{(i)} - \hat{\tau}^{(i)})^2 \right]. \tag{2.3}$$

Thus, this model also includes as parameters the *stiffnesses* K_θ, K_ϕ, and K_τ. Estimates for these stiffnesses are available from experimental values for the persistence length (specifying K_θ and K_ϕ, from sedimentation, light scattering, and cyclization experiments) and torsional modulus (specifying K_τ, from fluorescence anisotropy and cyclization). Implicit in Eq. (2.3) is the assumption that K_θ, K_ϕ and K_τ are the same for all base-pairs. This restriction is due to a lack of appropriate experimental data to model nonuniform stiffnesses, at least at the time of the development of this wedge-angle model. Given such experimental data, the continuum model described in this paper could easily be extended to include sequence-dependent stiffnesses. In addition, the energy (2.3) is assumed to be quadratic in the angles; if needed, higher order terms could be included, e.g., to incorporate asymmetric effects like DNA's preference to undertwist rather than overtwist.[17]

There is some ambiguity in delimiting the ends of the DNA molecule. In many applications, it is reasonable to declare that the molecule begins at the first base-pair and ends at the last base-pair. However, in cyclization (the focus of

TABLE I. Trifonov (Ref. 9) and De Santis (Ref. 10) unstressed angles for the rotation from base-pair i to base-pair $i+1$. The direction from base-pair i to base-pair $i+1$ will be the 5′-to-3′ direction on one DNA strand and the 3′-to-5′ direction on the other strand. By convention, the name of a base-pair gives the name of the base on the 5′-to-3′ strand followed by the name of its partner on the 3′-to-5′ strand (e.g., GC). Base-pairing and symmetry implies that only 10 of these sets of angles are independent; for example, the relationship between the GC/AT angles and the TA/CG angles is clear from the table.

Base-pair stack	$\hat{\theta}$(Trif)	$\hat{\phi}$(Trif)	$\hat{\tau}$(Trif)	$\hat{\theta}$(DeS)	$\hat{\phi}$(DeS)	$\hat{\tau}$(DeS)
AT/AT	−6.5	3.2	35.6	−5.4	−0.5	35.9
AT/CG	−0.9	−0.7	34.4	−2.5	−2.7	34.6
AT/GC	8.4	−0.3	27.7	−1.0	−1.6	35.6
AT/TA	2.6	0.0	31.5	−7.3	0.0	35.0
CG/AT	1.6	3.1	34.5	6.8	0.4	34.5
CG/CG	1.2	1.8	33.7	1.3	0.6	33.0
CG/GC	6.7	0.0	29.8	4.6	0.0	33.7
CG/TA	8.4	0.3	27.7	−1.0	1.6	35.6
GC/AT	−2.7	−4.6	36.9	2.0	−1.7	35.8
GC/CG	−5.0	0.0	40.0	−3.7	0.0	33.3
GC/GC	1.2	−1.8	33.7	1.3	−0.6	33.0
GC/TA	−0.9	0.7	34.4	−2.5	2.7	34.6
TA/AT	0.9	0.0	36.0	8.0	0.0	34.6
TA/CG	−2.7	4.6	36.9	2.0	1.7	35.8
TA/GC	1.6	−3.1	34.5	6.8	−0.4	34.5
TA/TA	−6.5	−3.2	35.6	−5.4	0.5	35.9

this paper), the two ends of a DNA molecule are joined with the usual inter-base-pair spacing between the last and first base-pairs. Accordingly, to capture the geometry of the last junction, we append a fictitious base-pair, and declare it to be the same type (A, C, G, or T) as the first base-pair (since the first base-pair will eventually follow the last in the cyclized DNA). This technicality is not a major change to the model, but is worth including for the short DNA molecules considered in Section X. We adopt the convention that the molecule begins at the origin of the first frame ($i=0$) and ends at the origin of the last frame ($i=N$). For cyclization, there are N base-pairs in the DNA molecule, because the last frame is fictitious.

III. THE ELASTIC ROD MODEL

The special Cosserat theory of elastic rods is a standard model in continuum mechanics.[11] We restrict attention to inextensible and unshearable rods, which are approximations analogous to the assumption in the wedge-angle model that each base-pair lies at the tip of the $\mathbf{d}_3$ axis of the previous base-pair. Extensibility and shearability can easily be included in the continuum model if needed.[11,12] However, we note that the cyclized DNA configurations described in Section X involve forces of 0.5–3.0 pN, which are much smaller than the reported threshold of approximately 60 pN[18,19] required to cause significant DNA extension. The configuration of an inextensible and unshearable rod is given by a continuous family of frames (mutually perpendicular unit vectors) $(\mathbf{d}_1(s),\mathbf{d}_2(s),\mathbf{d}_3(s))$, with the origins of the frames lying on the rod's centerline and $\mathbf{d}_3(s)$ given by the centerline's unit tangent vector. It is convenient to let s be the arc-length parameter of the centerline and to choose a length-scale so that $0\leqslant s\leqslant 1$. Under these assumptions, the centerline $\mathbf{r}(s)$ can be recovered from the frames using the relation

$$\frac{d\mathbf{r}}{ds}=\mathbf{d}_3(s). \tag{3.1}$$

The interpretations of the centerline and frames in modeling a DNA molecule are just as in the discrete model. The centerline $\mathbf{r}(s)$ runs through the middle of the double helix, passing through (or at least near) the centers of the base-pairs. The normal vector $\mathbf{d}_1(s)$ points from the centerline to the center of the major groove, thus tracking the helix twist.

It is easily verified that given a continuous family of (orthonormal) frames $(\mathbf{d}_1(s),\mathbf{d}_2(s),\mathbf{d}_3(s))$, there exists a 3-vector function $\mathbf{u}(s)$ such that

$$\mathbf{d}_i'(s)\equiv\frac{d}{ds}\mathbf{d}_i(s)=\mathbf{u}(s)\times\mathbf{d}_i(s),\quad i=1,2,3. \tag{3.2}$$

By convention, we define u_i to be the components of $\mathbf{u}$ in the director frame:

$$\mathbf{u}(s)=u_1(s)\mathbf{d}_1(s)+u_2(s)\mathbf{d}_2(s)+u_3(s)\mathbf{d}_3(s). \tag{3.3}$$

Inserting Eq. (3.3) into Eq. (3.2) leads to

$$\begin{aligned} u_1(s)&=-\mathbf{d}_3'(s)\cdot\mathbf{d}_2(s),\\ u_2(s)&=\mathbf{d}_3'(s)\cdot\mathbf{d}_1(s),\\ u_3(s)&=\mathbf{d}_1'(s)\cdot\mathbf{d}_2(s). \end{aligned} \tag{3.4}$$

Given $u_i(s)$ and the initial frame $(\mathbf{d}_1(0),\mathbf{d}_2(0),\mathbf{d}_3(0))$, one can solve the first-order differential equations (3.2) for the frames at all s, and then recover the centerline $\mathbf{r}(s)$ using Eq. (3.1).

Model parameters. As in the discrete model, one important input to the continuum model is a description of the unstressed shape of the rod, given by three functions $\hat{u}_1(s)$, $\hat{u}_2(s)$, and $\hat{u}_3(s)$. By standard convention in rod theory, the shape descriptors u_i are called *strains*, even when they refer to the unstressed shape. We will call the quantities $\hat{u}_i$ *unstressed strains*.

In addition to describing the unstressed shape of the rod, we need to quantify the energy penalty in deviating from that unstressed shape. We assume that the energy is quadratic in the strains

$$\begin{aligned} E=\int_0^1\Big[&\frac{1}{2}K_1(u_1(s)-\hat{u}_1(s))^2+\frac{1}{2}K_2(u_2(s)-\hat{u}_2(s))^2\\ &+\frac{1}{2}K_3(u_3(s)-\hat{u}_3(s))^2\Big]ds\equiv\int_0^1 L\ ds, \end{aligned} \tag{3.5}$$

where the integrand, i.e., the Lagrangian, is denoted by L. Hence the stiffnesses (K_1,K_2,K_3) are parameters for the continuum model. Note that as in the discrete case, we could extend this model to include s-dependence in (K_1,K_2,K_3) or terms of higher order than quadratic if required.

In summary, the continuum model requires the parameters $(\hat{u}_1(s),\hat{u}_2(s),\hat{u}_3(s))$ and (K_1,K_2,K_3) as inputs. In Sections V and VI, we present a method for determining these parameters from the wedge-angle parameters $(\hat{\theta}^{(i)},\hat{\phi}^{(i)},\hat{\tau}^{(i)})$ and (K_θ,K_ϕ,K_τ).

IV. DISCRETE AND CONTINUOUS EQUILIBRIUM CONDITIONS

We outline the conditions for static equilibrium of the discrete and continuous models. For the sake of concreteness, we focus on equilibria of twisted closed rings, which is the context of the cyclization application presented in Section X; see Figure 2. Other boundary value problems could be handled similarly.

A. The discrete rod

In the wedge-angle model for DNA described in Section II, each base-pair is represented by a frame, and the rotation from frame i to frame $i+1$ is given by three Euler angles $(\tau^{(i)},\phi^{(i)},\theta^{(i)})$, with unstressed values determined according to the base-pairs in positions i and $i+1$. Without loss of generality, coordinates may be chosen so that frame 0 has origin $\mathbf{r}=\mathbf{0}$, and directors $(\mathbf{d}_1^{(0)},\mathbf{d}_2^{(0)},\mathbf{d}_3^{(0)})$ equal to the standard coordinate axes $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$.

Cyclization requires that the origin of the final frame also lie at the origin

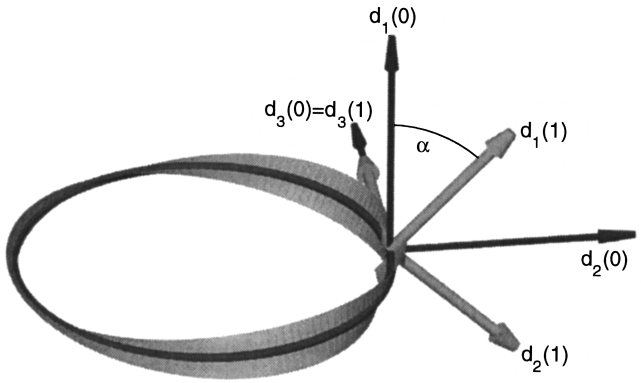

FIG. 2. Ring boundary conditions for a rod. The rod's centerline is indicated with a tube, and the normal vector $\mathbf{d}_1$ is indicated as a ribbon. At $s=0$, the frame has origin (0,0,0) and is oriented along the standard axes. At $s=1$, the frame has origin (0,0,0) and is rotated about the z axis by a prescribed angle α. The $s=0$ frame is drawn darker and at a larger scale in order to distinguish it from the $s=1$ frame.

$$\mathbf{r}^{(N)}=\mathbf{0}. \tag{4.1}$$

We have seen in Section II that each frame's origin lies at the tip of the previous frame's $\mathbf{d}_3$ vector, so Eq. (4.1) can be written as

$$\mathbf{d}_3^{(0)}+\mathbf{d}_3^{(1)}+\ldots+\mathbf{d}_3^{(N-1)}=\mathbf{0}.$$

Using Eq. (2.2), this becomes

$$(\mathbf{I}+\mathbf{R}^{(0)}+\mathbf{R}^{(0)}\mathbf{R}^{(1)}+\ldots+\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-2)})\mathbf{e}_3=\mathbf{0}. \tag{4.2}$$

For convenience, we label this constraint function by $\mathbf{g}$:

$$\mathbf{g}(\theta^{(0)},\phi^{(0)},\tau^{(0)},\ldots,\theta^{(N-1)},\phi^{(N-1)},\tau^{(N-1)})$$
$$\equiv(\mathbf{I}+\mathbf{R}^{(0)}+\mathbf{R}^{(0)}\mathbf{R}^{(1)}+\ldots+\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-2)})\mathbf{e}_3.$$

Further constraints prescribe the orientation of frame N

$$\mathbf{d}_1^{(N)}=\begin{bmatrix}\cos\alpha\\ \sin\alpha\\ 0\end{bmatrix},\quad \mathbf{d}_2^{(N)}=\begin{bmatrix}-\sin\alpha\\ \cos\alpha\\ 0\end{bmatrix},\quad \mathbf{d}_3^{(N)}=\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}, \tag{4.3}$$

which requires that the initial and final tangent vectors $\mathbf{d}_3$ line up and that there be an imposed angle α between the initial and final normal vectors $\mathbf{d}_1$; see Figure 2. Note that only $\alpha=2n\pi$, $n\in\mathbb{Z}$ gives continuity of the sugar–phosphate backbone, but we consider a more general configuration for computational convenience. Using Eq. (2.2), we can rewrite Eq. (4.3) as

$$\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}=\begin{bmatrix}\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \cos\alpha & 0\\ 0 & 0 & 1\end{bmatrix}, \tag{4.4}$$

which appears to entail 9 constraints. In fact, since $\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}$ is *a priori* a member of $SO(3)$ (the three-dimensional group of proper orthogonal rotations), these comprise only 3 *independent* constraints; e.g., it is an easy exercise to show that the 3 constraints which we call h_1, h_2, and h_3

$$h_1(\theta^{(0)},\phi^{(0)},\tau^{(0)},\ldots,\theta^{(N-1)},\phi^{(N-1)},\tau^{(N-1)})$$
$$\equiv\mathbf{e}_3^T\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}\mathbf{e}_2=0,$$
$$h_2(\theta^{(0)},\phi^{(0)},\tau^{(0)},\ldots,\theta^{(N-1)},\phi^{(N-1)},\tau^{(N-1)})$$
$$\equiv\mathbf{e}_1^T\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}\mathbf{e}_3=0, \tag{4.5}$$
$$h_3(\theta^{(0)},\phi^{(0)},\tau^{(0)},\ldots,\theta^{(N-1)},\phi^{(N-1)},\tau^{(N-1)})$$
$$\equiv\mathbf{e}_2^T\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}\mathbf{e}_1=\sin\alpha,$$

imply Eq. (4.4) given that we know $\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}\in SO(3)$. [Actually, Eq. (4.5) plus the condition that $\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}\in SO(3)$ implies 8 possible values for $\mathbf{R}^{(0)}\mathbf{R}^{(1)}\cdots\mathbf{R}^{(N-1)}$, namely

$$\begin{bmatrix}\pm\cos\alpha & -\sin\alpha & 0\\ \sin\alpha & \pm\cos\alpha & 0\\ 0 & 0 & 1\end{bmatrix},$$
$$\begin{bmatrix}\pm\cos\alpha & \sin\alpha & 0\\ \sin\alpha & \mp\cos\alpha & 0\\ 0 & 0 & -1\end{bmatrix},$$
$$\begin{bmatrix}0 & 1 & 0\\ \sin\alpha & 0 & \pm\cos\alpha\\ \pm\cos\alpha & 0 & -\sin\alpha\end{bmatrix},$$
$$\begin{bmatrix}0 & -1 & 0\\ \sin\alpha & 0 & \mp\cos\alpha\\ \pm\cos\alpha & 0 & \sin\alpha\end{bmatrix}.$$

However, these 8 elements are isolated in $SO(3)$, so in our search for discrete rod equilibria, as long as we have an initial guess which nearly satisfies Eq. (4.4), our computed equilibrium using the constraint (4.5) will in fact satisfy Eq. (4.4) and not one of the other seven possibilities.]

The discrete cyclization problem seeks critical points of the energy (2.3)

$$E=\sum_{i=0}^{N-1}\left[\frac{1}{2}K_\tau(\tau^{(i)}-\hat{\tau}^{(i)})^2+\frac{1}{2}K_\phi(\phi^{(i)}-\hat{\phi}^{(i)})^2+\frac{1}{2}K_\theta(\theta^{(i)}-\hat{\theta}^{(i)})^2\right]$$

subject to the position constraints (4.2) and the orientation constraints (4.5). Thus, critical points of a function of $3N$ variables $(\tau^{(i)},\phi^{(i)},\theta^{(i)})$ subject to 6 constraints must be found. Although the function E is simple, the constraints are quite complicated, since each $\mathbf{R}^{(i)}$ is a matrix with trigonometric entries depending on $(\tau^{(i)},\phi^{(i)},\theta^{(i)})$ (see Eq. (2.1)). Incorporation of the constraints in the standard way with Lagrange multipliers $\mathbf{n}$ and $\boldsymbol{\nu}$ leads to the nonlinear system

$$\nabla(E+\mathbf{n}\cdot\mathbf{g}+\boldsymbol{\nu}\cdot\mathbf{h})=\mathbf{0},$$

$$\mathbf{g}=\mathbf{0}, \tag{4.6}$$

$$\mathbf{h}=\langle 0,0,\sin\alpha\rangle,$$

which involves $3N+6$ equations for the $3N+6$ unknowns $\mathbf{n}$, $\boldsymbol{\nu}$, and $\{\theta^{(i)},\phi^{(i)},\tau^{(i)}\}$, for $i=0,\ldots,N-1$. The symbol ∇ denotes the gradient with respect to the angles $\{\theta^{(i)},\phi^{(i)},\tau^{(i)}\}$. The only closed-form solution we know for this system is the planar untwisted N-gon, e.g.,

$$\theta^{(i)}=2\pi/N,\quad \phi^{(i)}=\tau^{(i)}=0,\quad \nu_1=2\pi K_\theta/N,$$

$$\nu_2=\nu_3=n_1=n_2=n_3=0, \tag{4.7}$$

which satisfies Eq. (4.6) when $\alpha=0$ and $\hat{\theta}^{(i)}=\hat{\phi}^{(i)}=\hat{\tau}^{(i)}=0$; other solutions must be determined numerically. For realistic values of $(\hat{\theta}^{(i)},\hat{\phi}^{(i)},\hat{\tau}^{(i)})$ for DNA, we have been unable to numerically determine solutions to Eq. (4.6) except by using initial guesses derived from continuum solutions; Section IX discusses this issue in more detail.

B. The continuous rod

In this section, we outline a formulation of the equilibrium conditions of continuum rods, which leads to the boundary value problem (or BVP) consisting of the differential equations (4.9) and boundary conditions (4.11). This section defines the notation in the BVP and outlines its derivation. However, none of these details are required to understand the remainder of the article; one can proceed directly to the summary paragraph at the end of this section if desired. On the other hand, a more complete explanation of the equilibrium formulation can be found in Li and Maddocks.[13]

As described in Section III, a continuum (inextensible, unshearable) rod is described by a continuous family of frames $(\mathbf{d}_1(s),\mathbf{d}_2(s),\mathbf{d}_3(s))$. Each frame is an element of $SO(3)$, but rather than represent it by 3 Euler angles (as is done with the local rotations in the discrete model), it will be convenient instead to represent it by a 4-vector of Euler parameters $\mathbf{q}(s)$, which must obey $|\mathbf{q}|=1$. The Euler parameters provide the frame through the relations

$$\mathbf{d}_1=\begin{bmatrix} q_1^2-q_2^2-q_3^2+q_4^2\\ 2q_1q_2+2q_3q_4\\ 2q_1q_3-2q_2q_4\end{bmatrix},$$

$$\mathbf{d}_2=\begin{bmatrix} 2q_1q_2-2q_3q_4\\ -q_1^2+q_2^2-q_3^2+q_4^2\\ 2q_2q_3+2q_1q_4\end{bmatrix},$$

$$\mathbf{d}_3=\begin{bmatrix} 2q_1q_3+2q_2q_4\\ 2q_2q_3-2q_1q_4\\ -q_1^2-q_2^2+q_3^2+q_4^2\end{bmatrix}. \tag{4.8}$$

The strains $u_1(s)$, $u_2(s)$, and $u_3(s)$ can be expressed as functions of $\mathbf{q}(s)$ and $\mathbf{q}'(s)$. Thus, the Lagrangian in Eq. (3.5) is a function of $(\mathbf{q}(s),\mathbf{q}'(s),s)$, so a classic calculus of variations problem arises, namely to find critical points of the energy (3.5)

$$E=\int_0^1 L(\mathbf{q}(s),\mathbf{q}'(s),s)ds$$

subject to the pointwise constraints $\mathbf{r}'(s)=\mathbf{d}_3(\mathbf{q}(s))$ and $|\mathbf{q}(s)|=1$, and boundary conditions on $(\mathbf{r},\mathbf{q})$ at $s=0$ and $s=1$ (to be specified later in this section). The first-order stationarity conditions for this constrained Lagrangian system can be transformed to an unconstrained Hamiltonian system with Hamiltonian

$$H(\mathbf{r},\mathbf{q},\mathbf{n},\boldsymbol{\mu})=\sum_{j=1}^{3}\left(\frac{(\boldsymbol{\mu}^T\mathbf{B_j}\mathbf{q})^2}{8K_j}+\frac{\hat{u}_j}{2}\boldsymbol{\mu}^T\mathbf{B_j}\mathbf{q}\right)+\mathbf{n}^T\mathbf{d}_3(\mathbf{q}).$$

Here $\mathbf{n}$ (the force acting across a rod cross-section) is the 3-vector conjugate to $\mathbf{r}$, $\boldsymbol{\mu}$ is the 4-vector conjugate to $\mathbf{q}$, and the $\mathbf{B}_j$ are

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

$$\mathbf{B}_3=\begin{bmatrix}0&1&0&0\\-1&0&0&0\\0&0&0&1\\0&0&-1&0\end{bmatrix}.$$

Determining $(\mathbf{r}(s),\mathbf{q}(s),\mathbf{n}(s),\boldsymbol{\mu}(s))$ involves solving the standard Hamiltonian differential equations

$$\begin{aligned}
\mathbf{r}'&=\frac{\partial H}{\partial \mathbf{n}}=\mathbf{d}_3(\mathbf{q}),\\
\mathbf{q}'&=\frac{\partial H}{\partial \boldsymbol{\mu}}=\sum_{j=1}^{3}\left(\frac{\boldsymbol{\mu}^T\mathbf{B_j}\mathbf{q}}{2K_j}+\hat{u}_j\right)\frac{1}{2}\mathbf{B}_j\mathbf{q},\\
\mathbf{n}'&=-\frac{\partial H}{\partial \mathbf{r}}=\mathbf{0},\\
\boldsymbol{\mu}'&=-\frac{\partial H}{\partial \mathbf{q}}=-\left(\frac{\partial \mathbf{d}_3}{\partial \mathbf{q}}\right)^T\mathbf{n}+\sum_{j=1}^{3}\left(\frac{\boldsymbol{\mu}^T\mathbf{B_j}\mathbf{q}}{2K_j}+\hat{u}_j\right)\frac{1}{2}\mathbf{B}_j\boldsymbol{\mu},
\end{aligned} \tag{4.9}$$

where $\partial\mathbf{d}_3/\partial\mathbf{q}$ is found from (4.8)

$$\frac{\partial \mathbf{d}_3}{\partial \mathbf{q}}=\begin{bmatrix}2q_3&2q_4&2q_1&2q_2\\-2q_4&2q_3&2q_2&-2q_1\\-2q_1&-2q_2&2q_3&2q_4\end{bmatrix}.$$

So, to find the stationary points for the continuum energy functional (3.5), we solve the first-order system of differential equations (4.9) subject to appropriate boundary conditions. The boundary conditions are the continuous analogue of those used in the discrete problem

$$\mathbf{r}(0)=\mathbf{r}(1)=\mathbf{0},\quad \mathbf{d}_1(0)=\mathbf{e}_1,\quad \mathbf{d}_2(0)=\mathbf{e}_2,\quad \mathbf{d}_3(0)=\mathbf{e}_3,$$

$$\mathbf{d}_1(1)=\langle\cos\alpha,\sin\alpha,0\rangle,$$

$$\mathbf{d}_2(1)=\langle -\sin\,\alpha,\,\cos\,\alpha,0\rangle,\quad \mathbf{d}_3(1)=\langle 0,0,1\rangle,$$

which can be translated, using Eq. (4.8), into variables appearing in the Hamiltonian

$$\mathbf{r}(0)=\mathbf{r}(1)=\mathbf{0},\quad \mathbf{q}(0)=\langle 0,0,0,1\rangle,$$
$$\mathbf{q}(1)=\langle 0,0,-\sin(\alpha/2),-\cos(\alpha/2)\rangle. \tag{4.10}$$

Unfortunately, the BVP (4.9)+(4.10) has non-isolated solutions, which prohibits numerical solution of the BVP via continuation methods. [A solution $(\mathbf{r},\mathbf{q},\mathbf{n},\boldsymbol{\mu})$ is non-isolated if for the same parameter values there are other solutions arbitrarily close.] In particular, the BVP (4.9)+(4.10) is invariant to the symmetry $\boldsymbol{\mu}\rightarrow\boldsymbol{\mu}+\epsilon\mathbf{q}$. This symmetry is a consequence of using four Euler parameters to describe the locally three-dimensional group $SO(3)$. To remove this non-isolation, we replace the conditions (4.10) by equivalent boundary conditions which factor out the symmetry: in particular, the condition $q_4(0)=1$ (which is implied by the other seven boundary conditions on $\mathbf{q}$ and the fact that $|\mathbf{q}|^2$ is an integral of the Hamiltonian system) can be replaced by the condition $\mu_4(0)=0$, which selects a single representative of the family of possible $\boldsymbol{\mu}$ values. Accordingly, the boundary conditions used are

$$\begin{aligned}&\mathbf{r}(0)=\langle 0,0,0\rangle,\\ &\mathbf{r}(1)=\langle 0,0,0\rangle,\\ &q_1(0)=q_2(0)=q_3(0)=0,\\ &\mu_4(0)=0,\\ &\mathbf{q}(1)=\langle 0,0,-\sin(\alpha/2),-\cos(\alpha/2)\rangle.\end{aligned} \tag{4.11}$$

In summary, the continuum equilibrium condition for a twisted ring is the BVP consisting of the 14 differential equations (4.9) and the 14 boundary conditions (4.11), where α is the angle between $\mathbf{d}_1(0)$ and $\mathbf{d}_1(1)$. In the special case that $\hat{u}_1(s)=\hat{u}_2(s)=\hat{u}_3(s)\equiv 0$ and $K_1=K_2$ (an intrinsically straight and untwisted rod with equal bending stiffnesses, which we call the *perfect problem*), there is a closed-form solution for this BVP

$$\begin{aligned}&\mathbf{r}(s)=\frac{1}{2\pi}\begin{bmatrix}0\\ \cos(2\pi s)-1\\ \sin(2\pi s)\end{bmatrix},\quad \mathbf{q}(s)=\begin{bmatrix}\sin(\pi s)\cos(\alpha s/2)\\ -\sin(\pi s)\sin(\alpha s/2)\\ \cos(\pi s)\sin(\alpha s/2)\\ \cos(\pi s)\cos(\alpha s/2)\end{bmatrix},\\ &\mathbf{n}(s)=\begin{bmatrix}2\pi K_3\alpha\\ 0\\ 0\end{bmatrix},\quad \boldsymbol{\mu}(s)=\begin{bmatrix}4\pi K_1\cos(\pi s)\cos(\alpha s/2)-2\alpha K_3\sin(\alpha s)\sin(\alpha s/2)\\ -4\pi K_1\cos(\pi s)\sin(\alpha s/2)-2\alpha K_3\sin(\alpha s)\cos(\alpha s/2)\\ -4\pi K_1\sin(\pi s)\sin(\alpha s/2)+2\alpha K_3\cos(\alpha s)\cos(\alpha s/2)\\ -4\pi K_1\sin(\pi s)\cos(\alpha s/2)-2\alpha K_3\cos(\alpha s)\sin(\alpha s/2)\end{bmatrix}.\end{aligned} \tag{4.12}$$

The numerical determination of other solutions of this BVP is discussed in Section VII.

V. DETERMINING THE CONTINUUM STIFFNESSES

To motivate our conversion from angle-stiffnesses (K_θ,K_ϕ,K_τ) to strain-stiffnesses (K_1,K_2,K_3), we consider planar untwisted deformations of an intrinsically straight rod (cf., Weinberger[20]). Without loss of generality, let the planar deformations occur in the $x-z$ plane. Then in the discrete problem we use rotations about the y axis to get from one frame to the next, so $\tau^{(i)}=\theta^{(i)}=0$. The discrete energy is therefore

$$E=\sum_{i=0}^{N-1}\frac{1}{2}K_\phi(\phi^{(i)})^2.$$

We now define $\beta^{(i)}\equiv\Sigma_{j=0}^{i}\phi^{(j)}$ and $\beta^{(-1)}\equiv 0$ so that

$$E=\sum_{i=0}^{N-1}\frac{1}{2}K_\phi(\beta^{(i)}-\beta^{(i-1)})^2. \tag{5.1}$$

In the continuous problem, if we label by $\alpha(s)$ the clockwise angle the rod's tangent makes from vertical, then we have $\mathbf{d}_2=\langle 0,-1,0\rangle$, $\mathbf{d}_3=\langle\sin\alpha,0,\cos\alpha\rangle$, and $\mathbf{d}_1=\langle\cos\alpha,0,-\sin\alpha\rangle$. An easy computation shows that for this case, $u_1=u_3=0$ and $u_2(s)=d\alpha(s)/ds$. So, our continuous energy is

$$E=\int_0^1\frac{1}{2}K_2\left(\frac{d\alpha}{ds}(s)\right)^2 ds.$$

We change variables in the integral to $t\equiv Ns$

$$E=\int_0^N\frac{1}{2}K_2\left(\frac{d\alpha}{ds}\left(\frac{t}{N}\right)\right)^2\frac{dt}{N}.$$

Now define

$$\beta(t)\equiv\alpha\left(\frac{t}{N}\right)$$

for $0\leqslant t\leqslant N$, so that

$$\frac{d\beta}{dt}(t)=\frac{1}{N}\frac{d\alpha}{ds}\left(\frac{t}{N}\right).$$

Then

$$E=\int_0^N \frac{1}{2}K_2\left(N\frac{d\beta}{dt}(t)\right)^2\frac{dt}{N}=\int_0^N \frac{1}{2}NK_2\left(\frac{d\beta}{dt}(t)\right)^2 dt. \tag{5.2}$$

By inspection, it is clear that Eq. (5.1) is a discrete approximation of the integral (5.2) if we take $K_2=K_\phi/N$. We could repeat this process for a rod bent in the $y-z$ plane to find that $K_1=K_\theta/N$, or for a straight rod with twist to find that $K_3=K_\tau/N$.

Motivated by this argument, we adopt the conversion rules

$$K_1=K_\theta/N, \quad K_2=K_\phi/N, \quad K_3=K_\tau/N. \tag{5.3}$$

This should not be taken as a proof of these relationships, because in the full three-dimensional problem, the rotations R_θ, R_ϕ, and R_τ do not commute, so the problem will not decouple into the 3 simple cases discussed above. In DNA, two of the three angles are small, so R_θ, R_ϕ, and R_τ nearly commute, and the scaling is at least plausible. In Section IX, we test Eq. (5.3) by comparing discrete and continuous nonplanar ring equilibria, and we see that the conversion works very well in that context.

VI. DETERMINING THE CONTINUUM UNSTRESSED SHAPE

The wedge-angle model describes a DNA molecule as a sequence of frames, with the origin of frame $i+1$ at the end of $\mathbf{d}_3$ for frame i. Equivalently, we can think of this data as a centerline (the union $\mathbf{d}_3^{(0)}\cup\ldots\cup\mathbf{d}_3^{(N-1)}$) and a sequence of normal vectors (the $\mathbf{d}_1$ axes). We will consider first the conversion of the discrete centerline to a continuum centerline, and then the conversion of the discrete set of normal vectors to a continuous family of normal vectors.

A. Aligning the continuous and discrete centerlines

Though the discrete centerline ends at $\mathbf{r}^{(N)}$, it is useful to consider $\mathbf{r}^{(N+1)}$, the endpoint of $\mathbf{d}_3^{(N)}$. We want to fit a continuous function $\mathbf{r}(t)$ through the sequence of points $\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(N+1)}$, so that $\mathbf{r}(j)$ is a good approximation to $\mathbf{r}^{(j)}$ for $j=0,\cdots,N+1$. In general, however, the direction of $\mathbf{d}_3^{(j)}$ will be better approximated by $\mathbf{r}'(j+\frac{1}{2})$ than by $\mathbf{r}'(j)$; see Figure 3. Accordingly, we declare that the continuous rod begins at $t=\frac{1}{2}$ and ends at $t=N+\frac{1}{2}$, so that the discrete and continuous directors match well at the ends of the rods

$$\mathbf{d}_3^{(0)}\approx\frac{\mathbf{r}'(1/2)}{|\mathbf{r}'(1/2)|}, \quad \mathbf{d}_3^{(N)}\approx\frac{\mathbf{r}'(N+1/2)}{|\mathbf{r}'(N+1/2)|}.$$

These end-point directors figure importantly in the boundary conditions in Section IV. The exact procedure for fitting $\mathbf{r}(t)$ to $\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(N+1)}$ is the focus of Section VI B. Once this fit is done, we scale and reparametrize by arclength to obtain $\mathbf{r}(s)$, $0\leqslant s\leqslant 1$, where $s=0$ corresponds to $t=1/2$ and $s=1$ corresponds to $t=N+1/2$.

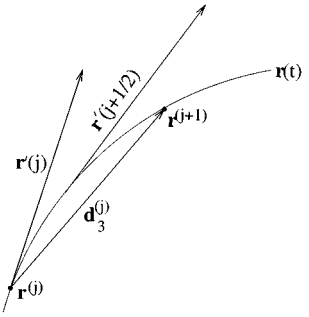

FIG. 3. The discrete model vector $\mathbf{d}_3^{(j)}$ is better approximated by the continuum tangent vector at the midpoint $\mathbf{r}'(j+1/2)$ than by the continuum tangent vector at the endpoint $\mathbf{r}'(j)$.

B. Determining a continuous centerline

Fitting $\mathbf{r}(t)$ to $\mathbf{r}^{(0)},\ldots,\mathbf{r}^{(N+1)}$ is a standard problem, often solved by polynomial interpolation or a least-squares polynomial approximation. Complicating matters, however, is the fact that for DNA models, the discrete set of points have irregular short-scale structure and generally smooth long-scale structure (see Figure 4). We would like to filter the small-scale structure, since it will prevent us from implementing long-scale discretization in the continuum problem (in addition, Figure 4 shows that it is plausible that this short-scale structure is not important for some long-scale properties). Thus, a least-squares approximation is preferable to interpolation.

We take the original t interval $0\leqslant t\leqslant N+1$ and divide it into k subintervals (of approximately equal size) and require that $\mathbf{r}(t)$ be a degree-m polynomial on each subinterval. We also require that $\mathbf{r}(t)$ be C^3 over the whole interval $0\leqslant t\leqslant N+1$, which imposes 4 continuity conditions (continuity of $\mathbf{r}$ and of its first three derivatives) at each point where the subintervals meet. Hence, we have $k(m+1)$ polynomial coefficients subject to $4(k-1)$ constraints. We then minimize

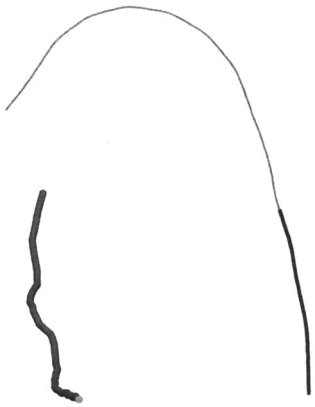

FIG. 4. Long-scale and short-scale structures of a discrete DNA centerline (08A17). The upper right shows a side-view of the entire 157 base-pair molecule with the first 47 base-pairs thickened, and the lower left shows an end-view of these 47 base-pairs. Despite the irregularity in the short-scale structure, the long-scale structure appears quite regular.

$$\sum_{j=0}^{N+1} (\mathbf{r}(j)-\mathbf{r}^{(j)})^2$$

subject to these constraints. This objective function is quadratic in the polynomial coefficients and the constraints are linear in the coefficients, so this least-squares minimization can be reduced to a system of linear equations.

This is our basic strategy for obtaining a continuum centerline. However, in practice we find that this method cannot take DNA data and return a centerline capturing its long-scale features without including its short-scale irregularities. For k (the number of subintervals) or m (the degree of the fitting polynomials) small, the long-scale features are not suitably approximated, and when k and m are increased to the point of capturing these long-scale features, the short-scale irregularities are already present in the approximate centerline. In short, the raw DNA data are sufficiently ''noisy'' that the simple least-squares approximation described above is not stable with respect to changes in k or m.

To forcibly remove the small-scale structure, we apply a filter to the raw data before computing the least-squares approximation. We replace $\mathbf{r}^{(j)}$ by a symmetric weighted average of its neighbors using the FILTFILT function from Matlab's signal processing toolbox[21]

$$\mathbf{r}^{(j)} \rightarrow \frac{\mathbf{r}^{(j-w+1)}}{w^2} + \frac{2\mathbf{r}^{(j-w+2)}}{w^2} + \cdots + \frac{(w-1)\mathbf{r}^{(j-1)}}{w^2} + \frac{w\mathbf{r}^{(j)}}{w^2} + \frac{(w-1)\mathbf{r}^{(j+1)}}{w^2} + \cdots + \frac{2\mathbf{r}^{(j+w-2)}}{w^2} + \frac{\mathbf{r}^{(j+w-1)}}{w^2}.$$

Two more technical details are involved in implementing the filter:

(1) Averaging filters naturally have difficulties at interval end points (since there are fewer neighbors over which to average). In fact, as the above definition is written, the filter is not defined for the $w-1$ entries on each end of the interval. The FILTFILT function actually pads the data on each end with $3(w-1)$ extra points using a reflection method to match the values and slopes of the data at the end points, and then removes these extra points after filtering (see pp. 1–17 of Matlab's Signal Processing manual[21]).

For our study of closed loops of DNA (Section X), we have a more suitable padding method available: we extend each end of the centerline with a piece of the other end of the centerline, its eventual neighbor in the cyclized molecule (we usually use a 50 bp pad on each end). For example, we extend the 150-base-pair DNA molecule $(50\text{bp})_1(50\text{bp})_2(50\text{bp})_3$ to $(50\text{bp})_3(50\text{bp})_1(50\text{bp})_2(50\text{bp})_3(50\text{bp})_1$ and build a padded discrete centerline using the unstressed angles for the padded DNA sequence. We then apply FILTFILT and the least-squares approximation, and then strip the filtered and fitted centerline back down to its central 150 base-pairs.

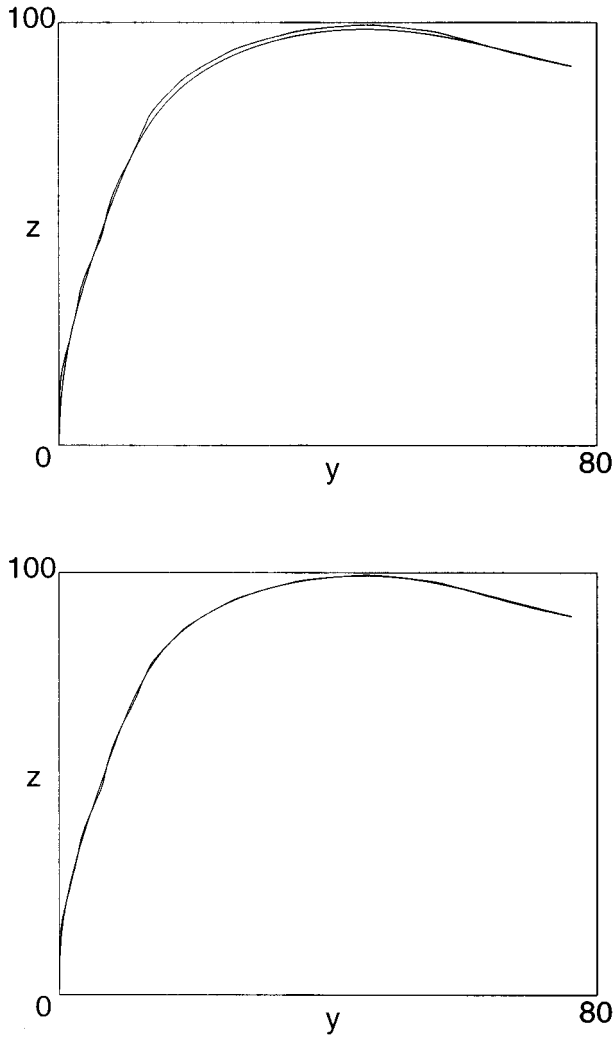

FIG. 5. Double-filtration (cf., Section VI B) of the centerline improves the fit. The top figure shows a data set and its once-filtered output. The bottom figure shows the same data set and its double-filtered output. Note that the straightening effect of the filter is decreased by double filtration. The data shown here are the projection in the y–z plane of the centerline of 08T15 DNA—see Section X.

(2) An averaging filter tends to straighten out data, decreasing long-scale centerline curvature. To remedy this, we apply the filter to compute an approximate averaged centerline $\mathbf{r}_{\text{av}}$ and subtract it from the original data to get a representation of the noise in the raw data $\mathbf{n} \equiv \mathbf{r} - \mathbf{r}_{\text{av}}$. We then apply the filter to $\mathbf{n}$ to get $\mathbf{n}_{\text{av}}$, and then add back in the approximate averaged centerline $\mathbf{r}_{\text{av}}$ to get our final approximation to the centerline. Figure 5 illustrates how this modification improves the quality of the approximation.

C. Determining a continuous field of normal vectors

After we filter the discrete centerline and make a least-squares approximation to find a continuous centerline $\mathbf{r}(t)$, we reparametrize by arclength s; let s_j denote the arclength parameter value corresponding to the value $t=j+1/2$. The tangent vectors $\mathbf{d}_3(s)$ can be computed by the inextensibility-unshearability condition (3.1)

$$\mathbf{d}_3(s)=\frac{d\mathbf{r}(s)}{ds}.$$

Due to the changes made by the filter and least-squares approximation, the original discrete set of normal vectors $\mathbf{d}_1^{(j)}$ will no longer be exactly perpendicular to their corresponding tangent vectors $\mathbf{d}_3(s_j)$, which were the best continuum fit to the original $\mathbf{d}_3^{(j)}$. Accordingly, we first project each $\mathbf{d}_1^{(j)}$ onto the plane perpendicular to $\mathbf{d}_3(s_j)$, and assign the result to be $\mathbf{d}_1(s_j)$. One might worry that this projection introduces significant errors into the model, but in practice, the filter and least-squares approximation change the tangent vectors very little, so the projection step has a minimal effect.

The crucial difficulty, at least in the DNA application, is that the normal vectors rotate rapidly about the centerline, making a full revolution approximately every 10.5 base-pairs. Hence, it is difficult to make an accurate interpolation or least-squares approximation of these normal vectors, since they change appreciably from base-pair to base-pair. Furthermore, when the centerline is curved and the normal vectors $\mathbf{d}_1$ spin rapidly, there will be rapid oscillations in u_1 and u_2. Consider, for example, a centerline $\mathbf{r}(s)=\langle R\cos(s/R),\ R\sin(s/R),0\rangle$ with the corresponding tangent vectors $\mathbf{d}_3(s)=\langle -\sin(s/R),\cos(s/R),0\rangle$, and choose normal vectors $\mathbf{d}_1(s)=\langle\cos(s/R)\cos(\gamma s/R),\sin(s/R)\cos(\gamma s/R),-\sin(\gamma s/R)\rangle$. We can then compute $u_2(s)=\mathbf{d}_3'(s)\cdot\mathbf{d}_1(s)=-\cos(\gamma s/R)/R$, which oscillates rapidly for large γ.

Because $u_1(s)$ and $u_2(s)$ vary rapidly in s, it is desirable to avoid computing with the true DNA frames. Fortunately, we can compute on a different set of frames which are not rapidly rotating, and then recover the true DNA results *analytically* from the transformed results. For each value of the arclength s, we define the transformed directors $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ as a rotation about the $\mathbf{d}_3$ axis of the original directors $\mathbf{d}_1$, $\mathbf{d}_2$, and $\mathbf{d}_3$:

$$\begin{bmatrix}\mathbf{D}_1^T(s)\\ \mathbf{D}_2^T(s)\\ \mathbf{D}_3^T(s)\end{bmatrix}=\begin{bmatrix}\cos(\Omega(s)) & \sin(\Omega(s)) & 0\\ -\sin(\Omega(s)) & \cos(\Omega(s)) & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{d}_1^T(s)\\ \mathbf{d}_2^T(s)\\ \mathbf{d}_3^T(s)\end{bmatrix}$$

$$\equiv\mathbf{M}(s)\begin{bmatrix}\mathbf{d}_1^T(s)\\ \mathbf{d}_2^T(s)\\ \mathbf{d}_3^T(s)\end{bmatrix}. \tag{6.1}$$

Here a superscript T denotes a transpose, so that $\mathbf{d}_i^T$ and $\mathbf{D}_i^T$ are three-dimensional row vectors. The function $\Omega(s)$ gives the angle of rotation (about $\mathbf{d}_3(s)$) to get the transformed frame from the original frame (at each arclength s). The relation (6.1) is easily inverted

$$\begin{bmatrix}\mathbf{d}_1^T(s)\\ \mathbf{d}_2^T(s)\\ \mathbf{d}_3^T(s)\end{bmatrix}=\mathbf{M}^T(s)\begin{bmatrix}\mathbf{D}_1^T(s)\\ \mathbf{D}_2^T(s)\\ \mathbf{D}_3^T(s)\end{bmatrix}.$$

We now determine how the strains transform, i.e., corresponding to the new set of directors $\mathbf{D}_i$, we have a strain vector $\mathbf{w}$ which is defined by

$$\mathbf{D}_i'=\mathbf{w}\times\mathbf{D}_i, \tag{6.2}$$

and we want to relate $\mathbf{w}$ to $\mathbf{u}$. We compute directly

$$\begin{bmatrix}(\mathbf{D}_1^T)'\\ (\mathbf{D}_2^T)'\\ (\mathbf{D}_3^T)'\end{bmatrix}=\Omega'\begin{bmatrix}-\sin\Omega & \cos\Omega & 0\\ -\cos\Omega & -\sin\Omega & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{d}_1^T\\ \mathbf{d}_2^T\\ \mathbf{d}_3^T\end{bmatrix}+\begin{bmatrix}\cos\Omega & \sin\Omega & 0\\ -\sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}(\mathbf{d}_1^T)'\\ (\mathbf{d}_2^T)'\\ (\mathbf{d}_3^T)'\end{bmatrix}=\Omega'\begin{bmatrix}-\sin\Omega & \cos\Omega & 0\\ -\cos\Omega & -\sin\Omega & 0\\ 0 & 0 & 0\end{bmatrix}$$

$$\times\begin{bmatrix}\cos\Omega & -\sin\Omega & 0\\ \sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{D}_1^T\\ \mathbf{D}_2^T\\ \mathbf{D}_3^T\end{bmatrix}+\begin{bmatrix}\cos\Omega & \sin\Omega & 0\\ -\sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 & u_3 & -u_2\\ -u_3 & 0 & u_1\\ u_2 & -u_1 & 0\end{bmatrix}\begin{bmatrix}\mathbf{d}_1^T\\ \mathbf{d}_2^T\\ \mathbf{d}_3^T\end{bmatrix}$$

$$=\begin{bmatrix}0 & \Omega' & 0\\ -\Omega' & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbf{D}_1^T\\ \mathbf{D}_2^T\\ \mathbf{D}_3^T\end{bmatrix}+\begin{bmatrix}\cos\Omega & \sin\Omega & 0\\ -\sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 & u_3 & -u_2\\ -u_3 & 0 & u_1\\ u_2 & -u_1 & 0\end{bmatrix}\begin{bmatrix}\cos\Omega & -\sin\Omega & 0\\ \sin\Omega & \cos\Omega & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{D}_1^T\\ \mathbf{D}_2^T\\ \mathbf{D}_3^T\end{bmatrix}$$

$$=\begin{bmatrix}0 & u_3+\Omega' & u_1\sin\Omega-u_2\cos\Omega\\ -u_3-\Omega' & 0 & u_1\cos\Omega+u_2\sin\Omega\\ -u_1\sin\Omega+u_2\cos\Omega & -u_1\cos\Omega-u_2\sin\Omega & 0\end{bmatrix}\begin{bmatrix}\mathbf{D}_1^T\\ \mathbf{D}_2^T.\\ \mathbf{D}_3^T\end{bmatrix} \tag{6.3}$$

Comparing to Eq. (6.2), we see that

$$w_1 = u_1 \cos\Omega + u_2 \sin\Omega,$$

$$w_2 = -u_1 \sin\Omega + u_2 \cos\Omega, \quad w_3 = u_3 + \Omega',$$

or

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{M} \begin{bmatrix} u_1 \\ u_2 \\ u_3 + \Omega' \end{bmatrix},$$

which implies

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \mathbf{M}^T \begin{bmatrix} w_1 \\ w_2 \\ w_3 - \Omega' \end{bmatrix}.$$

Similarly, the unstressed strains transform as

$$\begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 \end{bmatrix} = \mathbf{M} \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 + \Omega' \end{bmatrix}, \quad \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = \mathbf{M}^T \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \hat{w}_3 - \Omega' \end{bmatrix}.$$

In the case that $K_2 = K_1$, this transformation of directors preserves the form of the Lagrangian, which is critical for the computations described in Section IV

$$\begin{aligned} L &= \frac{K_1}{2}(u_1 - \hat{u}_1)^2 + \frac{K_1}{2}(u_2 - \hat{u}_2)^2 + \frac{K_3}{2}(u_3 - \hat{u}_3)^2 \\ &= \frac{K_1}{2} \begin{bmatrix} u_1 - \hat{u}_1 \\ u_2 - \hat{u}_2 \\ u_3 - \hat{u}_3 \end{bmatrix} \cdot \begin{bmatrix} u_1 - \hat{u}_1 \\ u_2 - \hat{u}_2 \\ u_3 - \hat{u}_3 \end{bmatrix} + \frac{K_3 - K_1}{2}(u_3 - \hat{u}_3)^2 \\ &= \frac{K_1}{2} \left(\mathbf{M}^T \begin{bmatrix} w_1 - \hat{w}_1 \\ w_2 - \hat{w}_2 \\ w_3 - \hat{w}_3 \end{bmatrix} \right) \cdot \left(\mathbf{M}^T \begin{bmatrix} w_1 - \hat{w}_1 \\ w_2 - \hat{w}_2 \\ w_3 - \hat{w}_3 \end{bmatrix} \right) \\ &\quad + \frac{K_3 - K_1}{2}(w_3 - \Omega' - (\hat{w}_3 - \Omega'))^2 \\ &= \frac{K_1}{2} \begin{bmatrix} w_1 - \hat{w}_1 \\ w_2 - \hat{w}_2 \\ w_3 - \hat{w}_3 \end{bmatrix} \cdot \begin{bmatrix} w_1 - \hat{w}_1 \\ w_2 - \hat{w}_2 \\ w_3 - \hat{w}_3 \end{bmatrix} \\ &\quad + \frac{K_3 - K_1}{2}(w_3 - \hat{w}_3)^2 \quad (\text{since } \mathbf{M}\mathbf{M}^T = \mathbf{I}) \\ &= \frac{K_1}{2}(w_1 - \hat{w}_1)^2 + \frac{K_1}{2}(w_2 - \hat{w}_2)^2 + \frac{K_3}{2}(w_3 - \hat{w}_3)^2. \end{aligned} \tag{6.4}$$

For the discrete data we are using, $K_\theta = K_\phi$, which implies that $K_2 = K_1$. Even if $K_2 \neq K_1$, it is possible that the rapid twist of the normal vectors will average the bend stiffnesses and give an effective isotropic rod, with $(K_2)_{\text{eff}} = (K_1)_{\text{eff}}$. Rigorous justification of this claim is still an open question, although numerical evidence suggests it to be true (see the Web site http://www.lcvm.umd.edu/~kehrbaum/research.html). In fact, it is likely in real DNA that $K_\theta \neq K_\phi$,[15] so that the experimentally determined $K_\theta = K_\phi$ in the discrete model really are the effective isotropic stiffnesses of the rapidly twisting DNA.

Since the Lagrangian has the same form for the transformed directors as for the original directors, the Hamiltonian also maintains the same form, so our computations will involve solving the same system of differential equations (4.9). The only change is in the boundary conditions. Specifically, the angle between the transformed directors $\mathbf{D}_1(0)$ and $\mathbf{D}_1(1)$, is not the same as the angle between the original directors $\mathbf{d}_1(0)$ and $\mathbf{d}_1(1)$. By Eq. (6.1), $\mathbf{D}_1(0)$ is a right-handed rotation of $\mathbf{d}_1(0)$ by an angle $\Omega(0)$ about $\mathbf{d}_3(0) = \mathbf{e}_3$. Similarly, $\mathbf{D}_1(1)$ is a rotation of $\mathbf{d}_1(1)$ by an angle $\Omega(1)$ about $\mathbf{d}_3(1) = \mathbf{e}_3$. Also, the boundary conditions (4.11) imply that $\mathbf{D}_1(1)$ is a right-handed rotation of $\mathbf{D}_1(0)$ by an angle α about $\mathbf{e}_3$. Together, these imply that the angle between $\mathbf{d}_1(0)$ and $\mathbf{d}_1(1)$ is $\alpha - (\Omega(1) - \Omega(0))$. If we want the original frame to be closed, i.e., $\mathbf{d}_1(0) = \mathbf{d}_1(1)$, we need $\alpha = \Omega(1) - \Omega(0) \pmod{2\pi}$.

D. The natural frame

We choose the transformed directors to be those of the *natural frame*.[22] Given a centerline $\mathbf{r}(s)$, the natural frame is an untwisted frame $(\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$ defined by a choice of $\mathbf{D}_1(0)$ and the condition

$$\hat{w}_3 = 0. \tag{6.5}$$

Equation (6.5), combined with Eqs. (3.2), (3.3), and (3.4), implies that

$$\begin{aligned} \mathbf{D}_1' &= (\hat{w}_1 \mathbf{D}_1 + \hat{w}_2 \mathbf{D}_2) \times \mathbf{D}_1 = -\hat{w}_2 \mathbf{D}_3 = -(\mathbf{D}_3' \cdot \mathbf{D}_1)\mathbf{D}_3 \\ &= -\{\mathbf{D}_3 [\mathbf{D}_3']^T\} \mathbf{D}_1. \end{aligned} \tag{6.6}$$

This is a first-order linear differential equation for the 3-vector $\mathbf{D}_1$; given the centerline $\mathbf{r}(s)$, the tangent vectors $\mathbf{D}_3(s) = \mathbf{d}_3(s)$ are determined by condition (3.1), so once an initial normal vector $\mathbf{D}_1(0)$ is chosen, Eq. (6.6) can be solved numerically with an initial value problem solver to find $\mathbf{D}_1(s)$. [In practice, we solve Eq. (6.6) with a sixth order hybrid Gear routine[23] and monitor the normalization of $\mathbf{D}_1$ and its perpendicularity to $\mathbf{D}_3$ to check the accuracy of the solution].

VII. CONTINUUM ROD COMPUTATIONS

For the remainder of this paper, we focus on the problem of computing the lowest-energy equilibrium configuration of a continuum rod. We first summarize the implementation of the procedure described in Sections V and VI to compute K_i, $\hat{w}_i$, and $\Omega(s)$ given wedge-angle model parameters (recall from Section VI C that $\hat{w}_i$ and $\Omega(s)$ together determine $\hat{u}_i$).

Given a DNA base-pair sequence, we pad the sequence on both ends and then use the chosen set of unstressed angles (e.g., Trifonov or De Santis) to build a discrete unstressed shape with centerline $\mathbf{r}^{(j)}$ and normal vectors $\mathbf{d}_1^{(j)}$. The centerline is double-filtered and a least-squares approximation then gives $\mathbf{r}(t)$ as described in Section VI B; to distinguish

$\mathbf{r}(t)$ from the scaled and arclength-parametrized centerline that is our ultimate goal, we relabel it as $\overline{\mathbf{r}}(t)$. The arclength function σ is defined as usual by

$$\sigma(\tau) \equiv \int_0^\tau \left| \frac{d\overline{\mathbf{r}}}{dt}(t) \right| dt,$$

whose inverse function we denote by $\tau(\sigma)$; it is straightforward to determine both $\sigma(\tau)$ and $\tau(\sigma)$ numerically. We denote the total arclength of $\overline{\mathbf{r}}$ by ℓ. Also, let σ_{init} and σ_{final} denote the values of σ at the ends of the true molecule (within the padded centerline), and define $s_{\text{init}} \equiv \sigma_{\text{init}}/\ell$, $s_{\text{final}} \equiv \sigma_{\text{final}}/\ell$.

The scaled and arclength-parametrized centerline is then defined by

$$\mathbf{r}(s) \equiv \frac{1}{\ell}\overline{\mathbf{r}}(\tau(\ell s)).$$

Using the fact that

$$\frac{d\tau}{d\sigma}(\ell s) = \frac{1}{\left| \frac{d\overline{\mathbf{r}}}{dt}(\tau(\ell s)) \right|},$$

it then follows from the chain rule that

$$\mathbf{d}_3(s) = \frac{d\mathbf{r}}{ds}(s) = \frac{\frac{d\overline{\mathbf{r}}}{dt}}{\left| \frac{d\overline{\mathbf{r}}}{dt} \right|},$$

$$\mathbf{d}_3'(s) = \frac{d^2\mathbf{r}}{ds^2}(s) = \ell \frac{\left| \frac{d\overline{\mathbf{r}}}{dt} \right|^2 \frac{d^2\overline{\mathbf{r}}}{dt^2} - \left(\frac{d\overline{\mathbf{r}}}{dt} \cdot \frac{d^2\overline{\mathbf{r}}}{dt^2} \right) \frac{d\overline{\mathbf{r}}}{dt}}{\left| \frac{d\overline{\mathbf{r}}}{dt} \right|^4},$$

where all derivatives of $\overline{\mathbf{r}}$ are evaluated at $\tau(\ell s)$. Having computed $\mathbf{d}_3$ and $\mathbf{d}_3'$, it is then possible to compute the natural-frame normal vectors $\mathbf{D}_1$ from Eq. (6.6) using the fact that $\mathbf{D}_3 = \mathbf{d}_3$. From $\mathbf{D}_1(s)$ we then compute the function $\Omega(s)$ which records the angle of rotation between $\mathbf{D}_1$ and the true DNA normal vectors $\mathbf{d}_1^{(j)}$. By the definition of the natural frame, $\hat{w}_3 = 0$, and $\hat{w}_1$ and $\hat{w}_2$, the unstressed strains with respect to the natural frame, can be computed using Eq. (3.4). Although $\hat{w}_1$ and $\hat{w}_2$ are thus determined for the padded molecule, it is routine to obtain their values for the true molecule: the unpadded centerline and normal vectors are given by

$$\mathbf{r}_{\text{unpadded}}(s) = \frac{\mathbf{r}_{\text{padded}}(s_{\text{init}} + s(s_{\text{final}} - s_{\text{init}}))}{s_{\text{final}} - s_{\text{init}}},$$

$$\mathbf{d}_{1,\text{unpadded}}(s) = \mathbf{d}_{1,\text{padded}}(s_{\text{init}} + s(s_{\text{final}} - s_{\text{init}}))$$

($\mathbf{r}$ must be rescaled in order for $|\mathbf{r}'| = 1$), and it is easy to verify that

$$\hat{w}_{1,\text{unpadded}}(s) = (s_{\text{final}} - s_{\text{init}})\hat{w}_{1,\text{padded}}(s),$$

$$\hat{w}_{2,\text{unpadded}}(s) = (s_{\text{final}} - s_{\text{init}})\hat{w}_{2,\text{padded}}(s).$$

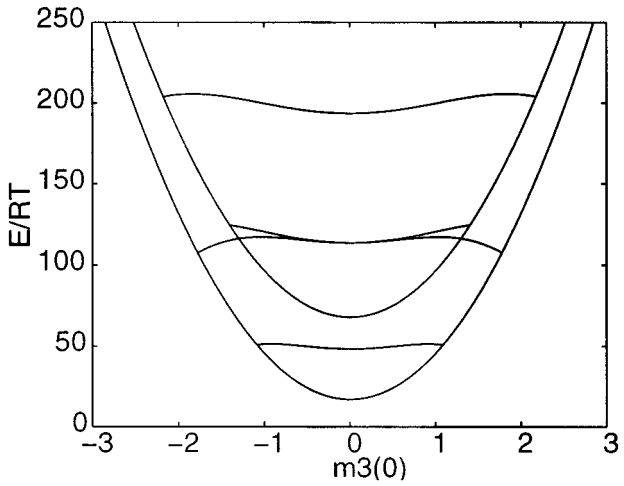

FIG. 6. Section of the perfect bifurcation diagram. Each point represents an equilibrium twisted ring for an intrinsically straight and untwisted rod with equal bending stiffnesses. As the imposed twist is varied, we get the intricate pattern of equilibrium solutions shown here; the energy and twist $m_3(0)$ of each equilibrium are plotted (for the particular case $K_1 = K_2 = RT/((158)(0.00734))$, $K_3 = 1.5K_1$ corresponding to the 158-base-pair DNA molecule 11T15 described in Section X).

These $\hat{w}_i$ are inserted for $\hat{u}_i$ into the right-hand side of Eq. (4.9) along with $K_1 = K_\theta/N$, $K_2 = K_\phi/N$, and $K_3 = K_\tau/N$, and we then seek the lowest-energy solution of the BVP [Eq. (4.9)+(4.11)] with $\alpha = \Omega(1) - \Omega(0) \pmod{2\pi}$ (as discussed in Section VI C). An efficient computational approach to this problem is parameter-continuation. Specifically, we use the package AUTO;[24] given a known solution of a parameter-dependent BVP, AUTO computes a family of solutions as a parameter (either in the differential equations or in the boundary conditions) is varied. For example, starting from the closed-form solution (4.12) to the perfect problem ($K_1 = K_2$, $\hat{w}_i = 0$), one can vary the imposed twist angle α to sweep out an intricately connected family of solutions.[12,13] For another recent use of numerical continuation in rod computations based upon a collocation discretization (as in AUTO), see Mahadevan and Keller,[25] where Mobius-band equilibria for intrinsically straight rods with rectangular cross-sections are computed beginning with a known solution for the circular cross-section case.

A piece of the set of solutions to the perfect problem is shown in Figure 6, in which the energy and local twist $m_3(0)$ are plotted for each solution; we call this set the *perfect bifurcation diagram*. Our focus here, however, is on computing equilibria for intrinsically curved rods, with $\hat{w}_1, \hat{w}_2 \neq 0$ (an *imperfect problem*). There are some intricacies associated with numerically breaking the perfect rod's symmetry, and we follow the procedure described in detail by Li and Maddocks,[13] Section VI. We introduce a homotopy parameter γ into the differential equations (4.9), replacing every instance of $\hat{w}_i(s)$ with $\gamma\hat{w}_i(s)$. Accordingly, $\gamma = 0$ corresponds to the perfect problem, for which Eq. (4.12) is known to be a solution, and $\gamma = 1$ corresponds to the imperfect problem of interest. Starting from Eq. (4.12) for some value of $\alpha = \alpha_0$, we use AUTO to increase γ from 0 to

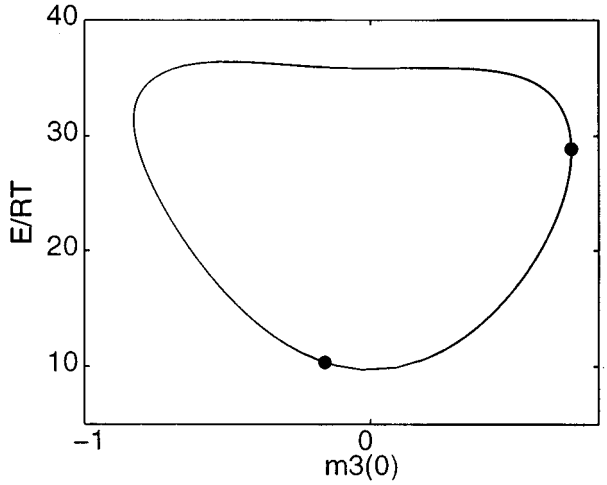

FIG. 7. Component of the bifurcation diagram for an imperfect rod modeling the DNA molecule 11T15. This component is a perturbation of the lowest-energy cycle in the perfect diagram. The large intrinsic bend of the unstressed rod (about 110°) allows the energies in this component of the imperfect diagram to be lower than those of the perfect rod, since less bending in addition to the unstressed curvature needs to occur in order to form a cycle. The solutions whose normal-vector fields close on themselves are marked with circles. As in Figure 6, we take $K_1=K_2=RT/((158)\times(0.00734))$ and $K_3=1.5K_1$.

1 (while holding α fixed) and then switch parameters to sweep out the imperfect diagram by varying α (while holding γ fixed). In contrast to the connected perfect diagram (Figure 6), the imperfect diagram contains many (apparently infinitely many) components; the component traced out by AUTO depends on the value of α_0 chosen. Also, recall that of

FIG. 8. Configuration of the rod equilibrium corresponding to the lower-energy circle in the bifurcation diagram of Figure 7. The rod centerline is indicated by a tube and the normal-vectors by a ribbon. The top figure shows the original normal vectors, which track the sugar–phosphate chains. The bottom figure shows the natural-frame normal vectors used in the computations (which are not periodic for this configuration).

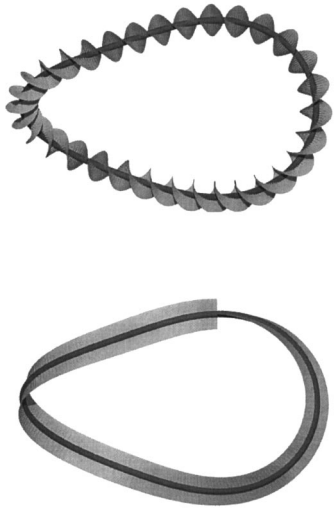

FIG. 9. Configuration of the rod equilibrium corresponding to the higher-energy circle in the bifurcation diagram of Figure 7. Interpretations are as in Figure 8.

the solutions on the imperfect diagram, only those with $\alpha=\Omega(1)-\Omega(0)\pmod{2\pi}$ have the normal-vector field $\mathbf{d}_1(s)$ close up on itself and thus represent cyclized DNA.

One component of an imperfect diagram is shown in Figure 7; the two solutions for which $\alpha=\Omega(1)-\Omega(0)\pmod{2\pi}$ are marked by circles, and the corresponding rod configurations are shown in Figures 8 and 9. This component was computed using $\alpha_0\approx 0$ ($\alpha_0=0$ is a degenerate point from which numerical continuation is not possible). It is a clear descendant of the lowest-energy cycle in the perfect diagram shown in Figure 6. This lowest-energy cycle in the perfect diagram actually splits into two components in the imperfect problem,[13] but we find in our applications that the other descendant lies at a higher energy than the component shown in Figure 7. With an apparent infinity of connected components to the imperfect diagram, we cannot be completely assured of finding the lowest-energy branch, but we take the ancestry from the perfect diagram, and experimentation with various α_0, as evidence that the $\alpha_0\approx 0$ component shown in Figure 7 contains the lowest-energy equilibrium configuration.

Having found a minimum-energy equilibrium for one value of K_3/K_1, one can easily generate minimum-energy equilibria for other K_3/K_1 by freezing α and declaring K_3/K_1 to be the active continuation parameter in AUTO. This fact is a significant benefit of the computational technique of parameter continuation.

VIII. NUMERICAL ROBUSTNESS OF CONTINUUM COMPUTATIONS

Section VI describes a procedure for determining continuum rod parameters from given discrete rod parameters, but within this procedure there are still several choices: the filter window-width w used in smoothing the discrete data, and the number of subintervals k and polynomial degree m used in the least-squares approximation to the resulting shape. In this section, we discuss how sensitive the computational results are to these choices. The goal, of course, is

that the final computations not depend significantly on the choices made, and within fairly wide limits this is indeed the case. A subtle aspect of this problem, at least for the DNA discrete data we have used, is that visually assessing the quality of the long-scale structure of the centerline is not a sufficient check of the stability of a model. Most procedures we investigated, even those which used interpolation instead of least-squares approximation, or did not include the enhancements described in Section VI B, produced continuous centerlines which appeared to be reasonable fits to the discrete centerline. The crucial point is the stability of the *curvatures* of the centerline, since the boundary-value problem for the continuous rod involves the parameters $\hat{u}_i$, which depend on the centerline curvatures (and the rotation of the normal vectors).

We took a 155-base-pair DNA molecule (08T15 as described in Section X) and varied w from 5 to 25, k from 10 to 20, and m from 4 to 6. For each triple (w,k,m), we produced a plot of the energy of the lowest-energy cyclized rod versus the ratio K_3/K_1 (as described in Section X, this is the central computation for comparison with experiment). As shown in Figure 10, when we vary w, k, and m over the ranges described above, the computed energies change by at most $0.1RT$, which is approximately a 1% variation and is well within the experimental error in energy determination, which is estimated at $0.7RT$ (see Crothers *et al.*[26] and Section X). We also remark that this energy variation is markedly smaller than that which we observed for other discrete-to-continuous procedures, e.g., if interpolation is used instead of a least-squares approximation, or if a least-squares approximation is used without a filter.

In addition to model parameters, one must choose parameters for the numerical discretization. AUTO implements collocation, which subdivides the arclength into K subintervals and approximates solutions by continuous piecewise-polynomials of degree M on this mesh of subintervals. For most computations we used $K=30$ and $M=6$ but have verified that computed energies vary by at most $0.01RT$, or 0.1%, in the range $15<K<60$ and $4<M<6$. In particular, the model errors described above (and the experimental errors mentioned in Section X) dominate these discretization errors by an order of magnitude.

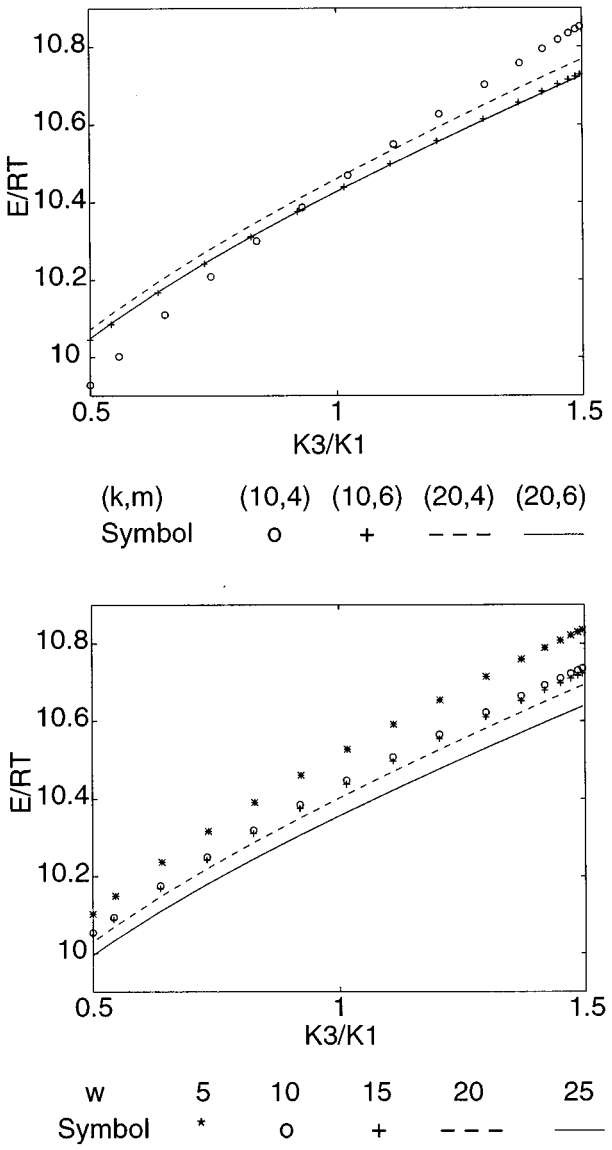

FIG. 10. Variation of computed energies with changes in filtering and fitting parameters for converting from the discrete to the continuum representation. In the top plot, the filter width is held fixed at $w=15$ and the fitting parameters (k,m) are varied. Except for $(k,m)=(10,4)$, which is probably too few polynomials of too low a degree, the energy variation is less than $0.05RT$, or 0.5%. In the bottom plot, the fitting parameters are held fixed at $(k,m)=(20,6)$ and the filter width w is varied. Here, $w=5$ appears to be too narrow a filter window, but the other choices again yield an energy variation of approximately $0.05RT$.

IX. COMPARISON OF CONTINUUM AND DISCRETE EQUILIBRIA

As one test of the accuracy of the continuum computations described in Section VII, we compare the computed continuous configurations and energies to the equilibrium configurations and energies of the original discrete problem which provided input parameters to the continuum rod model. As shown in Section IV A, determining discrete equilibria requires solution of the large nonlinear system (4.6), which necessitates a numerical iteration. Fortunately, a good initial guess for the discrete iteration can be extracted from the associated continuum solution. The determination of a discrete configuration is essentially an inversion of the derivation described in Section VI of the continuum shape from the discrete shape.

Given the continuum rod $\mathbf{r}(s)$ and $\mathbf{d}_1(s)$, $0<s<1$, we recover the positions of the base-pair centers by sampling evenly along $\mathbf{r}(s)$

$$\mathbf{r}^{(0)}=\mathbf{r}^{(N)}=\mathbf{r}\left(\frac{2N-1}{2N}\right),$$

$$\mathbf{r}^{(i)}=\mathbf{r}\left(\frac{2i-1}{2N}\right), \quad i=1,\ldots,N-1.$$

Notice the shift by arclength $s=-1/(2N)$ in accord with the alignment issues discussed in Section VI A. The $\mathbf{d}_3$ vectors are now given by the requirement that they extend from one base-pair center to the next

$$\mathbf{d}_3^{(i)}=\mathbf{r}^{(i+1)}-\mathbf{r}^{(i)}, \quad j=0,1,\ldots,N-1,$$

$$\mathbf{d}_3^{(N)}=\mathbf{d}_3^{(0)}.$$

Finally, the $\mathbf{d}_1$ vectors are recovered by sampling evenly along $\mathbf{d}_1(s)$, but now without the shift in arclength, again in accord with the alignment issues discussed in Section VI A

$$\mathbf{d}_1^{(i)}=\mathbf{d}_1\left(\frac{i}{N}\right), \quad i=0,1,\ldots,N.$$

The vectors $\mathbf{d}_3^{(i)}$ should be approximately of length $1/N$, and we should have approximate orthogonality between $\mathbf{d}_3^{(i)}$ and $\mathbf{d}_1^{(i)}$. We next project $\mathbf{d}_1^{(i)}$ onto the plane perpendicular to $\mathbf{d}_3^{(i)}$ (to obtain exact orthogonality) and then rescale $\mathbf{d}_1^{(i)}$ and $\mathbf{d}_3^{(i)}$ to have length one to complete the reconstruction of the discrete configuration (projecting $\mathbf{d}_1^{(i)}$ will change its length, so we restore its normalization at this point). Note that normalizing $\mathbf{d}_3^{(i)}$ will change the $\mathbf{r}^{(i)}$, but that is immaterial since solution of the discrete problem only involves relative rotation angles.

From the discrete frames, the relative rotation matrices are determined from Eq. (2.2)

$$\mathbf{R}^{(i)}=\begin{bmatrix}(\mathbf{d}_1^{(i)})^T\\(\mathbf{d}_2^{(i)})^T\\(\mathbf{d}_3^{(i)})^T\end{bmatrix}[\mathbf{d}_1^{(i+1)}\ \ \mathbf{d}_2^{(i+1)}\ \ \mathbf{d}_3^{(i+1)}]$$

and then angles can be determined using Eq. (2.1)

$$\theta^{(i)}=-\arcsin(\mathbf{R}_{(3,2)}^{(i)}),$$

$$\phi^{(i)}=-\arcsin(\mathbf{R}_{(3,1)}^{(i)}/\cos\ \theta^{(i)}),$$

$$\tau^{(i)}=-\arcsin(\mathbf{R}_{(1,2)}^{(i)}/\cos\ \theta^{(i)}).$$

These angles are then used as an initial guess in the solution of Eq. (4.6) using the constrained minimization function CONSTR in Matlab's Optimization Toolbox.[27]

We made this comparison for the 11 DNA molecules studied in Section X with Trifonov II unstressed angles (see Section X) and stiffnesses $K_\theta=K_\phi=1$ and $K_\tau=1.5$ (the absolute energy scaling is immaterial to the computations, so $K_\theta=1$ is chosen purely for convenience). The lowest-energy continuum configurations are computed as described in Section VII. Then, starting with the initial guess supplied by the continuum solution, each discrete constrained minimum was found in approximately 15 to 90 min on a Dec Alpha 3000 (the variation in run times is perhaps an indication of the sensitivity of the constrained minimization computation to the initial guess). The discrete and continuous energies agree to within 0.5%; see Table II. The discrete and continuum configurations for the molecule 12A09 are superimposed in Figure 11.

In our experience, solution of the discrete problem (4.6) without the initial guess provided by the continuum solution is difficult. If the initial guess is chosen to be the planar untwisted solution (4.7) or the unstressed shape, the CONSTR function fails to converge. Indeed, the convergence of the discrete iteration appears to be directly attributable to the accuracy of the initial guess derived from the continuum solution.

In addition, we have tried to solve Eq. (4.6) with parameter continuation (using AUTO) starting with the known untwisted planar solution (4.7) and with a homotopy parameter γ in the unstressed angles. However, this numerical continuation was not successful, perhaps because the rapid twist of the unstressed DNA is too different from the untwisted starting point. It is possible that if one could find a known solution with comparable twist, one could compute nearby discrete equilibria by parameter continuation, but our experience has been that parameter continuation within the wedge-angle model is impracticable.

TABLE II. Discrete and continuum energies for lowest-energy equilibrium rings, with Trifonov II angles (see Section X) and $K_\theta=K_\phi=1$, and $K_\tau=1.5$. The rms deviation between base-pair centers in the continuous and discrete equilibrium configurations is also given.

No.	Name	E_{disc}	E_{cont}	% Difference	rms configuration difference (Å)
9	12A09	0.10746	0.10740	0.05	0.59
10	09T09	0.11701	0.11714	0.11	0.64
8	13A09	0.09469	0.09457	0.13	0.58
11	17A11	0.09682	0.09654	0.29	0.63
1	11A17	0.09860	0.09817	0.43	0.58
7	15A09	0.08305	0.08280	0.30	0.58
4	11T15	0.07732	0.07730	0.03	0.61
2	09A17	0.07976	0.07947	0.36	0.58
3	08A17	0.07414	0.07390	0.32	0.59
6	08T15	0.07501	0.07487	0.22	0.61
5	09T15	0.07292	0.07284	0.11	0.60

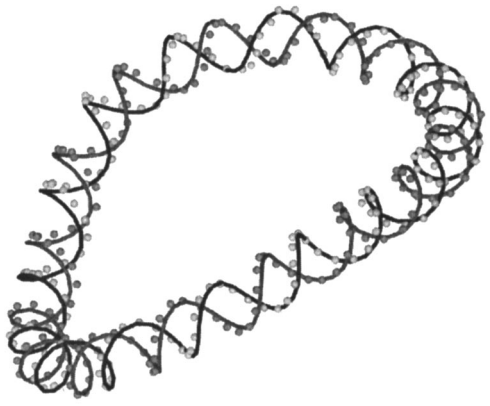

FIG. 11. Discrete and continuous equilibrium configurations for the 12A09 DNA molecule using Trifonov II angles (see Section X). The sugar-phosphate helices are reconstructed from the computed centerlines $\mathbf{r}$ and frames $(\mathbf{d}_1,\mathbf{d}_2,\mathbf{d}_3)$ using the idealized B-DNA coordinates: helix1 $=\mathbf{r}+2.521\mathbf{d}_2-0.769\mathbf{d}_1$, helix2$=\mathbf{r}-2.521\mathbf{d}_2-0.769\mathbf{d}_1$. The discrete helices are shown with spheres and the continuous helices with tubes.

X. APPLICATION: DNA CYCLIZATION

We modeled 11 DNA molecules with lengths ranging from 150 to 160 base-pairs, whose experimental cyclization probabilities span three orders of magnitude.[28] All the molecules contain, in sequence, the following segments:

(1) a PCR segment (essentially straight; a vestige of a polymerase chain reaction (PCR) primer),
(2) a CAP-binding site (bent by $\approx 10°$; a site where catabolite activator protein (CAP) can bind),
(3) adaptor II (essentially straight),
(4) a sequence of 6 A-tracts (total bend $\approx 108°$),
(5) adaptor I (essentially straight).

Although the DNA contains a CAP-binding site, no protein-binding is involved in this article. The effect of CAP binding on these molecules is investigated with experiments and Monte Carlo computations in Kahn and Crothers;[28] a continuum model incorporating protein-binding is a subject of future study.

Some molecules have the A-tract adenines on the opposite strand from the others; these two types are distinguished by labeling the molecules as ''A'' or ''T.'' The molecules are also distinguished by different lengths of adaptor I and adaptor II. Following the notation in Kahn and Crothers,[28] the principal 5 DNA molecules are labeled by their adaptor lengths and their A-tract orientation: 11A17, 11T15, 15A09, 09T09, and 17A11; the first number is the length of adaptor I and the second that of adaptor II. In addition, to study the effects of overall sequence length, shorter versions of some of these molecules were made by removing 2 or 3 base-pairs from the PCR segment to form what we call 09A17, 08A17 (from 11A17), 09T15, 08T15 (from 11T15), and 13A09, 12A09 (from 15A09). Note that these names are somewhat misleading in that it is not the adaptor segment but rather the PCR segment that has been shortened, via synthesis using PCR primers bearing 2 or 3 nucleotide deletions relative to the principal molecules. Due to a mutation, the 17A11 sample used in the experiment was actually missing a base-pair from one of its A-tracts; this change was included in all computations. Detailed structures and experimental information can be found in Kahn and Crothers.[8,28]

Experimental studies using T4 DNA ligase-mediated trapping of apposed DNA ends[26,29,30] do not directly measure the probabilities of cyclization, but rather the rate constants (k_c, k_d) for cyclization and dimerization (the bonding of one molecule's $5'$ end to another molecule's $3'$ end to form a DNA dimer). These in turn can be related to the equilibrium constants (K_c, K_d) for cyclization and dimerization[26,30] by

$$k_c/k_d = K_c/K_d .$$

This quantity is defined to be the Jacobsen–Stockmayer factor J:[31]

$$J \equiv K_c/K_d .$$

A useful interpretation is to think of J as ''the molar DNA concentration required to cause bimolecular joining to occur at the same rate as the corresponding cyclization reaction.''[26] Standard statistical mechanical expressions for the equilibrium constants K_c and K_d give[32]

$$J = \exp\left[-\frac{\Delta G_c^0 - \Delta G_d^0}{RT}\right],$$

where ΔG_c^0 and ΔG_d^0 are the standard molar free energy changes in the cyclization and dimerization reactions. Hence, the quantity

$$\Delta G_{\exp} \equiv \Delta G_c^0 - \Delta G_d^0 = \Delta H_c^0 - T\Delta S_c^0 - (\Delta H_d^0 - T\Delta S_d^0) \tag{10.1}$$

can be experimentally determined. We focus first on the enthalpy contribution. The term ΔH_d^0 is the enthalpy for the new chemical bonds at the $5'-3'$ connection. The term ΔH_c^0 contains the same enthalpy of the $5'-3'$ connection, and also the enthalpy change due to the rearranged molecular shape in cyclization. The enthalpy from the new bonds will cancel in the difference $\Delta H_{\exp} \equiv \Delta H_c^0 - \Delta H_d^0$ to leave only the enthalpy change due to the rearranged cyclized shape; it is exactly this enthalpy change $\Delta H_{\exp}$ which should be well-approximated by the strain energies in the rod model.

Accordingly, we can compare the experimentally determined J factors with strain energies computed in our continuum rod model. There are two important points to consider in making this comparison. First, the experimentally determined J factors, since they involve *free* energies, will contain entropy contributions not computed in the continuum model. Second, experimental difficulties imply that *relative* J factors among a set of molecules are generally more reliable than the absolute J factors.[26]

The discrete bending stiffnesses were taken to be

$$K_\theta = K_\phi = \frac{RT}{l/P}$$

with a helix-rise-per-base-pair $l = 3.4\times10^{-8}$ cm and persistence length $P = 463\times10^{-8}$ cm, to match the values used in associated Monte Carlo studies.[28] The Monte Carlo study took $K_\tau \approx 1.5K_\theta$, but we investigated the entire range $0.5 < K_\tau/K_\theta \approx K_3/K_1 < 1.5$ since various values in this range are reported in the literature.[1] The parameter-continuation techniques we employ make it particularly easy to compute the entire range of these solutions. For those more familiar with the notation used in Schlick,[1] the range $0.5 < K_3/K_1 < 1.5$ corresponds to a range of ''torsional moduli'' C of $0.9–2.8\times10^{-19}$ erg cm (where $C = K_\tau l/N_{\mathrm{Av}}$, with N_{Av} = Avogadro's number).

The unstressed shape of each discrete molecule was computed for each of three different sets of equilibrium wedge angles, namely Trifonov,[9] De Santis,[10] and a set of adapted Trifonov angles we call Trifonov II. Trifonov II angles were determined in the Monte Carlo study[28] to better match recent experimental data. The modifications are: first, the $\hat{\tau}$ values for different base-pair stacks are replaced with a single value of 34.45°, to match the helical repeat of 10.45 seen in studies of the periodicity of cyclization probability with varying DNA sequence length;[30] second, $\hat{\tau} = 34.85°$ is

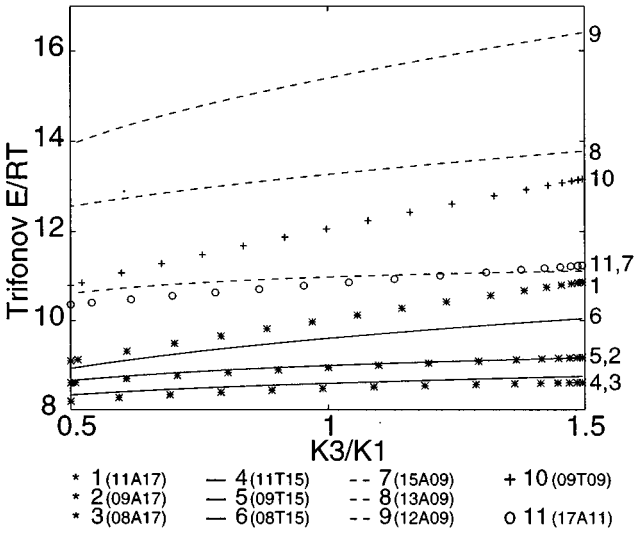

FIG. 12. Continuum model strain energies of the 11 DNA molecules in Table III for various values of K_3/K_1 (ratio of twist to bend stiffness), using Trifonov angles for the unstressed shape. When two or more strands are listed on one line at the right edge of the graph, the one with the higher energy at $K_3/K_1=1.5$ is listed first.

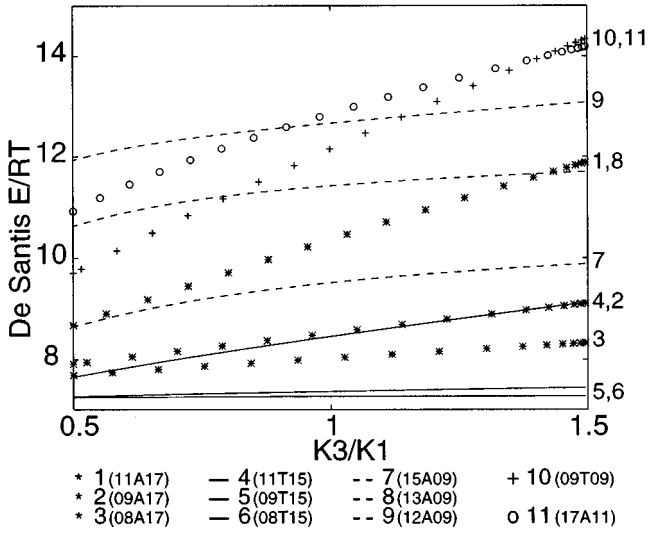

FIG. 13. Continuum model strain energies, using De Santis angles for the unstressed shape. Notation and curve labeling are as in Figure 12.

used in the A-tracts to match the helical repeat of 10.33 found in Crothers *et al.*;[26] finally, the Trifonov angles $\hat{\theta}$ and $\hat{\phi}$ are scaled by 0.61 to account for recent cyclization kinetic studies[33] indicating that the bending angle in the particular A-tract sequence used is approximately 18° rather than the 30° predicted by Trifonov angles.

From these discrete parameters, continuum computations of the lowest-energy equilibria were made for each of the three sets of unstressed angles. For each molecule, starting from discrete wedge-angles, it takes approximately 15 min on a Dec Alpha 3000 to generate equilibrium energies and configurations for 16 different values of K_3/K_1: 5 min to determine continuum parameters from the base-pair sequence, 5 min to determine an equilibrium solution for an initial value of K_3/K_1, and about 15 s for each subsequent value of K_3/K_1. Figures 12, 13, and 14 show our computed cyclization strain energies $\Delta E_{\text{model},i}$, $i=1,\ldots,11$ for each of the 11 DNA strands plotted against K_3/K_1. In addition, by computing an appropriate Hessian, we have verified that these equilibria generate local minima of the discrete problem, and thus are likely to be the configurations realized in experiments. Interestingly, for some molecules and some values of K_3/K_1, both continuum solutions on the low-energy branch of the imperfect diagram generate discrete configurations which are local minima; investigation of this phenomenon is ongoing.

We can offer a heuristic explanation for the different cyclization probabilities of these molecules and the effect of varying K_3/K_1. According to the bend phasing described in Table III, the A17 and T15 molecules are approximately C-shaped, while A09, T09, and A11 molecules are approximately S-shaped. The C-shaped molecules tend to cyclize readily; however, in order for the S-shaped molecules to close their centerlines, they must either (a) twist to rephase their bends or (b) bend against their intrinsic curvature. Figure 15 plots the bending contribution

$$\int_0^1 \left[\frac{1}{2}K_1(u_1(s)-\hat{u}_1(s))^2+\frac{1}{2}K_2(u_2(s)-\hat{u}_2(s))^2\right]ds$$

to the continuum energy (3.5); the higher bending energies of the A09, T09, and A11 molecules suggest that mechanism (b) dominates in their cyclization. For molecules with two large bends, such as CAP-bound S-shaped molecules, mechanism (a) seems more likely.

Another important effect is the additional twisting energy required to align the sugar–phosphate chains once the centerline is closed. Figure 16 plots the twist contribution

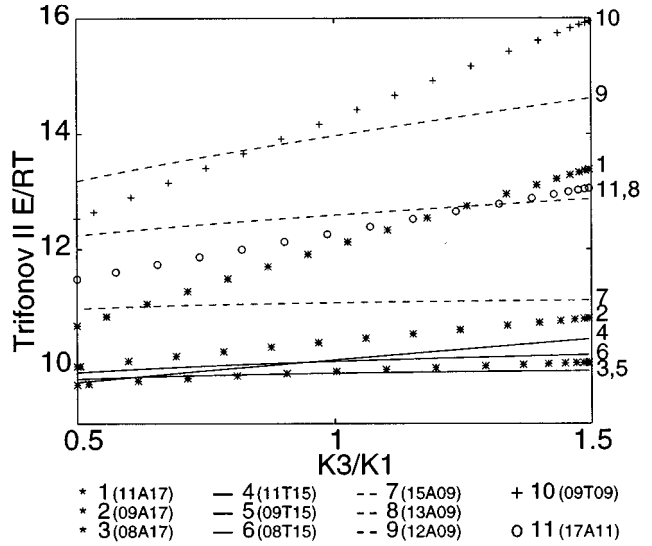

FIG. 14. Continuum model strain energies, using Trifonov II angles for the unstressed shape. Notation and curve labeling are as in Figure 12.

TABLE III. The DNA molecules: numbered as in Figures 12–19 and ranked according to their experimental free energies (Ref. 28). The torsional phasing gives the total number of turns of the sugar–phosphate chain, assuming 10.33 bp/turn in the A-tracts and 10.45 bp/turn elsewhere (Ref. 28). The bend phasing gives the total number of turns between the center of the CAP site bend to the center of the first A-tract; when this phase is an integer, the two bends (CAP site and A-tracts) are in phase, so they form a C-shaped molecule. Also shown are the discrepancies of various computed free energies from the experimental results; all computations are for $K_3/K_1=K_\tau/K_\theta=1.5$. The three continuum computations use parameters derived from different sets of unstressed wedge-angles, as described in Section X. These continuum computations also include fitting with a parameter (whose best-fit value is given in the last line) to include a constant free-energy-shift due to entropy.

					$(\Delta G_{\text{exp}}-\Delta G_{\text{model}})/RT$			
No.	Name	Torsional phasing	Bend phasing	$\Delta G_{\text{exp}}/RT$	Monte Carlo	Continuum Trifonov II	Continuum Trifonov	Continuum De Santis
9	12A09	14.7	4.3	21.0	0.4	0.6	−3.0	0.1
10	09T09	14.4	4.3	21.0	1.0	−0.5	1.4	−0.2
8	13A09	14.8	4.3	20.1	1.0	1.8	0.0	0.2
11	17A11	15.4	4.5	19.1	0.1	0.7	1.9	−3.0
1	11A17	15.4	5.1	17.9	0.4	−0.9	0.9	−0.6
7	15A09	15.0	4.3	17.2	−0.3	0.5	−0.6	−0.4
4	11T15	15.2	4.9	16.2	0.8	0.3	1.4	0.6
2	09A17	15.2	5.1	15.7	−0.2	−0.5	0.4	0.4
3	08A17	15.1	5.1	15.0	−0.8	−0.5	0.2	0.4
6	08T15	14.9	4.9	14.8	−0.8	−0.8	−1.7	1.4
5	09T15	15.0	4.9	14.7	−0.4	−0.6	−0.8	1.1
	ΔS determined in least-squares fit				N/A	5.2R	5.8R	6.1R

$$\int_0^1 \left[\frac{1}{2}K_3(u_3(s)-\hat{u}_3(s))^2\right] ds$$

to the continuum energy (3.5), which correlates well with the torsional phasings from Table III. The cycles which contain significant twist energy are generally more sensitive to the value of K_3/K_1, as shown by Figure 16. In summary, the ordering of the 11 cyclization rates can be explained to a first approximation by considering bend phasing and torsional phasing; however, in reality, cyclization involves a competition between twisting and bending, whose balance depends on the detailed unstressed shape and stiffnesses, as shown by the variation of the energies with K_3/K_1.

Table III reports the experimental cyclization free energies $\Delta G_{\text{exp},i}$, $i=1,\ldots,11$ (with experimental error of approximately $\pm 0.7RT$). A few conclusions are immediately apparent. First, for all three sets of angles, the ordering of the $\Delta E_{\text{model},i}$ matches the ordering of the $\Delta G_{\text{exp},i}$, taking into account the experimental error in the $\Delta G_{\text{exp},i}$. Even though the Trifonov and De Santis angles are quite different (see Table I), their computed energies are qualitatively similar,

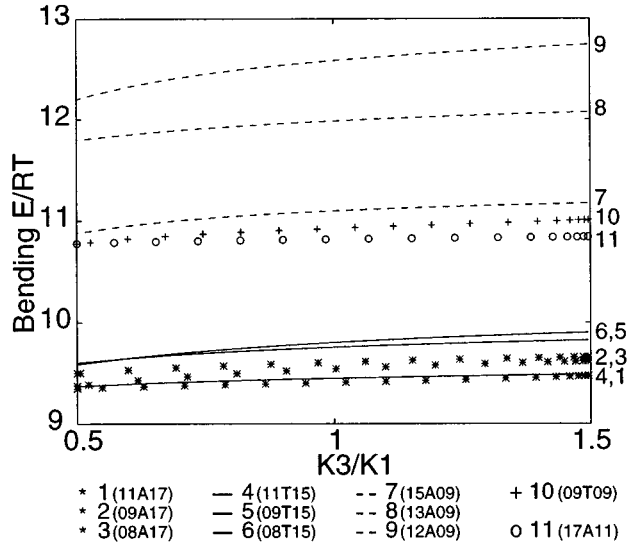

FIG. 15. Bending contribution to the continuum model strain energies, using Trifonov II angles for the unstressed shape. Notation and curve labeling are as in Figure 12.

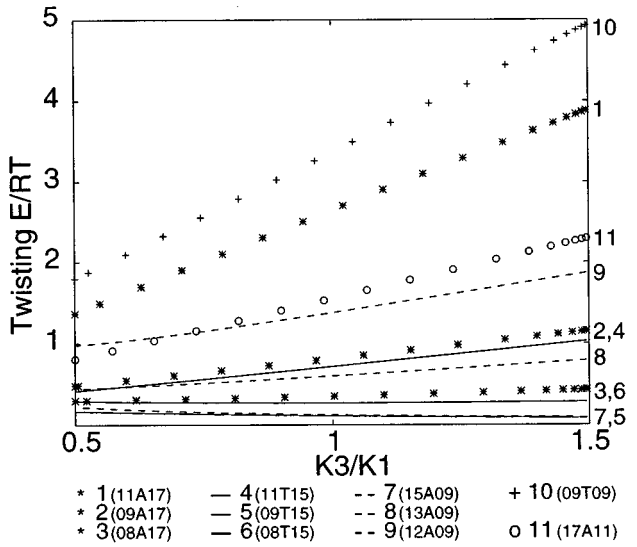

FIG. 16. Twisting contribution to the continuum model strain energies, using Trifonov II angles for the unstressed shape. Notation and curve labeling are as in Figure 12.

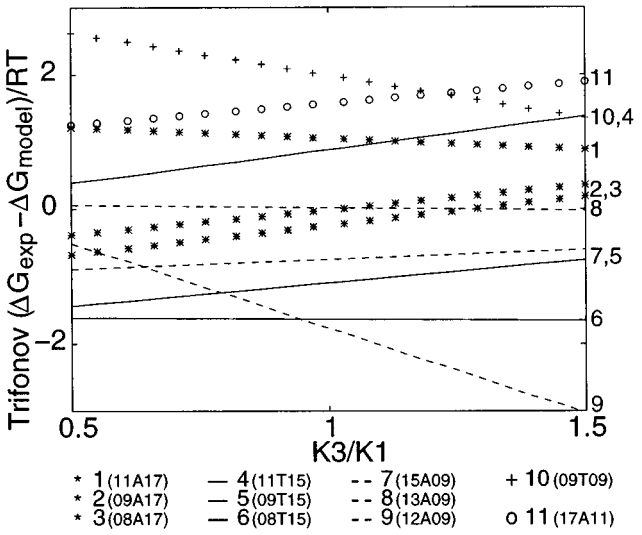

FIG. 17. Difference between experimental and rod model free energies for Trifonov angles, with energy shift due to entropy contribution included as a free parameter (the best-fit ΔS varies from $7.5R$ at $K_3/K_1=0.5$ to $5.8R$ at $K_3/K_1=1.5$). Curve labeling is as in Figure 12.

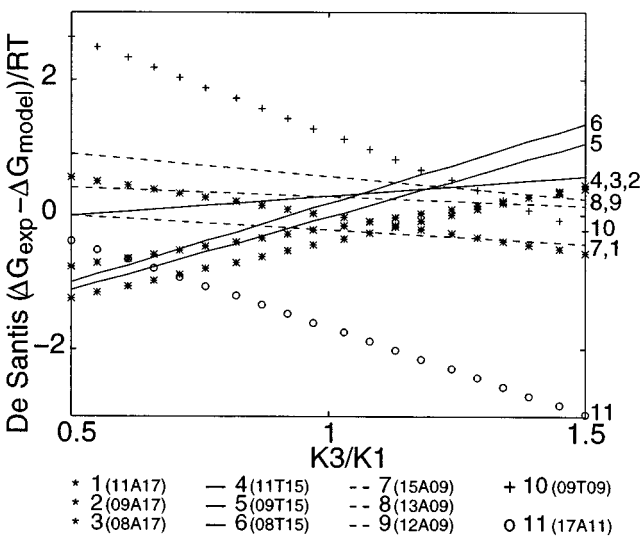

FIG. 18. Difference between experimental and rod model free energies for De Santis angles, with energy shift due to entropy contribution included as a free parameter (the best-fit ΔS varies from $8.6R$ at $K_3/K_1=0.5$ to $6.1R$ at $K_3/K_1=1.5$). Curve labeling is as in Figure 12.

perhaps because A-tracts figured importantly in their training sets.

In addition, the computed energies are uniformly lower than the experimental energies. Given the accurate energy computations in Section IX, this systematic inaccuracy is unlikely to be an error in the continuum computations, but is more likely due to the absence of entropic contributions in the elastic model. From Eq. (10.1), the entropic contribution to $\Delta G_{\rm exp}$ is $T\Delta S_d^0 - T\Delta S_c^0$ We can estimate this contribution with the following heuristic argument: each linear molecule has significant entropy, but for the short molecules considered here, the cyclized configuration has essentially zero configurational entropy; hence, $-T\Delta S_c^0 \approx TS_{\rm linear}$. Furthermore, dimerization leads to a decrease in translational and rotational entropy. In any event, detailed entropy computations are outside the scope of this article, so we merely assume that the entropy contribution to free energy is the same for all 11 molecules (since they are of similar length and shape). Accordingly, we assume a uniform shift to compute free energies from the continuum strain energies: $\Delta G_{{\rm model},i}=\Delta E_{{\rm model},i}+T\Delta S$, for $T\Delta S$ independent of i. We treat $T\Delta S$ as a free parameter, whose value is determined to give the best least-squares fit (at each value of K_3/K_1) between the 11 computed $\Delta G_{{\rm model},i}$ and experimental $\Delta G_{{\rm exp},i}$. The results of this best fit are shown in Figures 17, 18, and 19. The entropy shifts used varied from $5.2R$ to $8.6R$, depending on the value of K_3/K_1 and the angle set, with the values further described in the figure captions.

Of course, this cannot be taken as a true computation of $T\Delta S$ or as a true quantitative comparison of the rod computations to experiment; we merely assert that it is plausible that the elastic rod computations model the experiment accurately. Moreover, Table II shows that the continuum computations compute the exact equilibrium wedge-angle internal energy to within $0.05RT$ (0.5% error), so the discrepancies between continuum predictions and experiment in Figures 17, 18, and 19 must instead be due to experimental error, inaccuracies in the wedge-angle model, or the incorrectness of our assumption of a constant free-energy shift due to entropy.

Finally, we compare our continuum equilibrium energies to Monte Carlo computations on the same DNA molecules. Monte Carlo simulations, which build DNA configurations with probabilities weighted by the discrete energies of those configurations, can simulate the actual fluctuations of the DNA molecule, and counting the resulting configurations which satisfy the ring-closure constraint (to within a toler-

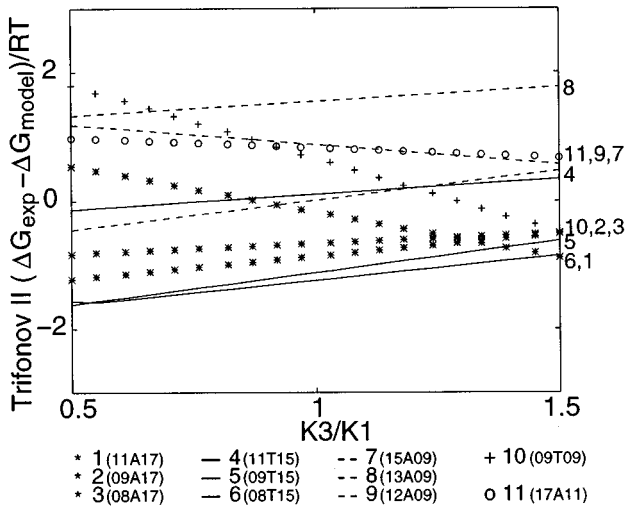

FIG. 19. Difference between experimental and rod model free energies for Trifonov II angles, with energy shift due to entropy contribution included as a free parameter (the best fit ΔS varies from $6.5R$ at $K_3/K_1=0.5$ to $6.1R$ at $K_3/K_1=5.2$). Curve labeling is as in Figure 12.

ance) can be used to determine the ring-closure probability.[26,28,32,34] This approach has been successful in reproducing cyclization free energies (including the entropic contribution) determined by experiment.[28] However, it is a computationally intensive procedure, which samples approximately 10^{10} configurations to get stable cyclization probabilities for some molecules (those which cyclize with low, but experimentally accessible, probability) and requires about 12 h on a Dec Alpha. For exploring highly strained molecules such as supercoils or for scanning through parameter space (especially for more complicated DNA geometries like those including multiple protein-induced bends), the Monte Carlo method may become prohibitively time-consuming.

Ring-closure probabilities, including entropic effects, have also been studied with statistical mechanical theories of ''wormlike chains.'' For example, Shimada and Yamakawa[35] compute DNA ring-closure probabilities using a wormlike chain theory with twisted but straight unstressed shape; Hagerman and Ramadevi[36] show that Monte Carlo results match the Shimada and Yamakawa computations for short (less than 500 bp) DNA molecules.

We compare in Table III the differences between computed and experimental free energies for all the methods described here: Monte Carlo using Trifonov II angles[28] and the continuum rod computations using Trifonov, De Santis, and Trifonov II angles. All computations are for $K_3/K_1 \approx K_\tau/K_\theta = 1.5$, since that is the value used in the Monte Carlo simulations, and parameter continuation of Monte Carlo results is not possible. The improvement of the Trifonov II angles over the original Trifonov angles is clear, and the De Santis results are quite close to experiment with the exception of the outlying 17A11 molecule.

A comparison of the Monte Carlo and the continuum Trifonov II results shows that, especially for the best cyclizers, which have the smallest Monte Carlo and experimental errors, the deviations from experiment are generally well correlated, of similar magnitude and in the same direction. This suggests again that these errors are more likely attributable to shortcomings of the discrete model and its parameters than to inaccuracies in the continuum and Monte Carlo computations. Our experience has been that the continuum and Monte Carlo computations are in good agreement even for small changes in base-pair sequence. For example, in the course of our research, we computed continuum energies for two 158-base-pair molecules (11A17 with two base-pairs removed from the PCR segment and 11A17 with two base-pairs removed from adaptor I) and predicted an energy difference of $0.4RT \pm 0.1RT$. Subsequent Monte Carlo computations found an energy difference in the same direction of $0.65RT \pm 0.3RT$, in agreement with the continuum predictions. These two molecules are very similar—one expects cyclization rates to depend primarily on the DNA length and on the adaptor II sequence which phases the two DNA bends, not on the PCR segment or adaptor I at the DNA ends—and yet the continuum model was able to detect their differences, even with the filtering and smoothing involved in the model.

XI. DISCUSSION

We have presented a procedure for taking the parameters of a discrete base-pair model for DNA and producing smoothed parameters for a continuum rod model. This idea attempts to reconcile the desire for large-scale, numerically efficient discretization in computations with the experimental fact that the sequence of a DNA molecule does have an effect on some of its large-scale physical properties. Some DNA properties do not depend significantly on the base-pair sequence, and the sequence-dependent techniques described here would not be necessary for those applications, although the continuum rod model can still provide an excellent computational tool. Similarly, some DNA properties are so inherently local that they could not be captured by a computation whose discretization length covers several base-pairs, and of course the smoothing algorithm proposed here would be unsuitable for those applications. For properties between these two extremes, such as the DNA bending results in Section X or the consequences of protein binding on DNA structure (to be investigated in a future study), the continuum rod model coupled with the smoothed sequence-dependence outlined in this paper provide an accurate and efficient means to compute sequence-dependent DNA deformations.

Even though the computation of continuum rod parameters involves significant smoothing and filtering, a surprisingly detailed level of base-pair information is retained in the averaged model. As shown in Table II, the continuum energies match the discrete energies to within 0.5%. Accordingly, even if computation of discrete equilibria is the stated objective, we believe that the introduction of the continuum model described here, taken with a discretization chosen for efficient numerics, and followed by reconstruction of the base-pair configuration, is an overall efficient computational approach.

The sequences studied in Section X differ very little, and yet the rod model is able to separate their cyclization energies in a way that is consistent with experiment. Certainly there are limitations to the continuum model. It relies on accurate knowledge of the discrete model parameters describing how base-pairs stack on each other and their resistance to bending and twisting. In some settings, the discrete-model nearest-neighbor assumption in stacking of base-pairs may be incorrect; longer range effects may also be important. What does seem to be true is that if the discrete model and its parameters are accurate, then the continuum model successfully captures many of its long-range behaviors and allows for more rapid computations for those long-range behaviors. In addition, the generality of the rod model could allow continuum descriptions of more intricate discrete base-pair models as they are developed.

Note added in proof. After acceptance of this article, we learned of the study by P. De Santis, M. Fua, M. Savino, C. Anselmi, and G. Bocchinfuso, [J. Phys. Chem. **100**, 9968 (1996)], which also considers DNA cyclization (including some of the sequences considered here). That work uses an extension of the statistical mechanical twisted-wormlike-chain theory (cf. Section X) that includes an approximation

to elasticity theory based on an assumption of constant twist and a Fourier analysis of sequence-dependent curvatures. Our work provides an extremely accurate computation of purely elastic energy, but our continuum model does not capture the entropic effects approximated by the De Santis *et al.* model.

ACKNOWLEDGMENTS

R.S.M. was supported by an NSF Postdoctoral Research Fellowship and the AFOSR. J.H.M. was supported by the AFOSR and ONR. J.D.K. was supported by startup funds from the University of Maryland. Supplementary graphics and information to this article can be found at the Website: http://www.lcvm.umd.edu/~rmanning/dna.html

[1] T. Schlick, Curr. Opinion Struct. Biol. **5**, 245 (1995).
[2] W. K. Olson, Curr. Opinion Struct. Biol. **6**, 242 (1996).
[3] W. R. Bauer, R. A. Lund, and J. H. White, Proc. Natl. Acad. Sci. USA **90**, 833 (1993).
[4] I. Tobias and W. K. Olson, Biopolymers **33**, 639 (1993).
[5] H. Yamakawa and M. Fujii, J. Chem. Phys. **64**, 5222 (1976).
[6] Y. Yang, I. Tobias, and W. K. Olson, J. Chem. Phys. **98**, 1673 (1993).
[7] C. L. Brooks, Curr. Opinion Struct. Biol. **5**, 211 (1995).
[8] J. D. Kahn and D. M. Crothers, Proc. Natl. Acad. Sci. USA **89**, 6343 (1992).
[9] A. Bolshoy, P. McNamara, R. E. Harrington, and E. N. Trifonov, Proc. Natl. Acad. Sci. USA **88**, 2312 (1991).
[10] P. De Santis, A. Palleschi, M. Savino, and A. Scipioni, Biophys. Chem. **42**, 147 (1992).
[11] S. Antman, *Nonlinear Problems of Elasticity* (Springer, New York, 1994).
[12] D. J. Dichmann, Y. Li, and J. H. Maddocks, in *Mathematical Approaches to Biomolecular Structures and Dynamics, The IMA Volumes in Mathematics and Its Applications*, edited by J. P. Mesirov, K. Schulten, and D. W. Sumners (Springer, New York, 1996), Vol. 86, p. 71.
[13] Y. Li and J. H. Maddocks, J. Comput. Phys. (submitted).
[14] T. E. Haran, J. D. Kahn, and D. M. Crothers, J. Mol. Biol. **244**, 135 (1994).
[15] C. R. Calladine and H. R. Drew, *Understanding DNA: The Molecule & How It Works* (Academic, San Diego, 1992).
[16] H. Goldstein, *Classical mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1980).
[17] P. R. Selvin, D. N. Cook, N. G. Pon, W. R. Bauer, M. P. Klein, and J. E. Hearst, Science **255**, 82 (1992).
[18] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. Viovy, D. Chatenay, and F. Caron, Science **271**, 792 (1996).
[19] S. Smith, Y. Cui, and C. Bustamante, Science **271**, 795 (1996).
[20] H. F. Weinberger, *Variational Methods for Eigenvalue Approximation* (Society for Industrial and Applied Mathematics, Philadelphia, 1974).
[21] T. P. Krauss, L. Shure, and J. N. Little, *Signal Processing Toolbox User's Guide* (The Math Works, Natick, MA, 1994).
[22] J. Langer and D. Singer, to appear in SIAM Review.
[23] C. W. Gear, SIAM J. Num. Anal. **2**, 69 (1965).
[24] E. J. Doedel, AUTO: Software for Continuation and Bifurcation Problems in Ordinary Differential Equations, Technical report, Applied Mathematics, California Institute of Technology, Pasadena, CA 91125, 1986. E.J. Doedel, H. B. Keller, and J. P. Kernévez, Int. J. Bifurcation Chaos **1**, 493,745 (1991).
[25] L. Mahadevan and J. B. Keller, Proc. R. Soc. London, Ser. A **440**, 149 (1993).
[26] D. M. Crothers, J. Drak, J. D. Kahn, and S. D. Levene, Methods Enzymol. **212**, 3 (1992).
[27] A. Grace, *Optimization Toolbox User's Guide* (The Math Works, Natick, MA, 1994).
[28] J. D. Kahn and D. M. Crothers, J. Mol. Biol (submitted).
[29] D. Shore, J. Langowski, and R. Baldwin, Proc. Natl. Acad. Sci. USA **78**, 4833 (1981).
[30] D. Shore and R. Baldwin, J. Mol. Biol. **170**, 983 (1983).
[31] H. Jacobsen and W. Stockmayer, J. Chem. Phys. **18**, 1600,1607 (1950).
[32] S. D. Levene and D. M. Crothers, J. Mol. Biol. **189**, 61 (1986).
[33] H.-S. Koo, J. Drak, J. A. Rice, and D. M. Crothers, Biochemistry **29**, 4227 (1990).
[34] S. D. Levene and D. M. Crothers, J. Biomol. Struct. Dynamics **1**, 429 (1983).
[35] J. Shimada and H. Yamakawa, Macromolecules **17**, 689 (1984).
[36] P. J. Hagerman and V. A. Ramadevi, J. Mol. Biol. **212**, 351 (1990).