Modeling the sequence-dependent diffusion coefficients of short DNA molecules

O. Gonzalez[1,a)] and J. Li[2]
[1]*Department of Mathematics, University of Texas, Austin, Texas 78712, USA*
[2]*Graduate Program in Computational and Applied Mathematics, University of Texas, Austin, Texas 78712, USA*

(Received 4 June 2008; accepted 8 September 2008; published online 28 October 2008)

A boundary element model for the computation of sequence-dependent hydrodynamic properties of short DNA molecules is introduced. The hydrated surface is modeled as a curved tube of uniform radius with ends capped by hemispheres, and the axis of the tube is a general space curve whose length and curvature are determined locally by the sequence using a rigid basepair model of double-helical DNA with parameters based on x-ray crystallography. Diffusion coefficients for families of random and periodic DNA sequences are computed and compared with theories for straight tubes and experimental data. Our results indicate that sequence-dependent curvature can have a measurable impact on both the translational and rotational diffusion coefficients, even for relatively short fragments of lengths less than about 150 basepairs, and that previous estimates of the hydrated radius of DNA are likely to be underestimates. Moreover, our results suggest a possible method for refining the rigid basepair model parameters for DNA in solution as well as the hydrated radius. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2992080]

I. INTRODUCTION

There are many experimental methods for probing the structure of macromolecules in solution. Examples range from more modern techniques such as light scattering,[1] fluorescence polarization,[2] and electric dichroism and birefringence[3] to more classic techniques such as centrifugation[4–6] and electrophoresis,[7] including newer variants such as membrane-confined[8] and capillary[9] electrophoresis, which employ a solvent rather than a gel support medium. When a macromolecule is relatively rigid and its surrounding medium can be modeled as a Newtonian fluid, there exist well-defined theories connecting experimental observables to the overall size and shape of the macromolecule. Hydrodynamic properties such as the translational and rotational diffusion coefficients of the macromolecule at infinite dilution play a central role in these theories. Indeed, experimental measurements often lead to direct estimates of these diffusion coefficients, which are then connected to the size and shape of a macromolecule through an appropriate hydrodynamic model. Recent advances in experimental, modeling, and computational techniques now allow the structure of macromolecules in solution to be studied with more resolution than ever before. Methods for solution, combined with methods for crystalline states such as x-ray diffraction, together provide a powerful means of exploring macromolecular structure in various applications to proteins[10–13] and DNA.[14–17]

A hydrodynamic model of a macromolecule involves two main modeling assumptions. The first assumption is that the macromolecule can be represented by an effective hydrated surface which represents not only the macromolecule itself but also any tightly bound solvent molecules. In classic treatments, this surface has traditionally been modeled as a simple shape such as an ellipsoid or a straight circular cylinder.[18] The second assumption is that the solvent medium exterior to the hydrated surface can be modeled as a Newtonian fluid. For the description of relatively slow diffusive translational and rotational motions of a relatively large macromolecule, the steady Stokes equations with no-slip boundary conditions have had great success and have become standard.[19,20] Except for the simplest of geometries such as spheres and ellipsoids, the Stokes equations are, in general, too complex to solve exactly and, hence, approximations must be made. Two common classes of approximations are boundary element methods,[12,21,22] which provide convergent approximations to the Stokes equations and the no-slip boundary condition, and bead-type methods,[13,23] which provide only heuristic approximations.

In this article, we introduce a boundary element model for the stable and accurate computation of sequence-dependent hydrodynamic properties of short DNA fragments. We model the hydrated surface as a curved cylinder or tube of uniform radius with ends capped by hemispheres. The axis of the tube is a general space curve whose length and curvature are determined locally by the sequence. This axis is constructed using a rigid basepair model of double-helical DNA in which the relative displacement and orientation between adjacent basepairs (dimer step) is described by a set of six parameters depending on the dimer composition.[24–26] For the case of B-DNA, parameters for all possible dimer compositions have been estimated from crystallographic data.[27] Using these parameters, we construct an axis and a corresponding tubular surface which twists and bends in space according to the sequence. This rigid surface is expected to

[a)]Electronic mail: og@math.utexas.edu.

provide a reasonable approximation for DNA fragments with lengths from about 20 to 150 basepairs. At longer lengths, the rigidity assumption becomes more of an issue since the persistence length of DNA is about 150 basepairs.[28] At shorter lengths, we expect an atomistic-type surface which captures local fine-scale features to be more appropriate.[16] Studies on straight tube models have shown that, for lengths greater than about 20 basepairs, the details of the capped ends have a negligible effect on global hydrodynamic properties such as the translational and rotational diffusion coefficients, provided the end-to-end length of the axial curve is the same.[29–31] Thus a tube model with ends capped by hemispheres is expected to produce nearly identical results as a model with ends capped by disks provided the length of the axial curve is matched. We prefer to work with hemispherical caps since they lead to a better conditioned hydrodynamic problem because sharp edges are avoided.

The basic hydrodynamic problem is to solve the steady Stokes equations with no-slip boundary conditions in the three-dimensional domain exterior to the hydrated surface. To this end, we employ a new stable, accurate, and singularity-free boundary integral formulation described elsewhere.[32] The formulation is a Fredholm equation of the second kind and employs a mixed combination of the classic double- and single-layer potentials defined on parallel surfaces. In contrast to formulations of the first kind based on the single-layer potential alone,[15,21,22] the mixed formulation is not subject to ill conditioning in the limit of fine discretizations, and also avoids issues associated with null vectors, which pose a difficulty independent of the conditioning problem.[33] By virtue of being singularity-free, the mixed formulation can be discretized using standard high-order Gauss quadrature rules. Moreover, because parametrizations of the tubular surface geometry are available, all integrals can be formulated on the exact curved surface without a further approximation by flat elements. Thus curvature effects in both the circumferential and axial directions can be accurately resolved without the need for fine meshes or heuristic area correction factors.[21] Numerical experiments on a variety of tubular surfaces with relatively large curvatures indicate that global hydrodynamic quantities such as resistance matrices and diffusion coefficients can be computed with a typical accuracy of about 0.1% using moderate meshes. Compared with typical experimental errors in these quantities, the proposed numerical method can be viewed as delivering nearly exact results. Other formulations of the second kind are possible,[12,20,34,35] but these typically require the use of product integration rules[36,37] or specialized coordinate transformations and projections[38] to accurately approximate weakly singular integrals, and some also require that one or more distinguished points within the hydrated surface be identified for the placement of point singularities.[12,34]

For purposes of validation, we use our boundary element method to accurately compute diffusion coefficients for a straight tube model of DNA and compare our results with various empirical formulas[17,31] and experimental data.[39–45] A classic model for the hydrated surface of an arbitrary DNA sequence of length up to about 150 basepairs is that of a straight circular tube. The length of the tube is determined by the rise per basepair, whose average value is about 3.4 Å,[46] and the radius of the tube is a prescribed constant, whose value is estimated to be between 10 Å[46] and 13 Å.[17] The first estimate is based on x-ray diffraction data, whereas the second is based on experimental data in solution and empirical formulas for the diffusion coefficients. Using our boundary element method, we compute the translational and rotational diffusion coefficients for a straight tube model as a function of length (number of basepairs) for different radius values. For the translational diffusion coefficient we observe a good agreement between the boundary element results and the empirical formulas, but for the rotational diffusion coefficient we observe more significant differences. Our boundary element results confirm the previous observations that a hydrated radius of about 10–13 Å is consistent with the experimental data on the translational diffusion coefficient,[17] whereas a radius of 13–17 Å is consistent with the experimental data on the rotational diffusion coefficient. We surmise that this difference may largely be due to the limited validity of the straight tube assumption, which is a critical assumption in various experimental techniques for measuring rotational diffusion.[1,18]

To illustrate the effects of the sequence, we compute the translational and rotational diffusion coefficients as a function of length for families of random and periodic DNA sequences using the rigid basepair model with crystallographic data.[27] Our results show that while the data for the sequence-dependent curved tube model exhibit the same general trends as that for the straight tube model, there is a noticeable systematic offset due to curvature effects. For each value of the radius, the straight model uniformly underestimates both the translational and rotational diffusion coefficients over the entire range of lengths compared to the sequence-dependent curved model. This result suggests that sequence-dependent curvature can have a measurable impact on the diffusion coefficients, even for relatively short fragments of lengths less than about 150 basepairs, and that previous estimates of the hydrated radius are likely to be underestimates. The variability of the sequence-dependent data also suggests a possible method for refining rigid basepair model parameters for DNA in solution. In particular, from experimental data on known sequences of known length, the hydrodynamic model introduced here could be used to numerically fit the relative displacement and rotation parameters for all the different possible dimer steps, as well as the hydrated radius. Indeed, the possibility of refining these structural parameters has been a major motivation for this work.

The presentation is structured as follows. In Sec. II we outline a rigid basepair model of DNA, translate crystallographic data[27] to the parametrization used here, and describe our method for constructing a tubular surface model for a given DNA sequence. In Sec. III we briefly outline the hydrodynamic equations and numerical boundary element procedure for computing the resistance and mobility matrices and diffusion coefficients for a given surface. In Sec. IV we present numerical results for both the straight and the sequence-dependent curved tube models and compare them with previous results and experimental data. In Sec. V we end with some concluding remarks.

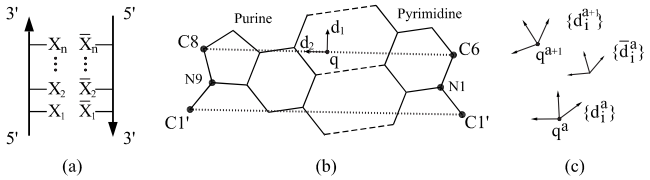

FIG. 1. Illustration of conventions used in the rigid basepair model of DNA. (a) Indexing scheme for basepairs $(X,\bar{X})_a$ $(a=1,\ldots,n)$. (b) Reference point q and frame $\{d_i\}$ for an arbitrary basepair $(X,\bar{X})$. (c) Relative rotation and displacement between adjacent basepairs $(X,\bar{X})_a$ and $(X,\bar{X})_{a+1}$ with associated middle frame $\{\bar{d}_i^a\}$.

II. MODEL

Here we outline a rigid basepair model for describing the three-dimensional sequence-dependent structure of a DNA molecule. We introduce the quantities necessary to define the configuration of a basepair, define coordinates for describing the relative rotations and displacements between adjacent basepairs, translate crystallographic data[27] to the parametrization used here, and describe our method for constructing a tubular surface model for a given DNA sequence.

A. Basepairs, configurations

We consider right-handed double-helical DNA in which bases T, A, C, and G are attached to two oriented antiparallel backbone strands and form only the standard Watson–Crick pairs (A,T) and (C,G). Choosing one backbone strand as a reference, a DNA molecule consisting of n basepairs is identified with a sequence of bases $X_1X_2\ldots X_n$, listed in the $5'-3'$ direction along the strand, where $X_a\in\{\mathrm{T},\mathrm{A},\mathrm{C},\mathrm{G}\}$. The basepairs associated with this sequence are denoted by $(X,\bar{X})_1,(X,\bar{X})_2,\ldots,(X,\bar{X})_n$, where $\bar{X}$ is defined as the Watson–Crick complement of X in the sense that $\bar{A}=T$ and $\bar{C}=G$. The notation $(X,\bar{X})_a$ for a basepair indicates that base X is attached to the reference strand while $\bar{X}$ is attached to the opposite strand as illustrated in Fig. 1(a).

We adopt a standard model of DNA[24–26] in which each basepair is modeled as a flat rigid object. The configuration of an arbitrary basepair $(X,\bar{X})_a$ is specified by giving the location of a reference point $q^a\in\mathbb{R}^3$ fixed in the basepair and the orientation of a right-handed orthonormal frame $\{d_i^a\}$, $d_i^a\in\mathbb{R}^3$ $(i=1,2,3)$, attached to the basepair. The reference point and the frame vectors are defined according to the Tsukuba convention.[26] The vector d_1^a points in the direction of the major groove along the perpendicular bisector of the $C1'$-$C1'$ axis of an ideal basepair, whereas d_2^a points in the direction of the reference strand and is parallel to the $C1'$-$C1'$ axis. As a result, $d_3^a=d_1^a\times d_2^a$ is perpendicular to the plane of $(X,\bar{X})_a$ and normally points in the direction of $(X,\bar{X})_{a+1}$. The reference point q^a is located at the intersection of the perpendicular bisector of the $C1'$-$C1'$ axis with the axis defined by the pyrimidine C6 and the purine C8 atoms as illustrated in Fig. 1(b).

There are four possible basepairs $(X,\bar{X})_a$ corresponding to the choice $X_a\in\{\mathrm{T},\mathrm{A},\mathrm{C},\mathrm{G}\}$. In a rigid basepair model, the positions of the nonhydrogen atoms in each of these basepairs with respect to q^a and $\{d_i^a\}$ are considered to be constant. As a result, once the reference point and frame of a basepair are specified, so too are the positions of all of its nonhydrogen atoms. Estimated values for these positions for basepairs in their ideal forms have been tabulated by Olson *et al.*[26] Thus the configuration of a DNA molecule consisting of n basepairs is completely defined by the reference points q^a and frames $\{d_i^a\}$ $(a=1,\ldots,n)$.

B. Rotation, displacement coordinates

In the present model, the three-dimensional shape of a DNA molecule is determined by the relative rotation and displacement between adjacent basepairs as illustrated in Fig. 1(c). The relative rotation and displacement between $(X,\bar{X})_a$ and $(X,\bar{X})_{a+1}$ can be described in the general form

$$d_j^{a+1}=\sum_{i=1}^{3}\Lambda_{ij}^a d_i^a,\quad q^{a+1}=q^a+\sum_{i=1}^{3}\eta_i^a\bar{d}_i^a, \tag{1}$$

where $\Lambda^a\in\mathbb{R}^{3\times3}$ is a rotation matrix which describes the orientation of frame $\{d_i^{a+1}\}$ with respect to $\{d_i^a\}$, $\eta^a\in\mathbb{R}^3$ is a coordinate vector which describes the position of q^{a+1} with respect to q^a, and $\{\bar{d}_i^a\}$ is a right-handed orthonormal frame between $\{d_i^a\}$ and $\{d_i^{a+1}\}$. The frame $\{\bar{d}_i^a\}$ is often referred to as a middle frame and will be defined below. From Eq. (1) we deduce that the entries in Λ^a and η^a are given by

$$\Lambda_{ij}^a=d_i^a\cdot d_j^{a+1},\quad \eta_i^a=\bar{d}_i^a\cdot(q^{a+1}-q^a). \tag{2}$$

A rotation matrix Λ^a can be parametrized by a coordinate vector $\theta^a\in\mathbb{R}^3$ in a variety of ways. In this work, we parametrize rotation matrices using the Cayley (also referred to as Euler–Rodrigues or Gibbs) formula[47]

$$\Lambda^a=\mathrm{cay}[\theta^a]:=I+\frac{4}{4+|\theta^a|^2}\left([\theta^a\times]+\frac{1}{2}[\theta^a\times]^2\right), \tag{3}$$

where I is the identity matrix and $[\theta^a\times]$ denotes the skew-symmetric matrix

$$[\theta^a\times]=\begin{pmatrix}0&-\theta_3^a&\theta_2^a\\ \theta_3^a&0&-\theta_1^a\\ -\theta_2^a&\theta_1^a&0\end{pmatrix}. \tag{4}$$

The Cayley formula can be explicitly inverted as

$$\theta^a=\mathrm{cay}^{-1}[\Lambda^a]:=\frac{2}{\mathrm{tr}[\Lambda^a]+1}\mathrm{vec}[\Lambda^a-(\Lambda^a)^T], \tag{5}$$

where $\mathrm{tr}[\Lambda^a]$ and $(\Lambda^a)^T$ denote the trace and the transpose of Λ^a and, for an arbitrary skew-symmetric matrix A, we define $\mathrm{vec}[A]=(A_{32},A_{13},A_{21})$. Equations (3) and (5) provide a one-to-one correspondence between rotation matrices Λ^a and coordinates θ^a, provided that $\mathrm{tr}[\Lambda^a]\neq-1$. Matrices for which $\mathrm{tr}[\Lambda^a]=-1$ can be shown to correspond to a rotation through π-radians (180°), which are unlikely to occur between adjacent basepairs in our application.

The Cayley parametrization of a rotation matrix has a straightforward geometrical interpretation. The matrix Λ^a in Eq. (3) corresponds to a right-handed rotation about a unit vector ν^a through an angle $\phi^a\in[0,\pi)$, where

$$\nu^a = \frac{1}{|\theta^a|}\sum_{i=1}^{3} \theta_i^a d_i^a, \tag{6a}$$

$$\phi^a = 2\ \arctan\left(\frac{|\theta^a|}{2}\right). \tag{6b}$$

From Eq. (6a) we deduce that a simple rotation about the frame vector d_1^a is obtained when $\theta^a=(\theta_1^a,0,0)$, where the angle of rotation is determined by Eq. (6b). Similar conclusions can be drawn for simple rotations about the other frame vectors. In general, a rotation about a given unit vector $m^a=\Sigma_{i=1}^3 \mu_i^a d_i^a$ through a given angle $\psi^a \in [0,\pi)$ is obtained when

$$\theta_i^a = 2\ \tan\left(\frac{\psi^a}{2}\right)\mu_i^a. \tag{7}$$

The middle frame $\{\bar{d}_i^a\}$ used to describe relative displacements can now be defined. Let Λ^a be the relative rotation matrix for frame $\{d_i^{a+1}\}$ with respect to $\{d_i^a\}$. Then the coordinates θ^a, axis ν^a, and angle ϕ^a associated with this rotation are as given in Eqs. (5), (6a), and (6b). The middle frame is here defined by a relative rotation about the same axis ν^a, but through an angle of $\phi^a/2$. Using Eq. (7) we obtain

$$\bar{d}_j^a = \sum_{i=1}^{3} \mathrm{cay}_{ij}[\bar{\theta}^a] d_i^a, \quad \bar{\theta}_i^a = 2\ \tan\left(\frac{\phi^a}{4}\right)\frac{\theta_i^a}{|\theta^a|}. \tag{8}$$

Thus the relative rotation and displacement between an arbitrary pair of adjacent basepairs $(X,\bar{X})_a$ and $(X,\bar{X})_{a+1}$ are completely described by the coordinates θ^a and η^a. The definitions above satisfy all the qualitative guidelines set forth in the Cambridge convention,[24] including the symmetry conditions associated with a change in reference strand. Accordingly, we refer to θ^a as tilt-roll-twist coordinates and η^a as shift-slide-rise coordinates. For a given configuration $\{q^a,d_i^a\}$ of basepair $(X,\bar{X})_a$, the above definitions establish a one-to-one correspondence between $\{\theta^a,\eta^a\}$ and the configuration $\{q^{a+1},d_i^{a+1}\}$ of basepair $(X,\bar{X})_{a+1}$. Notice that θ^a are not conventional angular coordinates as employed by many authors. Rather, they are abstract coordinates defined via the parametrization in Eq. (3). These abstract coordinates can be put into correspondence with conventional angular ones, and are nearly identical in the case of small rotations when the angular ones are measured in radians.

C. Crystallographic estimates

In the simplest model of sequence-dependent DNA structure, the coordinate set $\{\theta^a,\eta^a\}$ is completely determined by the composition of basepairs $(X,\bar{X})_a$ and $(X,\bar{X})_{a+1}$. Thus there are 16 different coordinate sets corresponding to the 16 possible basepair combinations defined by $X_a,X_{a+1} \in \{\mathrm{T},\mathrm{A},\mathrm{C},\mathrm{G}\}$. Crystal structures of B-form DNA were analyzed by Olson *et al.*[27] In that work, estimates for a set of relative rotation and displacement coordinates were reported for all of the 16 possible combinations of adjacent basepairs. These estimates were translated into the coordinates defined here and are summarized in Tables I and II as a function of

TABLE I. Estimates of tilt-roll-twist in the Cayley coordinates.

	T	A	C	G	
	0.00	−2.56	−0.18	−3.09	$\theta_1\times10^2$
A	1.98	1.28	1.27	8.17	$\theta_2\times10^2$
	5.23	6.32	5.64	5.72	$\theta_3\times10^1$
	2.56	0.00	2.76	−0.92	$\theta_1\times10^2$
T	1.28	6.09	3.49	8.66	$\theta_2\times10^2$
	6.32	6.85	6.56	6.76	$\theta_3\times10^1$
	0.18	−2.76	0.00	−0.18	$\theta_1\times10^2$
G	1.27	3.49	0.55	6.55	$\theta_2\times10^2$
	5.64	6.56	6.04	5.91	$\theta_3\times10^1$
	3.09	0.92	0.18	0.00	$\theta_1\times10^2$
C	8.17	8.66	6.55	9.92	$\theta_2\times10^2$
	5.72	6.76	5.91	6.52	$\theta_3\times10^1$

X_a (rows) and X_{a+1} (columns). Notice that only 10 of the 16 entries in each table are independent, as required by symmetry considerations based on the Watson–Crick pairing rules. In particular, each entry below the main diagonal in each table is a simple multiple of the corresponding entry above. Notice that the parameters θ^a in Table I are dimensionless, whereas the parameters η^a in Table II have dimensions of length.

In this work, we use Tables I and II to define the three-dimensional structure of a DNA molecule. Given a molecule with sequence $X_1X_2\ldots X_n$, there are $n-1$ dimer steps X_aX_{a+1}. To each such step we associate a unique coordinate set $\{\theta^a,\eta^a\}$ from the tables. For a given configuration $\{q^a,d_i^a\}$ of basepair $(X,\bar{X})_a$, the configuration $\{q^{a+1},d_i^{a+1}\}$ of basepair $(X,\bar{X})_{a+1}$ is determined by $\{\theta^a,\eta^a\}$. Thus the three-dimensional shape of a molecule is built up recursively beginning from an arbitrary choice of configuration for the first basepair. The three-dimensional structure constructed from the data in Tables I and II is to be interpreted as an approximation of the relaxed ground-state shape of a molecule. Sufficiently short molecules are expected to remain close to this shape under appropriate experimental conditions. At long lengths, the shape of a DNA molecule is expected to be highly variable due to its intrinsic flexibility.

D. Hydrated surface

We model the hydrated surface of a DNA molecule as a curved cylinder or tube of uniform radius with ends capped

TABLE II. Estimates of shift-slide-rise (Å) in the Cayley midframe.

	T	A	C	G	
	0.00	−0.35	1.35	0.90	$\eta_1\times10^1$
A	−5.95	−0.82	−5.82	−2.68	$\eta_2\times10^1$
	3.31	3.27	3.36	3.34	$\eta_3\times10^0$
	0.35	0.00	2.85	−0.94	$\eta_1\times10^1$
T	−0.82	0.31	0.79	5.04	$\eta_2\times10^1$
	3.27	3.42	3.37	3.33	$\eta_3\times10^0$
	−1.35	−2.85	0.00	0.56	$\eta_1\times10^1$
G	−5.82	0.79	−3.81	−2.37	$\eta_2\times10^1$
	3.36	3.37	3.40	3.42	$\eta_3\times10^0$
	−0.90	0.94	−0.56	0.00	$\eta_1\times10^1$
C	−2.68	5.04	−2.37	3.80	$\eta_2\times10^1$
	3.34	3.33	3.42	3.39	$\eta_3\times10^0$

by hemispheres. The axis of this tube is a space curve defined by the basepair reference points. For a given radius r and sequence $S=X_1X_2\ldots X_n$, we denote the tubular surface by $\Gamma(r,S)$ and the axial curve by $\gamma(S)$. We model $\gamma(S)$ as a biarc curve,[48] which is a space curve formed by a concatenation of circular arcs. A main motivation for this choice is that it leads to a simple construction of the surface $\Gamma(r,S)$. In particular, $\Gamma(r,S)$ is constructed by superimposing a toroidal segment of radius r on each arc of $\gamma(S)$.

In the simplest approach, the curve $\gamma(S)$ is defined by interpolating the basepair reference points q^a $(a=1,\ldots,n)$. However, to maintain smoothness of the surface $\Gamma(r,S)$, the maximum curvature of the curve $\gamma(S)$ must be bounded by $1/r$, otherwise the surface $\Gamma(r,S)$ will locally self-intersect and develop kinks. While such irregularities in $\Gamma(r,S)$ would likely have a negligible effect on global hydrodynamic quantities, they can significantly increase the computational effort required to solve the hydrodynamic equations. Thus the problem of defining $\gamma(S)$ can be described as that of finding a biarc curve which best fits the reference points q^a $(a=1,\ldots,n)$ subject to the restriction $\kappa_{\gamma(S)}\leq 1/r$, where $\kappa_{\gamma(S)}$ denotes the maximum curvature of $\gamma(S)$.

Guided by the above characterization we devised an iterative procedure for determining the axial curve $\gamma(S)$. Given a control parameter $0<\zeta<1$, the procedure begins with an initial set of nodes $q_k^a=q^a$, constructed as described in Sec. II, and an interpolating curve $\gamma_k(S)$, defined as the unique biarc interpolant of q_k^a,[48] with tangent vectors computed from a cubic spline. If $\kappa_{\gamma_k(S)}\leq\zeta/r$, the procedure is terminated and $\gamma(S)=\gamma_k(S)$. Otherwise, segments of high curvature on $\gamma_k(S)$ are identified, and the indices of all interior high curvature nodes are collected into a set J_k. To smoothen the curve we perform the local Gaussian convolution

$$q_{k+1}^a=\begin{cases}\frac{1}{4}q_k^{a-1}+\frac{1}{2}q_k^a+\frac{1}{4}q_k^{a+1}, & a\in J_k\\ q_k^a, & a\notin J_k.\end{cases}$$

The nodes q_{k+1}^a are then interpolated as before to obtain $\gamma_{k+1}(S)$. The procedure is repeated until the curvature condition is satisfied. By design, the procedure smooths local fine-scale kinks of the axial curve while preserving large-scale features, where the scale is given by r. In practice, we used the parameter value $\zeta=1/2$ and found that for realistic values of r (for example, $r=13$ Å) the procedure is terminated for most sequences after only one or two iterations. The resulting smoothened axial curves were found to differ from the unsmoothened curves by less than 0.1% in the l_2-norm. The result of this procedure is a biarc curve $\gamma(S)$ which closely follows the path of the basepair reference points q^a (see Fig. 2). The curve begins at q^1, ends at q^n, and satisfies the curvature condition $\kappa_{\gamma(S)}\leq 1/r$.

The surface $\Gamma(r,S)$ is defined by superimposing a toroidal segment of radius r on each circular arc of the biarc curve $\gamma(S)$. We close the ends of $\Gamma(r,S)$ using hemispherical caps of radius r. In the simplest approach, the hemispheres would be centered at the points q^1 and q^n, which would cause the surface to extend beyond the first and last basepairs. To eliminate this effect, we translate the centers of the hemispheres along $\gamma(S)$ by a distance r from each end. Thus the points q^1 and q^n are located, approximately, at the apexes of the hemispheres. The resulting surface $\Gamma(r,S)$ can be shown to be continuously differentiable because $\gamma(S)$ is continuously differentiable and satisfies the curvature condition, and because the ends of $\Gamma(r,S)$ are capped by hemispheres. The overall procedure is illustrated in Fig. 2. For reference, the figure also shows a straight tube model $\Gamma_{\mathrm{str}}(r,S)$. The straight tube is constructed in a similar way from equally spaced reference points placed along a line with spacing 3.4 Å, which corresponds to a commonly accepted value for the average rise per basepair.[46] This value is in close agreement with the average value of rise obtained from Table II, which is 3.36 Å.

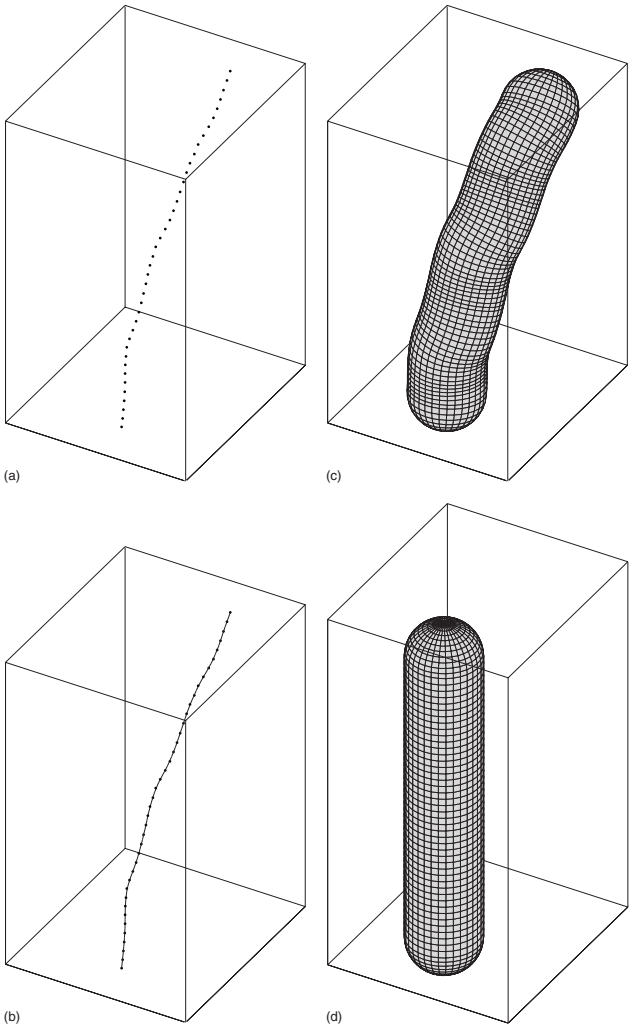

FIG. 2. Illustration of curved and straight tube models for the 40 basepair sequence $(CTTTTAAAGAG)_3CTTTTAA$ with radius $r=13$ Å. (a) Reference points q^a. (b) Reference points q^a and smoothened axial curve $\gamma(S)$. (c) Curved tube model $\Gamma(r,S)$. (d) Straight tube model $\Gamma_{\mathrm{str}}(r,S)$.

III. THEORY

Here we briefly outline the standard theory for modeling the hydrodynamics of a rigid molecule in an incompressible Newtonian fluid at infinite dilution. We state the basic flow equations, define various associated flow quantities, and introduce the resistance and mobility matrices and diffusion coefficients for a rigid body of arbitrary shape. We discuss

various properties of the diffusion coefficients relevant to our study and then briefly outline our numerical procedure, which is described in detail elsewhere.

A. Flow equations

We consider the slow diffusive motion of a relatively large rigid macromolecular body in an incompressible Newtonian fluid of absolute viscosity $\mu>0$. In a body-fixed frame, we denote the hydrated surface by Γ, the interior body domain by B, and the exterior fluid domain by B_e. Given a velocity field $v:\Gamma\to\mathbb{R}^3$, we seek to determine the hydrodynamic loads exerted by the fluid on the body. According to standard theory,[19,20] these loads are determined by the fluid velocity field $u:B_e\to\mathbb{R}^3$ and pressure field $p:B_e\to\mathbb{R}$ which satisfy the classic Stokes equations

$$\mu u_{i,jj}-p_{,i}=0, \quad x\in B_e, \tag{9a}$$

$$u_{i,i}=0, \quad x\in B_e, \tag{9b}$$

$$u_i=v_i, \quad x\in\Gamma, \tag{9c}$$

$$u_i,p\to 0, \quad |x|\to\infty. \tag{9d}$$

Equation (9a) is the local balance law of linear momentum for the fluid and Eq. (9b) is the local incompressibility constraint. Equation (9c) is the no-slip boundary condition which states that the fluid and the body velocities coincide at each point of the hydrated surface. (For the case of small macromolecules, a different type of boundary condition may be more appropriate.[21,49]) The limits in Eq. (9d) are boundary conditions that are consistent with the fluid being at rest at infinity. Unless mentioned otherwise, all vector quantities are referred to a body-fixed frame and indices take values from 1 to 3. Here and throughout we will use the usual conventions that a pair of repeated indices implies summation, and that indices appearing after a comma denote partial derivatives.

B. Body velocities, hydrodynamic loads

When the body B is rigid, the boundary velocity field v in Eq. (9c) takes the general form

$$v=V+\Omega\times(x-c), \tag{10}$$

where $\times$ denotes the cross product. Here V is the linear velocity of a given reference point c, and Ω is the angular velocity of the body-fixed frame. Under mild assumptions on the surface Γ,[33] the system in Eqs. (9a)–(9d) has a unique solution (u,p) for given (V,Ω). The fluid stress field $\sigma:B_e\to\mathbb{R}^{3\times 3}$ associated with this solution is defined by

$$\sigma_{ij}=-p\delta_{ij}+\mu(u_{i,j}+u_{j,i}), \tag{11}$$

where δ_{ij} is the Kronecker delta symbol, and the traction field $f:\Gamma\to\mathbb{R}^3$ exerted by the fluid on the body surface (force per unit area) is defined by

$$f=\sigma\nu, \tag{12}$$

where ν is the outward unit normal on Γ. The hydrodynamic loads of the fluid on the body are found by integrating this traction field over the body surface. In particular, the resultant force F and torque T, about the reference point c, are given by

$$F=\int_\Gamma f(x)dA_x, \quad T=\int_\Gamma (x-c)\times f(x)dA_x, \tag{13}$$

where dA_x denotes an infinitesimal area element at $x\in\Gamma$. The resultant loads (F,T) are purely hydrodynamic in the sense that they vanish when the body velocities (V,Ω) vanish. Resultant loads due to hydrostatic effects, that is, buoyancy loads caused by a body force field, may be accounted for separately.

C. Resistance, mobility matrices

Linearity of Eqs. (9)–(13) implies the existence of matrices $L_a\in\mathbb{R}^{3\times 3}$ $(a=1,\ldots,4)$ such that

$$F=-L_1V-L_3\Omega, \quad T=-L_2V-L_4\Omega. \tag{14}$$

These matrices are called the hydrodynamic resistance matrices associated with the body B. They are intrinsic properties of the shape of B and are proportional to the fluid viscosity μ. With the exception of L_1, they depend on the choice of reference point c appearing in Eqs. (10) and (13).[19,20,50] Self-adjointness of the differential operator associated with the system in Eqs. (9a)–(9d) implies that the overall resistance matrix

$$L=\begin{pmatrix} L_1 & L_3 \\ L_2 & L_4 \end{pmatrix} \tag{15}$$

is symmetric.[19,20,51] Furthermore, an energy dissipation inequality associated with the system in Eqs. (9a)–(9d) implies that L is also positive-definite.

Throughout our developments we denote the inverse of L by M so that

$$M=\begin{pmatrix} M_1 & M_3 \\ M_2 & M_4 \end{pmatrix}=\begin{pmatrix} L_1 & L_3 \\ L_2 & L_4 \end{pmatrix}^{-1}. \tag{16}$$

The block entries $M_a\in\mathbb{R}^{3\times 3}$ are called the mobility matrices associated with B and are related to the resistance matrices $L_a\in\mathbb{R}^{3\times 3}$ through the expressions

$$M_1=L_1^{-1}+L_1^{-1}L_3S^{-1}L_2L_1^{-1}, \quad M_3=-L_1^{-1}L_3S^{-1},$$

$$M_2=-S^{-1}L_2L_1^{-1}, \quad M_4=S^{-1}, \tag{17}$$

where $S=L_4-L_2L_1^{-1}L_3$. Since L is symmetric positive-definite, we find that M is also symmetric positive-definite. In particular, the block entries L_1, L_4, M_1, and M_4 are all symmetric positive-definite. It can be shown that, with the exception of M_4, the mobility matrices also depend on the choice of the reference point c.[50] In view of Eqs. (14) and (16), we have

$$V=-M_1F-M_3T, \quad \Omega=-M_2F-M_4T. \tag{18}$$

The resistance matrix L and the mobility matrix M provide a fundamental relation between velocities and loads for a rigid body immersed in a Newtonian fluid. For a body subject to prescribed velocities (V,Ω), the resultant hydrodynamic loads (F,T) predicted by the solution of the system in Eqs. (9a)–(9d) are unique and given by $(F,T)=-L(V,\Omega)$.

Conversely, for a body subject to prescribed hydrodynamic loads (F,T), the resulting velocities (V,Ω) must also be unique and given by $(V,\Omega)=-M(F,T)$.

The matrices L and M satisfy a simple transformation law under a change in reference point. Let c and c' be two arbitrary reference points with associated matrices (L,M) and (L',M'), and let $r=c'-c$ be the vector from c to c'. Then[50]

$$L' = Q^T L Q, \quad M' = Q^{-1} M Q^{-T}, \tag{19}$$

where Q and its inverse Q^{-1} are matrices given in block form by, employing the notation from Sec. II B,

$$Q = \begin{pmatrix} I & [r\times] \\ 0 & I \end{pmatrix}, \quad Q^{-1} = \begin{pmatrix} I & -[r\times] \\ 0 & I \end{pmatrix}. \tag{20}$$

D. Diffusion, mobility coefficients

As derived by Brenner,[52,53] the translational and rotational diffusion coefficients for a rigid body immersed in a Newtonian fluid are

$$D_t = kT\vartheta_t, \quad D_r = kT\vartheta_r, \tag{21}$$

where k is the Boltzmann constant, T is the absolute temperature of the fluid, and ϑ_t and ϑ_r are mobility coefficients defined by

$$\vartheta_t = \tfrac{1}{3}\mathrm{tr}(M_1), \quad \vartheta_r = \tfrac{1}{3}\mathrm{tr}(M_4). \tag{22}$$

The diffusion coefficients D_t and D_r quantify the mean-square linear and angular displacements of the body per unit time in free Brownian motion. In contrast, the mobility coefficients ϑ_t and ϑ_r quantify the mean linear and angular velocities of the body per unit load under the assumption of hydrodynamic force and torque balance. The relations in Eq. (21) can be viewed as generalizations to a rigid body of arbitrary shape of the classic Stokes–Einstein relations for bodies of spherical or ellipsoidal shape.[18,54,55]

The diffusion coefficients D_t and D_r play central roles in the interpretation of various different types of experimental data. The orientational averages implicit in the above definitions account for the rotational aspect of the Brownian motion of a macromolecule in solution. In experiments in which a macromolecule has no preferred orientation, these orientationally averaged coefficients are expected to provide accurate descriptions of translational and rotational diffusive motions. (Although only one such mode of motion may be measurable in any given experiment.) On the other hand, in experiments in which a macromolecule has a preferred orientation, the above coefficients would not be expected to provide accurate descriptions unless M_1 and M_4 are nearly isotropic, as in the case of a nearly spherical macromolecule. When M_1 and M_4 are far from isotropic, the translational and rotational diffusive motions of a macromolecule with a preferred orientation are better described by individual components of M_1 and M_4 which correspond to the directions in which diffusion is measured.

For the analysis of translational diffusion data we use the orientationally averaged coefficient D_t defined in Eq. (21). However, due to the nature of the corresponding experimental data, for the analysis of rotational diffusion we use a transversely averaged coefficient defined by

$$D_r^{\perp} = kT\vartheta_r^{\perp}, \quad \vartheta_r^{\perp} = \tfrac{1}{2}\mathrm{tr}^{\perp}(M_4), \tag{23}$$

where $\mathrm{tr}^{\perp}(M_4)$ is the sum of the smallest two eigenvalues of M_4. This coefficient characterizes diffusive rotational motion which occurs transverse to the axis of a nearly cylindrical body. In particular, this average explicitly excludes the largest eigenvalue of M_4, which characterizes diffusive rotational motion parallel to the axis. The definition in Eq. (23) can be viewed as a natural generalization to the case of curved cylinders of the transverse rotational diffusion coefficient of straight cylinders.[12,17]

In this work, we follow standard convention and refer all diffusion coefficients to the center of diffusion.[12,50] This point is the unique point at which the mobility matrices M_2 and M_3 are equal and, consequently, symmetric because the overall matrix M is symmetric. It also has the interesting property that, among all points in space, it is the unique point at which the coefficients ϑ_t and ϑ_r achieve minimum values.[12,50] (This is trivially true for ϑ_r since it is independent of the reference point.) The center of diffusion can also be described as that point at which the translational and rotational diffusive motions are most independent. For symmetric bodies such as spheres, ellipsoids, and straight cylinders, the center of diffusion is coincident with the center of volume. However, for bodies of arbitrary shape, these two points are generally distinct. The use of the center of diffusion as the reference point guarantees that the predicted diffusion coefficients for a body are not overstated.

The center of diffusion of a body can be determined provided the mobility matrix is known at some given reference point. In particular, if c is a given reference point with associated mobility matrix M, then the center of diffusion c' is given by, employing the notation from Sec. II B,

$$c' = c + (\mathrm{tr}[M_4]I - M_4)^{-1}\,\mathrm{vec}(M_3 - M_3^T). \tag{24}$$

This expression is equivalent to those in previous works,[12,50] but may be more convenient because it avoids the principal axes frame employed there. Once c' is known, the associated mobility matrix M' can be found using Eq. (19), and the associated diffusion coefficients D_t' and $D_r'^{\perp}$ can be obtained from relations analogous to Eqs. (21) and (23). The resulting expressions take the form

$$D_r'^{\perp} = D_r^{\perp},$$

$$D_t' = D_t - \frac{kT}{3}\mathrm{vec}(M_3 - M_3^T)\cdot(\mathrm{tr}[M_4]I - M_4)^{-1}\,\mathrm{vec}(M_3 - M_3^T). \tag{25}$$

E. Numerical procedure

The basic hydrodynamic problem is to solve the steady Stokes system in Eqs. (9a)–(9d) in the three-dimensional domain exterior to the hydrated surface $\Gamma(r,S)$ of a given DNA molecule. To this end, we employ a boundary element method described in detail elsewhere.[32] Choosing an arbi-

trary body-fixed frame for the surface $\Gamma(r,S)$, we first use a numerical integration procedure to determine its center of volume, which we take as our initial reference point c. For a given linear velocity V and angular velocity Ω, the boundary element method yields an approximate resultant hydrodynamic force F and torque T. From the relation $(F,T)=-L(V,\Omega)$, we deduce that the columns of $-L$ are the resultant loads corresponding to unit velocities, that is,

$$-L=\left(\left.\begin{pmatrix}F\\T\end{pmatrix}\right|_{\substack{V=(1,0,0)\\ \Omega=(0,0,0)}} \cdots \left.\begin{pmatrix}F\\T\end{pmatrix}\right|_{\substack{V=(0,0,0)\\ \Omega=(0,0,1)}}\right). \tag{26}$$

Thus the determination of L for $\Gamma(r,S)$ requires six independent computations. From L we compute the mobility matrix $M=L^{-1}$ by matrix inversion, and then compute the center of diffusion c', the associated mobility matrix M', and the diffusion coefficients D_t' and $D_r'^{\perp}$ as described above.

The boundary element method we employ is defined by four parameters: the size h of curved quadrilateral elements used to represent the surface $\Gamma(r,S)$, the number q^2 of Gauss quadrature points used in each element, and two parameters $\lambda\in(0,1)$ and $\phi\in(0,\phi_{\Gamma(r,S)})$ which control the conditioning of the method. Here $\phi_{\Gamma(r,S)}$ is a purely geometrical parameter which in the present case is equal to r. Previous studies[32] indicate that accurate results on tubular geometries can be obtained with moderate values of h using $q=1$, $\lambda=1/2$, and $\phi/\phi_{\Gamma(r,S)}=1/2$. In our numerical experiments below, we choose an h value that is small enough so that the relative errors in computed hydrodynamic quantities are below 0.1% uniformly over the entire range of lengths considered. Thus numerical errors can be considered negligible.

IV. RESULTS

Here we use our boundary element method to numerically compute translational and rotational diffusion coefficients for the straight and sequence-dependent curved tube models introduced in Sec. II. For convenience, all results are presented in terms of the dimensionless coefficients

$$\bar{D}_t'=\frac{\mu\ell D_t'}{kT},\quad \bar{D}_r'^{\perp}=\frac{\mu\ell^3 D_r'^{\perp}}{kT}, \tag{27}$$

where ℓ is a characteristic length scale equal to 34 Å. We compute diffusion coefficients for straight tubes and compare our results with various empirical formulas and experimental data. We then compute diffusion coefficients for curved tubes corresponding to random and periodic DNA sequences and compare these results with those for straight tubes.

A. Straight tube model

Figure 3 illustrates convergence results for our numerical method as a function of the element size parameter h and the number of basepairs n for a straight tube of fixed radius. The parameter h is proportional to the average size of a quadrilateral element in a given mesh. The results summarized in this figure were used to select and justify the numerical parameters $(q,\lambda,\phi/\phi_\Gamma)$ and the element size h that were used in all subsequent calculations.

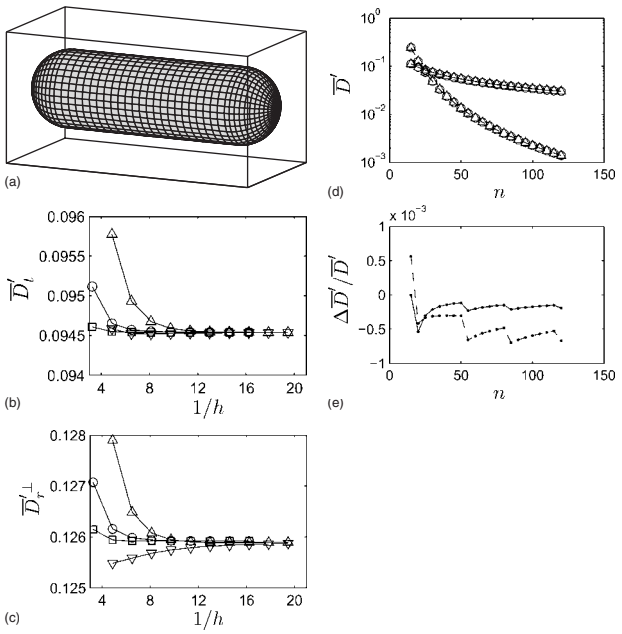

FIG. 3. Convergence results for straight tubes of fixed radius. (a) Tube with $r=10$ Å and $n=20$ discretized with elements of reference size h_*. [(b) and (c)] $\bar{D}_t'$ and $\bar{D}_r'^{\perp}$ vs $1/h$ for tube shown in (a). The scale on the horizontal axis is such that $1/h_*=18$. Upward-pointing triangles and circles denote results for $q=1$ and $q=2$ with $(\lambda,\phi/\phi_\Gamma)=(1/2,1/4)$. Downward-pointing triangles and squares denote results for $q=1$ and $q=2$ with $(\lambda,\phi/\phi_\Gamma)=(1/2,1/2)$. (d) $\bar{D}_t'$ (solid) and $\bar{D}_r'^{\perp}$ (dashed) vs n for tubes with $r=10$ Å. Triangles denote results for elements of size h_*. Circles denote results for elements of size $h_{**}\doteq h_*/2$. Both results were obtained with $q=1$ and $(\lambda,\phi/\phi_\Gamma)=(1/2,1/2)$. (e) $\Delta\bar{D}_t'/\bar{D}_t'$ (solid) and $\Delta\bar{D}_r'^{\perp}/\bar{D}_r'^{\perp}$ (dashed) vs n, where $\Delta\bar{D}'$ is the difference between results with element sizes h_* and h_{**}.

Plot (a) in Fig. 3 illustrates a straight tube model with $r=10$ Å, $n=20$, and a moderate element size h_*, which we take as a reference. Plots (b) and (c) show convergence results for the translational and rotational diffusion coefficients for the tube shown in (a). Results are given for two different $q\times q$ Gauss quadrature rules: $q=1$ and $q=2$. Moreover, results are given for two different pairs of conditioning parameters: $(\lambda,\phi/\phi_\Gamma)=(1/2,1/4)$ and $(\lambda,\phi/\phi_\Gamma)=(1/2,1/2)$. For reference, the scale on the horizontal axis is such that $1/h_*=18$. The data illustrate that, for the different choices of quadrature rule and conditioning parameters, the diffusion coefficients converge to limiting values on moderately refined meshes. The convergence characteristics are generally better for $q=2$ than for $q=1$, and are generally better for $(\lambda,\phi/\phi_\Gamma)=(1/2,1/2)$ than for $(\lambda,\phi/\phi_\Gamma)=(1/2,1/4)$. Because larger values of q require increased computational effort, we select the parameters $q=1$ and $(\lambda,\phi/\phi_\Gamma)=(1/2,1/2)$ for all subsequent calculations.

Plot (d) of Fig. 3 shows the diffusion coefficients $\bar{D}_t'$ and $\bar{D}_r'^{\perp}$ as a function of the number of base pairs n for a straight tube with radius $r=10$ Å. Results are given for two element sizes: h_* and h_{**} (h_{**} is approximately half of h_*). The data show that, for the entire range of n considered, reducing the element size by a factor of about 2 causes no significant changes in $\bar{D}_t'$ and $\bar{D}_r'^{\perp}$. Plot (e) shows the relative differences in each of $\bar{D}_t'$ and $\bar{D}_r'^{\perp}$ obtained with element sizes h_* and h_{**} as a function of n. The data show that the relative

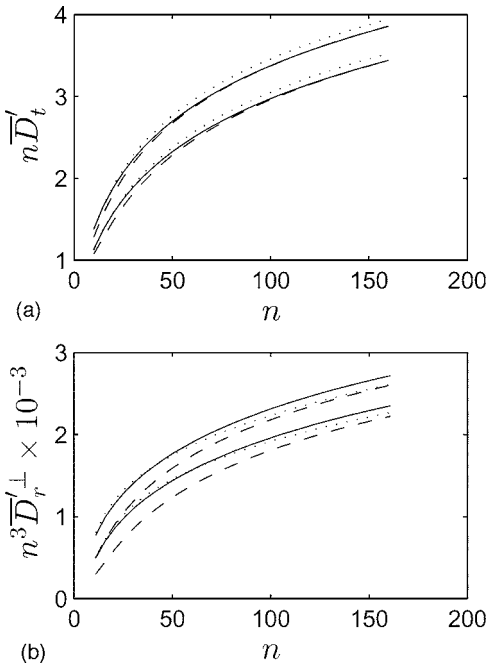

FIG. 4. Comparison of boundary element computations with empirical formulas for the diffusion coefficients of a straight tube. [(a) and (b)] Comparisons for $\bar{D}'_t$ and $\bar{D}'^{\perp}_r$. Solid curves correspond to boundary element computations, dashed curves correspond to the empirical formulas derived by Tirado *et al.* (Ref. 17), and dotted curves correspond to the empirical formulas derived by Yoshizaki and Yamakawa (Ref. 31). In each plot, the upper group of curves correspond to results for a tube of radius $r=10$ Å, and the lower group corresponds to results for $r=15$ Å.

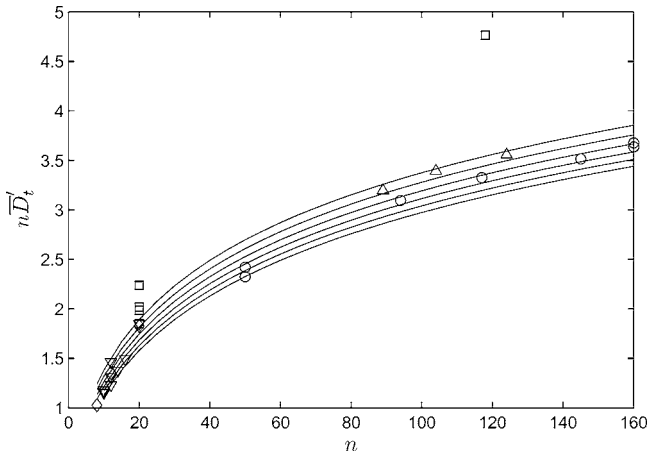

FIG. 5. Comparison of boundary element computations and experimental data on the translational diffusion coefficient for DNA sequences of various lengths. The solid curves denote boundary element results from the straight tube model with different radii: $r=10,11,\ldots,15$ Å ordered from upper to lower curve. The different open symbols denote different sets of experimental data: downward triangles (Ref. 39), diamonds (Ref. 40), circles (Ref. 43), upward triangles (Ref. 44), and squares (Ref. 45).

difference in each coefficient is well under 0.1% uniformly in n. Based on these results, we expect the element size h_* to be sufficiently small for tubes with $r \geq 10$ Å for the range of n we consider. Notice that tubes with larger values of r would have lower circumferential curvature, which would be even more adequately resolved by elements of size h_*. While the above results pertain to straight tubes, similar convergence results and accuracies hold for curved tubes with moderate axial curvatures.[32]

Figure 4 presents a comparison between our boundary element computations and two sets of empirical formulas[17,31] for the diffusion coefficients of a straight circular tube. Shown in the figure are plots of $n\bar{D}'_t$ and $n^3\bar{D}'^{\perp}_r$ versus n for two values of the radius: $r=10$ Å and $r=15$ Å. We choose to plot $n\bar{D}'_t$ and $n^3\bar{D}'^{\perp}_r$ rather than $\bar{D}'_t$ and $\bar{D}'^{\perp}_r$ since they enhance the sensitivity and allow results for different radius values to be more easily distinguished. The empirical formulas derived by Tirado *et al.*[17] are based on a curve fit of numerical results obtained from a bead-type model in which a cylinder is represented by a stack of rings of beads. In that model, the cylinders are left hollow and open, with no cap or plate at either end. The empirical formulas derived by Yoshizaki and Yamakawa[31] are based on a combination of numerical results and slender-body theory.[56,57] In that work, interpolation formulas were derived which were consistent with slender-body theory in the limit of long lengths and with numerical boundary element results in the limit of short lengths (the boundary element data were limited to tubes with length to diameter ratios $L/d<5$, or equivalently, $n<38$ basepairs on our scale).

The data in Fig. 4 show that, for each of the two values of r, our boundary element computations are in general agreement with the empirical formulas for $\bar{D}'_t$, but that more significant differences are visible between all three approaches for $\bar{D}'^{\perp}_r$. By design, the empirical formulas derived by Yoshizaki and Yamakawa[31] closely follow our boundary element results at shorter lengths, but show some slight discrepancy at longer lengths. This discrepancy is likely due to the approximations inherent in slender-body theory, and would presumably decrease in the limit of long length where slender-body theory is applicable. In contrast, the empirical formulas derived by Tirado *et al.*[17] are in better agreement with our results at longer lengths than at shorter lengths. Because of the fundamental differences between the bead-type and Stokes models, it is ultimately difficult to understand the differences observed in the figure. Nevertheless, the results suggest that the bead-type model provides a reasonable approximation of the more precise Stokes model at longer lengths where end effects are less important.

Figure 5 shows a comparison between our boundary element computations and experimental data on the translational diffusion coefficient for DNA sequences of various lengths. The curves in the figure are the results of our boundary element computations using a straight tube model with six different values of the radius: $r=10,11,\ldots,15$ Å. The symbols in the figure correspond to different sets of experimental data obtained by different techniques: velocity sedimentation,[39,43] dynamic light scattering,[40,44] and capillary electrophoresis.[45] Differences due to experimental conditions are expected to be minimal since all the data were normalized to standard conditions.[18] The observed disagreement between the electrophoresis data and those from other techniques is at present not well understood.[45] With the exception of this electrophoresis data, nearly all the experimental values fall within the region covered by the computed curves.

The experimental data in Fig. 5 for sequences of lengths

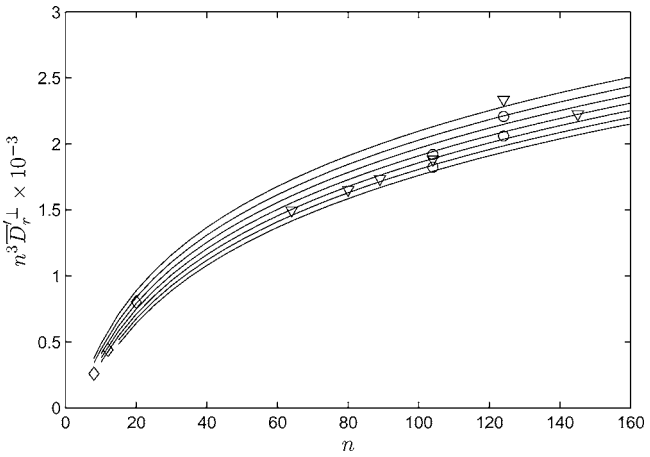

FIG. 6. Comparison of boundary element computations and experimental data on the rotational diffusion coefficient for DNA sequences of various lengths. The solid curves denote boundary element results from the straight tube model with different radii: $r=12,13,\ldots,18$ Å ordered from upper to lower curve. The different open symbols denote different sets of experimental data: diamonds (Ref. 40), downward triangles (Ref. 41), and circles (Ref. 42).

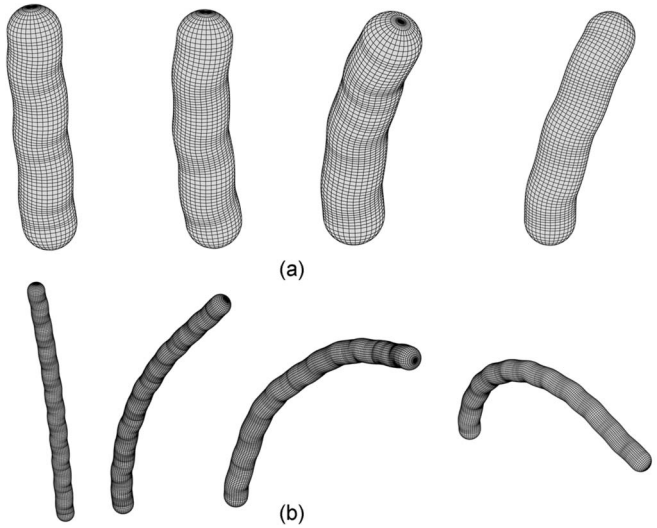

FIG. 7. Sequence-dependent curved tube models of different DNA sequences. (a) Four different sequences of $n=40$ basepairs. (b) Four different sequences of $n=120$ basepairs.

greater than 50 basepairs can be seen to be consistent with two distinct radius ranges. The light scattering data[44] are closely consistent with $r=10-11$ Å and the sedimentation data[43] are consistent with $r=12-13$ Å. This difference could be a characteristic of the different experimental techniques, or simply the result of experimental error. In this respect, it is important to note that the reported sequence lengths are themselves subject to error. Alternatively, the difference might be due to sequence effects, for example, sequence-dependent axial curvature, which are ignored in the interpretation of the experimental measurements. The data for sequences of about 50 basepairs or less show more scatter relative to the computed curves, although they are still generally consistent. In this short length range, we expect that any sequence-dependent curvature effect would play a lesser role and that local irregularities in the actual hydrated surface would become increasingly important. As a result, a straight tube model of uniform radius with ends capped by hemispheres may be overly simplistic. Indeed, for extremely short sequences, we expect that an atomistic-type surface model is likely more appropriate.

Figure 6 shows a comparison between our boundary element computations and experimental data on the rotational diffusion coefficient for DNA sequences of various lengths. The curves in the figure are the results of our boundary element computations using a straight tube model with seven different values of the radius: $r=12,13,\ldots,18$ Å. The symbols in the figure correspond to different sets of experimental data obtained by different techniques: dynamic light scattering[40] and transient electric birefringence.[41,42] As before, the data are normalized to standard conditions.[18] Notice that nearly all the experimental values fall within the region covered by the computed curves, but here the computed curves correspond to larger values of r. Thus, in contrast to the translational data, the bulk of the rotational data is consistent with $r=13-17$ Å. This difference can likely be attributed to sequence-dependent axial curvature, which is ignored in the interpretation of the experimental measurements, and experimental errors, which are expected to be higher since rotational diffusion coefficients are more difficult to measure. However, the systematic nature of the difference toward larger values of r suggests that a systematic effect such as curvature is more likely than experimental error.

B. Sequence-dependent curved tube model

Figure 7 shows curved tube models for various different DNA sequences. Shown are models for four different 40 basepair sequences and four different 120 basepair sequences, all with a tube radius of $r=13$ Å. These tube models were constructed as described in Sec. II using the crystallographic data in Tables I and II. The figure illustrates the possible variability in overall axial curvature due to sequence effects.

To study the effects of sequence on the diffusion coefficients as a function of n, we considered sample sets of the form $\Sigma_n=\{S_n^1,\ldots,S_n^m\}$, where m is a fixed number and each S_n^j is a sequence of length n. For different values of n the sample Σ_n was constructed by a culling procedure. We first constructed a set $\Sigma_n'=\{S_n^1,\ldots,S_n^{m'}\}$, $m'\gg m$, consisting of an arbitrary mixture of completely random sequences, periodic sequences with random repeating units, and periodic sequences containing A-tracts of varying lengths. The radius of gyration R of each $S_n^j\in\Sigma_n'$ was computed and the minimum and maximum values $R_{\min}$ and $R_{\max}$ were recorded. The set Σ_n was then constructed by selecting appropriate elements of Σ_n' to obtain an approximately uniform distribution in R between $R_{\min}$ and $R_{\max}$. In our numerical experiments, we considered lengths $15\leq n\leq 120$ (in increments of 5), and for each value of n we constructed sample sets of size $m=20$. The four sequences shown in Fig. 7(a) are typical elements from the sample set Σ_{40} and the four sequences shown in Fig. 7(b) are typical elements from the sample set Σ_{120}.

Figure 8 shows the translational diffusion coefficient of each sequence in the sample set Σ_n for each value of n com-

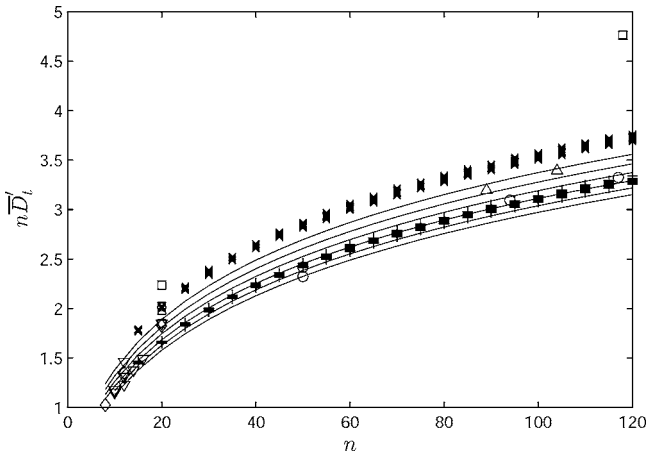

FIG. 8. Comparison of numerical results from sequence-dependent curved tube model and straight tube model for translational diffusion coefficient. Cross and plus symbols denote results from the sequence-dependent curved model with different radii: (crosses) $r=10$ Å and (pluses) $r=15$ Å. The solid curves denote results from the straight tube model with different radii: $r=10,11,\ldots,15$ Å ordered from upper to lower curve. The open symbols denote experimental data as in Fig. 5.

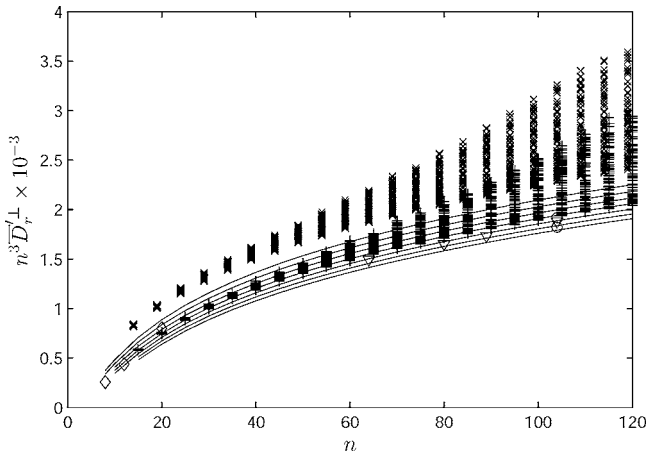

FIG. 9. Comparison of numerical results from sequence-dependent curved tube model and straight tube model for rotational diffusion coefficient. Cross and plus symbols denote results from the sequence-dependent curved model with different radii: (crosses) $r=12$ Å, (pluses) $r=18$ Å. For clarity, the crosses have been shifted horizontally one unit to the left. The solid curves denote results from the straight tube model with different radii: $r=12,13,\ldots,18$ Å ordered from upper to lower curve. The open symbols denote experimental data as in Fig. 6.

puted using the curved tube model. Results are shown for two values of the radius: $r=10$ Å and $r=15$ Å. For comparison, the six curves and the experimental data from Fig. 5 are included for the length range considered here. It can be seen that for $r=10$ Å and for $r=15$ Å the diffusion coefficients obtained from the straight tube model are uniformly below those obtained from the sequence-dependent curved tube model. Similar results are expected for each of the intermediate radius values. Thus, for given values of n and r, the data show that the straight tube model underestimates the diffusion coefficient $\bar{D}_t'$ relative to the curved model. Equivalently, for given values of n and $\bar{D}_t'$, the data show that the straight tube model underestimates the radius r relative to the curved model. This latter observation is relevant to previous studies in which the hydrated radius of DNA has been estimated from given data on n and $\bar{D}_t'$. In particular, it suggests that previous estimates of the hydrated radius based on the straight tube model are likely to be underestimates.

Figure 9 shows the rotational diffusion coefficient of each sequence in the sample set Σ_n for each value of n computed using the curved tube model. Results are shown for two values of the radius: $r=12$ Å and $r=18$ Å. For comparison, the seven curves and the experimental data from Fig. 6 are included as before for the length range considered here. Just as for the translational diffusion coefficient, it can be seen that for $r=12$ Å and for $r=18$ Å the rotational diffusion coefficients obtained from the straight tube model are uniformly below those obtained from the sequence-dependent curved tube model. Similar results are expected for each of the intermediate radius values. For given values of n and r, the data show that the straight tube model underestimates the diffusion coefficient $\bar{D}_r'^{\perp}$ relative to the curved model. Equivalently, for given values of n and $\bar{D}_r'^{\perp}$, the data show that the straight tube model underestimates the radius r relative to the curved model. The difference between the straight and the curved models is significantly more pronounced for the rotational diffusion coefficient than for the translational diffusion coefficient.

V. CONCLUDING REMARKS

The data in Figs. 8 and 9 may have an elegant mathematical explanation. Indeed, the data suggest that among all curved tubes of fixed length and radius the translational and rotational diffusion coefficients are minimal for a straight tube. Moreover, that for straight tubes of fixed length, both diffusion coefficients increase as the tube radius decreases. While such results seem plausible and even intuitive, we are not aware of any general results along these lines. We remark that such results would not only be of intrinsic mathematical interest, but would also be useful in the analysis and interpretation of hydrodynamic data on tubelike structures. Here such results would lend further support to our statement that previous estimates of the hydrated radius based on a straight tube model are likely to be underestimates.

The variability of the sequence-dependent data in Figs. 8 and 9 suggests a possible method for refining rigid basepair model parameters for DNA in solution. In particular, from experimental data on known sequences of known length, the sequence-dependent model introduced here could be used to numerically fit the relative displacement and rotation parameters for all the different possible dimer steps, as well as the hydrated radius. The computational effort required in such an endeavor would be large, but would be well within the reach of modern techniques. It is an open question whether the diffusion coefficients are sufficiently sensitive and whether current experimental techniques are sufficiently accurate to make a meaningful fit possible. However, through a proper selection of sequences to maximize sensitivity, such a fit may ultimately be feasible. We remark that, at the present time, we are not aware of any experimental results on the

sequence-dependent variability of the diffusion coefficients of DNA fragments at the length scales considered here.

Our computational model for the diffusion coefficients of DNA fragments is based on a fundamental assumption of rigidity and neglects thermally induced bending fluctuations. For fragments of about 150 basepairs, we expect that such fluctuations induce a local axial curvature of the order 10^{-3} rad/Å (root mean square). For comparison, we note that the straight tube model has zero curvature by definition, whereas the curved tube model yields a sequence-averaged curvature of the order of 10^{-2} rad/Å. Previous studies[58] related to the straight tube model indicate that these fluctuations have a negligible effect on the translational diffusion coefficient at this length scale. In view of the curvature estimates above, we expect any such effects to also be negligible and even smaller for the curved model. The effects of fluctuations on the rotational diffusion coefficient have been less well studied. Based on the results in Figs. 8 and 9, we expect the effects to be larger compared to the translational diffusion coefficient, but we are not aware of any quantitative results along these lines.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the generous support of the National Science Foundation DMS-0706951.

[1] B. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology and Physics* (Wiley, New York, 1976).
[2] B. Valeur, *Molecular Fluorescence: Principles and Applications* (Wiley-VCH, Weinheim, 2002).
[3] E. Fredericq and C. Houssier, *Electric Dichroism and Electric Birefringence* (Clarendon, Oxford, 1973).
[4] H. Fujita, *Foundations of Ultracentrifugal Analysis* (Wiley, New York, 1975).
[5] H. Schachman, *Ultracentrifugation in Biochemistry* (Academic, New York, 1959).
[6] *Modern Analytical Ultracentrifugation*, edited by T. Schuster and T. Laue (Birkhauser, Boston, 1994).
[7] D. Hawcroft, *Electrophoresis* (Oxford University Press, Oxford, 1997).
[8] T. Laue, T. Ridgeway, J. Wooll, H. Shepard, T. Moody, T. Wilson, J. Chaires, and D. Stevenson, J. Pharm. Sci. **85**, 1331 (1996).
[9] D. Baker, *Capillary Electrophoresis* (Wiley, New York, 1995).
[10] S. Allison and V. Tran, Biophys. J. **68**, 2261 (1995).
[11] S. Aragon and D. Hahn, Biophys. J. **91**, 1591 (2006).
[12] D. Brune and S. Kim, Proc. Natl. Acad. Sci. U.S.A. **90**, 3835 (1993).
[13] J. Garcia de la Torre, M. Huertas, and B. Carrasco, Biophys. J. **78**, 719 (2000).
[14] S. Allison, C. Chen, and D. Stigter, Biophys. J. **81**, 2558 (2001).
[15] S. Allison and S. Mazur, Biopolymers **46**, 359 (1998).
[16] S. Mazur, C. Chen, and S. Allison, J. Phys. Chem. B **105**, 1100 (2001).
[17] M. Tirado, C. Martinez, and J. Garcia de la Torre, J. Chem. Phys. **81**, 2047 (1984).
[18] C. Cantor and P. Schimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980), Pt. II.
[19] J. Happel and H. Brenner, *Low Reynolds Number Hydrodynamics: With Special Applications to Particulate Media* (Kluwer, Boston, 1983).
[20] S. Kim and S. Karrila, *Microhydrodynamics: Principles and Selected Applications* (Butterworth-Heinemann, Boston, 1991).
[21] S. Aragon, J. Comput. Chem. **25**, 1191 (2004).
[22] G. Youngren and A. Acrivos, J. Fluid Mech. **69**, 377 (1975).
[23] J. Garcia de la Torre and V. Bloomfield, Q. Rev. Biophys. **14**, 81 (1981).
[24] R. E. Dickerson, M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C.-S. Tung, E. von Kitzing, A. H.-J. Wang, and V. B. Zhurkin, J. Mol. Biol. **205**, 787 (1989).
[25] M. El Hassan and C. Calladine, J. Mol. Biol. **251**, 648 (1995).
[26] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman, J. Mol. Biol. **313**, 229 (2001).
[27] W. Olson, A. Gorin, X. Lu, L. Hock, and V. Zhurkin, Proc. Natl. Acad. Sci. U.S.A. **95**, 11163 (1998).
[28] J. Schellman and S. Harvey, Biophys. Chem. **55**, 95 (1995).
[29] M. Tirado and J. Garcia de la Torre, J. Chem. Phys. **71**, 2581 (1979).
[30] M. Tirado and J. Garcia de la Torre, J. Chem. Phys. **73**, 1986 (1980).
[31] T. Yoshizaki and H. Yamakawa, J. Chem. Phys. **72**, 57 (1980).
[32] O. Gonzalez, "On stable, complete and singularity-free boundary integral formulations of exterior Stokes flow," SIAM J. Appl. Math. (to be published).
[33] O. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow* (Gordon and Breach, New York, 1963).
[34] H. Power and L. Wrobel, *Boundary Integral Methods in Fluid Mechanics* (Computational Mechanics, Southampton, 1995).
[35] C. Pozrikidis, *Boundary Integral and Singularity Methods for Linearized Viscous Flow* (Cambridge University Press, Cambridge, England, 1992).
[36] K. Atkinson, *The Numerical Solution of Integral Equations of the Second Kind* (Cambridge University Press, Cambridge, England, 1997).
[37] R. Kress, *Linear Integral Equations* (Springer-Verlag, Berlin, 1989).
[38] L. Ying, G. Biros, and D. Zorin, J. Comput. Phys. **219**, 247 (2006).
[39] G. Bonifacio, T. Brown, G. Conn, and A. Lane, Biophys. J. **73**, 1532 (1997).
[40] W. Eimer and R. Pecora, J. Chem. Phys. **94**, 2324 (1991).
[41] J. Elias and D. Eden, Macromolecules **14**, 410 (1981).
[42] P. Hagerman, Biopolymers **20**, 1503 (1981).
[43] R. Kovacic and K. van Holde, Biochemistry **16**, 1490 (1977).
[44] M. Mandelkern, J. Alias, D. Eden, and D. Crothers, J. Mol. Biol. **152**, 153 (1981).
[45] N. Stellwagen, S. Magnusdottir, C. Gelfi, and P. Righetti, Biopolymers **58**, 390 (2001).
[46] R. Sinden, *DNA Structure and Function* (Academic, San Diego, 1994).
[47] P. Hughes, *Spacecraft Attitude Dynamics* (Wiley, Boston, 1983).
[48] T. Sharrock, *The Mathematics of Surfaces II* (Oxford University Press, New York, 1987), pp. 395–411.
[49] S. Allison, Macromolecules **32**, 5304 (1999).
[50] S. Harvey and J. Garcia de la Torre, Macromolecules **13**, 960 (1980).
[51] G. Galdi, *Handbook of Mathematical Fluid Mechanics* (Elsevier, Amsterdam, 2002), Vol. 1, pp. 653–792.
[52] H. Brenner, J. Colloid Sci. **20**, 104 (1965).
[53] H. Brenner, J. Colloid Interface Sci. **23**, 407 (1967).
[54] A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover, New York, 1956).
[55] S. Koenig, Biopolymers **14**, 2421 (1975).
[56] R. Johnson, J. Fluid Mech. **99**, 411 (1980).
[57] J. Keller and S. Rubinow, J. Fluid Mech. **75**, 705 (1976).
[58] H. Yamakawa and M. Fujii, Macromolecules **6**, 407 (1973).