Modeling the sequence-dependent diffusion coefficients of short DNA molecules

O. Gonzalez*       J. Li†

August 21, 2008

Abstract

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-basepair, 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.

Keywords. DNA curvature, hydrodynamic modeling, basepair parameters, hydrated radius

1 Introduction

There are many experimental methods for probing the structure of macromolecules in solution. Examples range from more modern techniques such as light scattering [9], fluorescence polarization [54] and electric dichroism and birefringence [20] to more classic techniques such as centrifugation [21, 45, 47] and electrophoresis [29], including newer variants such as membrane-confined [38] and capillary [8] 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

*Department of Mathematics, The University of Texas at Austin, og@math.utexas.edu
†Graduate Program in Computational and Applied Mathematics, The University of Texas at Austin, junli@ices.utexas.edu
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 [4, 6, 13, 24] and DNA [2, 3, 40, 53].

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 [14]. 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 [27, 33]. 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 [5, 13, 58], which provide convergent approximations to the Stokes equations and the no-slip boundary condition, and bead-type methods [23, 24], 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 [15, 18, 41]. For the case of B-DNA, parameters for all possible dimer compositions have been estimated from crystallographic data [42]. 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 provide a reasonable approximation for DNA fragments with length 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-basepair [46]. At shorter lengths, we expect an atomistic-type surface which captures local fine-scale features to be more appropriate [40]. Studies on straight tube models have shown that, for lengths greater than about 20-basepair, 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 [51, 52, 57]. Thus a tube model with ends capped by hemispheres is expected to produce nearly identical results as a model with ends capped by discs 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 the new stable, accurate and singularity-free boundary integral formulation introduced in [25]. 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 [3, 5, 58], 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 [37]. By virtue of being singularity-free, the mixed formulation can be discretized using standard, high-order Gauss quadrature rules. Moreover, because parameterizations 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 as proposed in [5]. 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 to 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 as described in [13, 33, 43, 44], but these typically require the use of product integration rules [7, 36] or specialized coordinate transformations and projections [56] 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 [13, 43].

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 formulae [53, 57] and experimental data [10, 16, 19, 26, 35, 39, 50]. A classic model for the hydrated surface of an arbitrary DNA sequence of length up to about 150-basepair is that of a straight, circular tube. The length of the tube is determined by the rise per basepair, whose average value is about 3.4Å [49], and the radius of the tube is a prescribed constant, whose value is estimated between 10Å [49] and 13Å [53]. The first estimate is based on X-ray diffraction data, whereas the second is based on experimental data in solution and empirical formulae 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 formulae, but for the rotational diffusion coefficient we observe more significant differences. Our boundary element results confirm previous observations that a hydrated radius of about 10 to 13 Å is consistent with the experimental data on the translational diffusion coefficient [53], whereas a radius of 13 to 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 [9, 14].

To illustrate the effects of 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 from [42]. Our results show that, while the data for the sequence-dependent curved tube model exhibits 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-basepair, 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 Section 2 we outline a rigid basepair model of DNA, translate the crystallographic data from [42] to the parameterization used here and describe our method for constructing a tubular surface model for a given DNA sequence. In Section 3 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 Section 4 we present numerical results for both the straight and sequence-dependent curved tube models and compare them with previous results and experimental data. In Section 5 we end with some concluding remarks.

2 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 the crystallographic data from [42] to the parameterization used here and describe our method for constructing a tubular surface model for a given DNA sequence.

2.1 Basepairs, configurations

We consider right-handed, double-helical DNA in which bases T, A, C and G are attached to two, oriented, anti-parallel 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 \( X_1 X_2 \cdots X_n \), listed in the 5' to 3' direction along the strand, where \( X_a \in \{ T,A,C,G \} \). The basepairs associated with this sequence are denoted by \( (X,\overline{X})_1, (X,\overline{X})_2, \ldots, (X,\overline{X})_n \), where \( \overline{X} \) is defined as the Watson-Crick complement of \( X \) in the sense that \( A = T \) and \( C = G \). The notation \( (X,\overline{X})_a \) for a basepair indicates that base \( X \) is attached to the reference strand while \( \overline{X} \) is attached to the opposite strand as illustrated in Figure 1(a).

There are four possible basepairs \( (X,\overline{X})_a \) corresponding to the choice \( X_a \in \{ T,A,C,G \} \). We adopt a standard model of DNA [15, 18, 41] in which each basepair \( (X,\overline{X})_a \) 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 frame vectors are defined according to the Tsukuba convention [41]. 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 this C1'-C1' axis. As a result, \( d_3^a = d_1^a \times d_2^a \) is perpendicular to the plane of \( (X,\overline{X})_a \) and normally points in the direction of \( (X,\overline{X})_{a+1} \). 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 Figure 1(b).
{T, A, C, G}. In a rigid basepair model, the positions of the non-hydrogen 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 non-hydrogen atoms. Estimated values for these positions for basepairs in their ideal forms are tabulated in [41]. Thus the configuration of a DNA molecule consisting of \( n \) basepairs is completely defined by the reference points \( q^a \) and frames \( \{d_i^a\} \) (\( a = 1, \ldots, n \)).

2.2 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 Figure 1(c). The relative rotation and displacement between \( (X, x^a) \) and \( (X, x^{a+1}) \) can be described in the general form

\[
d_{i}^{a+1} = \sum_{i=1}^{3} A_{ij}^{a} d_{i}^{a}, \quad q^{a+1} = q^{a} + \sum_{i=1}^{3} \eta_{i}^{a} d_{i}^{a},
\]

(1)

where \( A^{a} \in \mathbb{R}^{3 \times 3} \) 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 \( \{d_{i}^{a}\} \) is a right-handed, orthonormal frame between \( \{d_{i}^{a}\} \) and \( \{d_{i}^{a+1}\} \). The frame \( \{d_{i}^{a}\} \) is often referred to as a middle frame and will be defined below. From (1) we deduce that the entries in \( A^{a} \) and \( \eta^{a} \) are given by

\[
A_{ij}^{a} = d_{i}^{a} \cdot d_{j}^{a+1}, \quad \eta_{i}^{a} = d_{i}^{a} \cdot (q^{a+1} - q^{a}).
\]

(2)

A rotation matrix \( A^{a} \) can be parameterized by a coordinate vector \( \theta^{a} \in \mathbb{R}^{3} \) in a variety of ways. In this work, we parameterize rotation matrices using the Cayley (also referred to as Euler-Rodrigues or Gibbs) formula [30]

\[
A^{a} = \text{cay}[\theta^{a}] := I + \frac{4}{4 + |\theta^{a}|^{2}} (|\theta^{a}| \times) + \frac{1}{2} |\theta^{a}|^{2},
\]

(3)

where \( I \) is the identity matrix and \( |\theta^{a}| \times \) denotes the skew-symmetric matrix

\[
|\theta^{a}| \times = \begin{pmatrix}
0 & -\theta_{3}^{a} & \theta_{2}^{a} \\
\theta_{3}^{a} & 0 & -\theta_{1}^{a} \\
-\theta_{2}^{a} & \theta_{1}^{a} & 0
\end{pmatrix}.
\]

(4)

The Cayley formula can be explicitly inverted as

\[
\theta^{a} = \text{cay}^{-1}[A^{a}] := \frac{2}{\text{tr}[A^{a}] + 1} \text{vec}[A^{a} - (A^{a})^{T}],
\]

(5)

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

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

$$\nu^a = \frac{1}{|\theta^a|} \sum_{i=1}^3 \theta^a_i d^a_i, \quad \phi^a = 2 \arctan \left( \frac{|\theta^a|}{2} \right). \quad (6)$$

From (6), we deduce that a simple rotation about the frame vector $d^a_1$ is obtained when $\theta^a = (\theta^a_1, 0, 0)$, where the angle of rotation is determined by (6). Similar conclusions can be drawn for simple rotations about the other frame vectors. In general, a rotation about a given unit vector $m^a = \sum_{i=1}^3 \mu^a_i d^a_i$ through a given angle $\psi^a \in [0, \pi)$ is obtained when

$$\theta^a_i = 2 \tan \left( \frac{\psi^a}{2} \right) \mu^a_i. \quad (7)$$

The middle frame $\{d^a_i\}$ used to describe relative displacements can now be defined. Let $A^a$ be the relative rotation matrix for frame $\{d^a_{i+1}\}$ with respect to $\{d^a_i\}$. Then the coordinates $\theta^a$, axis $\nu^a$ and angle $\phi^a$ associated with this rotation are as given in (5) and (6). The middle frame is here defined by a relative rotation about the same axis $\nu^a$, but through an angle of $\phi^a/2$. Using (7) we obtain

$$d^a_j = \sum_{i=1}^3 \text{cay}_{ij} [\theta^a] d^a_i, \quad \theta^a_i = 2 \tan \left( \frac{\phi^a}{4} \right) \frac{\theta^a_i}{|\theta^a|}. \quad (8)$$

Thus the relative rotation and displacement between an arbitrary pair of adjacent basepairs $(X, X)_a$ and $(X, X)_{a+1}$ is completely described by the coordinates $\theta^a$ and $\eta^a$. The definitions above satisfy all the qualitative guidelines set forth in the Cambridge convention [15], including the symmetry conditions associated with a change of reference strand. Accordingly, we refer to $\theta^a$ as Tilt-Roll-Twist coordinates and $\eta^a$ as Shift-Slide-Rise coordinates. For a given configuration $\{q^a, d^a_i\}$ of basepair $(X, X)_a$, the above definitions establish a one-to-one correspondence between $\{\theta^a, \eta^a\}$ and the configuration $\{q^a+1, d^a_{i+1}\}$ of basepair $(X, X)_{a+1}$. Notice that $\theta^a$ are not conventional angular coordinates as employed by many authors. Rather, they are abstract coordinates defined via the parameterization in (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.

2.3 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 $(X, X)_a$ and
Thus there are 16 different coordinate sets corresponding to the 16 possible basepair combinations defined by $X_a, X_{a+1} \in \{T, A, C, G\}$. Crystal structures of B-form DNA were analyzed in [42]. 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 1 and 2 as a function of $X_a$ (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 $\theta^a$ in Table 1 are dimensionless, whereas the parameters $\eta^a$ in Table 2 have dimensions of length.

In this work, we use Tables 1 and 2 to define the three-dimensional structure of a DNA molecule. Given a molecule with sequence $X_1X_2\cdots X_n$, 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 $\{q^a, d^a\}$ of basepair $(X, \overline{X})_a$, the configuration $\{q^{a+1}, d^{a+1}\}$ of basepair $(X, \overline{X})_{a+1}$ 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 1 and 2 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.

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>A</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.00</td>
<td>-2.56</td>
<td>-0.18</td>
<td>-3.09</td>
</tr>
<tr>
<td>A</td>
<td>1.98</td>
<td>1.28</td>
<td>1.27</td>
<td>8.17</td>
</tr>
<tr>
<td></td>
<td>5.23</td>
<td>6.32</td>
<td>5.64</td>
<td>5.72</td>
</tr>
<tr>
<td></td>
<td>2.56</td>
<td>0.00</td>
<td>2.76</td>
<td>-0.92</td>
</tr>
<tr>
<td>T</td>
<td>1.28</td>
<td>6.09</td>
<td>3.49</td>
<td>8.66</td>
</tr>
<tr>
<td></td>
<td>6.32</td>
<td>6.85</td>
<td>6.56</td>
<td>6.76</td>
</tr>
<tr>
<td>G</td>
<td>0.18</td>
<td>-2.76</td>
<td>0.00</td>
<td>-0.18</td>
</tr>
<tr>
<td></td>
<td>1.27</td>
<td>3.49</td>
<td>0.55</td>
<td>6.55</td>
</tr>
<tr>
<td></td>
<td>5.64</td>
<td>6.56</td>
<td>6.04</td>
<td>5.91</td>
</tr>
<tr>
<td>C</td>
<td>3.09</td>
<td>0.92</td>
<td>0.18</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>8.17</td>
<td>8.66</td>
<td>6.55</td>
<td>9.92</td>
</tr>
<tr>
<td></td>
<td>5.72</td>
<td>6.76</td>
<td>5.91</td>
<td>6.52</td>
</tr>
</tbody>
</table>
Table 2: Estimates of Shift-Slide-Rise (Å) in Cayley mid-frame.

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>A</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>-0.35</td>
<td>1.35</td>
<td>0.90</td>
</tr>
<tr>
<td>A</td>
<td>-5.95</td>
<td>-0.82</td>
<td>-5.82</td>
<td>-2.68</td>
</tr>
<tr>
<td></td>
<td>3.31</td>
<td>3.27</td>
<td>3.36</td>
<td>3.34</td>
</tr>
<tr>
<td>T</td>
<td>0.35</td>
<td>0.00</td>
<td>2.85</td>
<td>-0.94</td>
</tr>
<tr>
<td></td>
<td>-0.82</td>
<td>0.31</td>
<td>0.79</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>3.27</td>
<td>3.42</td>
<td>3.37</td>
<td>3.33</td>
</tr>
<tr>
<td>G</td>
<td>-1.35</td>
<td>-2.85</td>
<td>0.00</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>-5.82</td>
<td>0.79</td>
<td>-3.81</td>
<td>-2.37</td>
</tr>
<tr>
<td></td>
<td>3.36</td>
<td>3.42</td>
<td>3.40</td>
<td>3.42</td>
</tr>
<tr>
<td>C</td>
<td>-0.90</td>
<td>0.94</td>
<td>-0.56</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>-2.68</td>
<td>5.04</td>
<td>-2.37</td>
<td>3.80</td>
</tr>
<tr>
<td></td>
<td>3.34</td>
<td>3.33</td>
<td>3.42</td>
<td>3.39</td>
</tr>
</tbody>
</table>

2.4 Hydrated surface

We model the hydrated surface of a DNA molecule as a curved cylinder or tube of uniform radius with ends capped by hemispheres. The axis of this tube is a space curve defined by the basepair reference points. For a given radius $r$ and sequence $S = X_1 X_2 \cdots X_n$, we denote the tubular surface by $\Gamma(r, S)$ and the axial curve by $\gamma(S)$. We model $\gamma(S)$ as a bi-arc 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 base-pair 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 bi-arc 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 $q^a_0 = q^a$, constructed as described in Section 2, and an interpolating curve $\gamma_k(S)$, defined as the unique bi-arc interpolant of $q^a_k$ [48], with tangent vectors computed from a cubic spline. If $\kappa_{\gamma_k}(S) \leq \zeta/r$, the procedure terminates 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

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

The nodes $q^a_{k+1}$ 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, $r = 13\,\text{Å}$) the procedure 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 bi-arc curve $\gamma(S)$ which closely follows the path of the basepair reference points $q^a$ (see Figure 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 bi-arc 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 Figure 2. For reference, the figure also shows a straight tube model $\Gamma_{str}(r,S)$. The straight tube is constructed in a similar way from equally-spaced reference points placed along a line with spacing $3.4\,\text{Å}$, which corresponds to a commonly accepted value for the average rise per basepair [49]. This value is in close agreement with the average value of rise obtained from Table 2, which is 3.36Å.

3 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.
3.1 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 $\Gamma$, the interior body domain by $B$ and the exterior fluid domain by $B_e$. Given a velocity field $v: \Gamma \rightarrow \mathbb{R}^3$, we seek to determine the hydrodynamic loads exerted by the fluid on the body. According to standard theory [27, 33], these loads are determined by the fluid velocity field $u: B_e \rightarrow \mathbb{R}^3$ and pressure field $p: B_e \rightarrow \mathbb{R}$ which satisfy the classic Stokes equations

\begin{align}
\mu u_{i,jj} - p_{,i} &= 0, \quad x \in B_e \\
 u_{i,i} &= 0, \quad x \in B_e \\
 u_i &= v_i, \quad x \in \Gamma \\
 u_i, p &\rightarrow 0, \quad |x| \rightarrow \infty.
\end{align}

Equation (9)$_1$ is the local balance law of linear momentum for the fluid and (9)$_2$ is the local incompressibility constraint. Equation (9)$_3$ 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 [1, 5].) The limits in (9)$_4$ are boundary conditions which 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 one to three. 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.
3.2 Body velocities, hydrodynamic loads

When the body $B$ is rigid, the boundary velocity field $v$ in equation (9) takes the general form

$$v = V + \Omega \times (x - c), \quad (10)$$

where $\times$ denotes the cross product. Here $V$ is the linear velocity of a given reference point $c$, and $\Omega$ is the angular velocity of the body-fixed frame. Under mild assumptions on the surface $\Gamma$ [37], the equations in (9) have a unique solution $(u,p)$ for given $(V,\Omega)$. The fluid stress field $\sigma : B \rightarrow \mathbb{R}^{3 \times 3}$ associated with this solution is defined by

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

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

$$f = \sigma \nu, \quad (12)$$

where $\nu$ is the outward unit normal on $\Gamma$. 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

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

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 $(V,\Omega)$ vanish. Resultant loads due to hydrostatic effects, that is, buoyancy loads caused by a body force field, may be accounted for separately.

3.3 Resistance, mobility matrices

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

$$F = -L_1 V - L_3 \Omega, \quad T = -L_2 V - L_4 \Omega. \quad (14)$$

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 $\mu$. With the exception of $L_1$, they depend on the choice of reference point $c$ appearing in equations (10) and (13) [27, 28, 33]. Self-adjointness of the differential operator associated with equations (9) implies that the overall resistance matrix

$$L = \begin{pmatrix} L_1 & L_3 \\ L_2 & L_4 \end{pmatrix} \quad (15)$$

is symmetric [22, 27, 33]. Furthermore, an energy dissipation inequality associated with equations (9) implies that $L$ is also positive-definite.
Throughout our developments we denote the inverse of $L$ by $M$ so that

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

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

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

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

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 $c$ [28]. In view of (14) and (16) we have

$$V = -M_1F - M_3T, \quad \Omega = -M_2F - M_4T. \quad (18)$$

The resistance matrix $L$ and 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 $(V, \Omega)$, the resultant hydrodynamic loads $(F, T)$ predicted by the solution of (9) are unique and given by $(F, T) = -L(V, \Omega)$. Conversely, for a body subject to prescribed hydrodynamic loads $(F, T)$, the resulting velocities $(V, \Omega)$ 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 of 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 [28]

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

where $Q$ and its inverse $Q^{-1}$ are matrices given in block form by, employing the notation from Section 2.2,

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

### 3.4 Diffusion, mobility coefficients

As derived in [11, 12], the translational and rotational diffusion coefficients for a rigid body immersed in a Newtonian fluid are

$$D_t = kT \vartheta_t, \quad D_r = kT \vartheta_r, \quad (21)$$

where $k$ is the Boltzmann constant, $T$ is the absolute temperature of the fluid, and $\vartheta_t$ and $\vartheta_r$ are mobility coefficients defined by

$$\vartheta_t = \frac{1}{3} \text{tr}(M_1), \quad \vartheta_r = \frac{1}{3} \text{tr}(M_4). \quad (22)$$
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 $\vartheta_t$ and $\vartheta_r$ 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 (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 [14, 17, 34].

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 (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

$$D_r^\perp = kT \vartheta_r^\perp, \quad \vartheta_r^\perp = \frac{1}{2} \text{tr}^\perp(M_4),$$

where $\text{tr}^\perp(M_4)$ 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 (23) can be viewed as a natural generalization to the case of curved cylinders of the transverse rotational diffusion coefficient of straight cylinders employed in [13, 53].

In this work, we follow standard convention and refer all diffusion coefficients to the center of diffusion [13, 28]. 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 $\vartheta_t$ and $\vartheta_r$ achieve minimum values [13, 28]. (This is trivially true for $\vartheta_r$ 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 like 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 Section 2.2,

\[
c' = c + (\text{tr}[M_4 I - M_4])^{-1} \text{vec}(M_3 - M_3^T).
\]

This expression is equivalent to those in [13, 28], 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 (19), and the associated diffusion coefficients \( D'_{t} \) and \( D'_{r \perp} \) can be obtained from relations analogous to (21) and (23). The resulting expressions take the form

\[
D'_{r \perp} = D_{r \perp},
\]

\[
D'_{t} = D_{t} - \frac{kT}{3} \text{vec}(M_3 - M_3^T) \cdot (\text{tr}[M_4 I - M_4])^{-1} \text{vec}(M_3 - M_3^T).
\]

3.5 Numerical procedure

The basic hydrodynamic problem is to solve the steady Stokes equations (9) in the three-dimensional domain exterior to the hydrated surface \( \Gamma(r, S) \) of a given DNA molecule. To this end, we employ the boundary element method described in detail in [25]. Choosing an arbitrary 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 \( \Omega \), 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,

\[
-L = \left( \begin{array}{c} F \\ T \end{array} \right) \bigg| \begin{array}{c} V = (1, 0, 0) \\ \Omega = (0, 0, 0) \end{array} \cdots \left( \begin{array}{c} F \\ T \end{array} \right) \bigg| \begin{array}{c} V = (0, 0, 0) \\ \Omega = (0, 0, 1) \end{array} \right) .
\]

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 \( r \). The results in [25] 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 \( h \) 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.

4 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 Section 2. For convenience, all results are presented in terms of the dimensionless coefficients

\[ \overline{D}_t = \frac{\mu \ell D'_t}{kT}, \quad \overline{D}_r^\perp = \frac{\mu \ell^3 D'_r^\perp}{kT}, \]

(27)

where \( \ell \) is a characteristic length scale equal to 34Å. We compute diffusion coefficients for straight tubes and compare our results with various empirical formulae 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.

4.1 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 \((q, \lambda, \phi/\phi_T)\) and the element size \( h \) that were used in all subsequent calculations.

Plot (a) of Figure 3 illustrates a straight tube model with \( r = 10 \text{Å}, \ n = 20 \) 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: \((\lambda, \phi/\phi_T) = (1/2, 1/4)\) and \((\lambda, \phi/\phi_T) = (1/2, 1/2)\). For reference, the scale on the horizontal axis is such that \( 1/h_* = 18 \). The data illustrates 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 \((\lambda, \phi/\phi_T) = (1/2, 1/2)\) than for \((\lambda, \phi/\phi_T) = (1/2, 1/4)\). Because larger values of \( q \) require increased computational effort, we select the parameters \( q = 1 \) and \((\lambda, \phi/\phi_T) = (1/2, 1/2)\) for all subsequent calculations.

Plot (d) of Figure 3 shows the diffusion coefficients \( \overline{D}_t \) and \( \overline{D}_r^\perp \) as a function of the number of base pairs \( n \) for a straight tube with radius \( r = 10 \text{Å} \). Results are given for two element sizes: \( h_* \) and \( h_{**} \) (\( h_{**} \) is approximately half of \( h_* \)). The data shows that, for the entire range of \( n \) considered, reducing the element
Figure 3: Convergence results for straight tubes of fixed radius. (a): Tube with \( r = 10\text{Å} \) and \( n = 20 \) discretized with elements of reference size \( h_* \). (b),(c): \( \overline{D}_{t} \) and \( \overline{D}_{r}^{\perp} \) versus \( 1/h \) 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 \( (\lambda, \phi/\phi_T) = (1/2, 1/4) \). Downward-pointing triangles and squares denote results for \( q = 1 \) and \( q = 2 \) with \( (\lambda, \phi/\phi_T) = (1/2, 1/2) \). (d): \( \overline{D}_t \) (solid) and \( \overline{D}_r^{\perp} \) (dashed) versus \( n \) for tubes with \( r = 10\text{Å} \). Triangles denote results for elements of size \( h_* \). Circles denote results for elements of size \( h_\star \approx h_*/2 \). Both results obtained with \( q = 1 \) and \( (\lambda, \phi/\phi_T) = (1/2, 1/2) \). (e): \( \Delta \overline{D}_t/\overline{D}_t \) (solid) and \( \Delta \overline{D}_r^{\perp}/\overline{D}_r^{\perp} \) (dashed) versus \( n \), where \( \Delta \overline{D} \) is difference between results with element sizes \( h_* \) and \( h_\star \).
size by a factor of about two causes no significant changes in $D_t'$ and $D_r'$. Plot (e) shows the relative differences in each of $D_t'$ and $D_r'$ obtained with element sizes $h_*$ and $h_{**}$ as a function of $n$. The data shows that the relative 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 $r \geq 10\text{Å}$ 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 [25].

Figure 4 presents a comparison between our boundary element computations and the empirical formulae from [53, 57] for the diffusion coefficients of a straight, circular tube. Shown in the figure are plots of $nD_t'$ and $nD_r'$ versus $n$ for two values of the radius: $r = 10\text{Å}$ and $r = 15\text{Å}$. We choose to plot $nD_t'$ and $nD_r'$ rather than $D_t'$ and $D_r'$ since it enhances the sensitivity and allows results for different radius values to be more easily distinguished. The empirical formulae from [53] 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 formulae from [57] are based on a combination of numerical results and slender-body theory [31, 32]. In that work, interpolation formulae 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 Figure 4 shows that, for each of the two values of $r$, our boundary element computations are in general agreement with the empirical formulae for $D_t$, but that more significant differences are visible between all three approaches for $D_r^\perp$. By design, the empirical formulae from [57] for $D_t$ and $D_r^\perp$ 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 formulae from [53] for $D_t$ and $D_r^\perp$ 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: $r = 10, 11, \ldots, 15\text{Å}$. The symbols in the figure correspond to different sets of experimental data obtained by different techniques: velocity sedimentation [10, 35], dynamic light scattering [16, 39] and capillary electrophoresis [50]. Differences due to experimental conditions are expected to be minimal since all the data were normalized to standard conditions [14]. The observed disagreement between the electrophoresis data and that from other techniques is at present not well-understood [50]. 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 Figure 5 for sequences of lengths greater than 50-basepair can be seen to be consistent with two distinct radius ranges. The light scattering data from [39] is closely consistent with $r = 10 - 11\text{Å}$ and the sedimentation data from [35] is consistent with $r = 12 - 13\text{Å}$. This difference could be 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 is ignored in the interpretation of the experimental measurements. The data for sequences of about 50-basepair 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 effects 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 hemispherical end caps 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: $r = 12, 13, \ldots, 18\,\text{Å}$. The symbols in the figure correspond to different sets of experimental data obtained by different techniques: dynamic light scattering [16] and transient electric birefringence [19, 26]. As before, the data is normalized to standard conditions [14]. 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 $r = 13 - 17\,\text{Å}$. 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
Figure 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 [16], downward triangles [19], circles [26].

Difference towards larger values of \( r \) suggests that a systematic effect such as curvature is more likely than experimental error.

4.2 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 \( r = 13 \) Å. These tube models were constructed as described in Section 2 using the crystallographic data in Tables 1 and 2. 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 \( \Sigma_n = \{S_1^n, \ldots, S_m^n\} \), where \( m \) is a fixed number and each \( S_j^n \) is a sequence of length \( n \). For different values of \( n \) the sample \( \Sigma_n \) was constructed by a culling procedure. We first constructed a set \( \Sigma'_n = \{S_1^n, \ldots, S_m'^n\} \), \( m' \gg m \), consisting of an arbitrary mixture of completely random sequences, periodic sequences with random repeating units, and peri-
Figure 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.

Periodic sequences contain A-tracts of varying lengths. The radius of gyration $R$ of each $S_n \in \Sigma_n$ was computed, and the minimum and maximum values $R_{\text{min}}$ and $R_{\text{max}}$ were recorded. The set $\Sigma_n$ was then constructed by selecting appropriate elements of $\Sigma'_n$ to obtain an approximately uniform distribution in $R$ between $R_{\text{min}}$ and $R_{\text{max}}$. In our numerical experiments, we considered lengths $15 \leq n \leq 120$ (in increments of five), and for each value of $n$ we constructed sample sets of size $m = 20$. The four sequences shown in Figure 7(a) are typical elements from the sample set $\Sigma_{40}$, and the four sequences shown in Figure 7(b) are typical elements from the sample set $\Sigma_{120}$.

Figure 8 shows the translational diffusion coefficient of each sequence in the sample set $\Sigma_n$ for each value of $n$ computed using the curved tube model. Results are shown for two values of the radius: $r = 10\,\text{Å}$ and $r = 15\,\text{Å}$. For comparison, the six curves and experimental data from Figure 5 are included for the length range considered here. It can be seen that, for $r = 10\,\text{Å}$ and for $r = 15\,\text{Å}$, the diffusion coefficients obtained from the straight tube model are uniformly below those obtained from the sequence-dependent curved tube
Figure 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 $r = 10\text{Å}$, pluses $r = 15\text{Å}$. The solid curves denote results from the straight model with different radii: $r = 10, 11, \ldots, 15\text{Å}$ ordered from upper to lower curve. The open symbols denote experimental data as in Figure 5.

model. Similar results are expected for each of the intermediate radius values. Thus, for given values of $n$ and $r$, the data shows that the straight model underestimates the diffusion coefficient $\overline{D}_t$ relative to the curved model. Equivalently, for given values of $n$ and $\overline{D}_t$, the data shows that the straight 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 $\overline{D}_t$. 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 $\Sigma_n$ for each value of $n$ computed using the curved tube model. Results are shown for two values of the radius: $r = 12\text{Å}$ and $r = 18\text{Å}$. For comparison, the seven curves and experimental data from Figure 6 are included as before for the length range considered here. Just as for the translational diffusion coefficient, it can be seen that, for $r = 12\text{Å}$ and for $r = 18\text{Å}$, the rotational diffusion coefficients obtained from the straight tube model are uniformly
Figure 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 $r = 12\,\text{Å}$, pluses $r = 18\,\text{Å}$. For clarity, the crosses have been shifted horizontally one unit to the left. The solid curves denote results from the straight model with different radii: $r = 12, 13, \ldots, 18\,\text{Å}$ ordered from upper to lower curve. The open symbols denote experimental data as in Figure 6.

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 shows that the straight model underestimates the diffusion coefficient $D_r^\perp$ relative to the curved model. Equivalently, for given values of $n$ and $D_r^\perp$, the data shows that the straight model underestimates the radius $r$ relative to the curved model. The difference between the straight and curved models is significantly more pronounced for the rotational diffusion coefficient than for the translational diffusion coefficient.

5 Concluding remarks

The data in Figures 8 and 9 may have an elegant mathematical explanation. Indeed, the data suggests that, among all curved tubes of fixed length and radius, the translational and rotational diffusion coefficients are minimal for a straight
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 tube-like 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 Figures 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-basepair, 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 $10^{-2}$ rad/Å. Previous studies [55] related to the straight 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 Figures 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.

References


