THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Multiscale Modeling of Double-Helical DNA and RNA: A Unification through Lie Groups

Kevin C. Wolfe,[†] Whitney A. Hastings,[‡] Samrat Dutta,[§] Andrew Long,[∥] Bruce A. Shapiro,[⊥] Thomas B. Woolf,[#] Martin Guthold,[§] and Gregory S. Chirikjian*[,†]

[†]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland, United States

[‡]National Institutes of Health, Frederick, Maryland, United States

[§]Department of Physics, Wake Forest University, Winston-Salem, North Carolina, United States

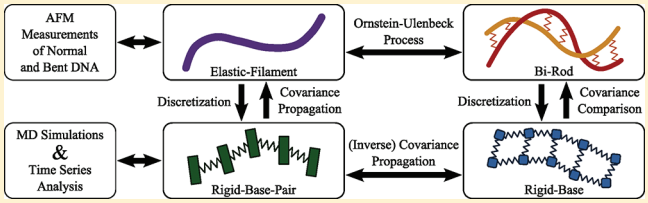
[∥]Department of Mechanical Engineering, Northwestern University, Evanston, Illinois, United States

[⊥]Center for Cancer Research Nanobiology Program, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, Maryland, United States

[#]Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

Supporting Information

ABSTRACT: Several different mechanical models of double-helical nucleic-acid structures that have been presented in the literature are reviewed here together with a new analysis method that provides a reconciliation between these disparate models. In all cases, terminology and basic results from the theory of Lie groups are used to describe rigid-body motions in a coordinate-free way, and when necessary, coordinates are introduced in a way in which simple equations result. We consider double-helical DNAs and RNAs which, in their unstressed referential state, have backbones that are either straight, slightly precurved, or bent by the action of a protein or other bound molecule. At the coarsest level, we consider worm-like chains with anisotropic bending stiffness. Then, we show how bi-rod models converge to this for sufficiently long filament lengths. At a finer level, we examine elastic networks of rigid bases and show how these relate to the coarser models. Finally, we show how results from molecular dynamics simulation at full atomic resolution (which is the finest scale considered here) and AFM experimental measurements (which is at the coarsest scale) relate to these models.



1. INTRODUCTION

DNA and RNA double helices are fairly stiff and are often treated as homogeneous elastic rods. However, double helices with different base compositions are not structurally or dynamically equal, thus leading to studies of rigid-base and rigid-base-pair models. In this paper, we focus on a unified Lie-group framework that reconciles the most typical levels of coarse-graining in DNA and RNA modeling—from molecular dynamics, to rigid-base, to rigid-base-pair, to elastic filament models. We provide a survey of common models recast in terms of Lie groups while also demonstrating new methods for reconciliation between the disparate models. Figure 1 illustrates the different levels of modeling that are compared and contrasted in this paper, and some of the ways that they are related. It also depicts several connections that are made between these models and experiments/simulations that have been performed, namely, molecular dynamics simulations and measurements using atomic force microscopy.

The structure of DNA and RNA plays an important role in the function of these molecules, and hence has been studied extensively.[1,2] Different double helix structures can allow for different compaction ratios within the cell, initiate supercoiling, affect the local concentration of the solution (e.g., by interacting with ions and water), and create a variety of nanostructures. Studying the properties of these double helices can help determine these structure–function relationships. A standard helical structure and its implementation into nanostructures is shown in Figure 2. Such structures, as well as naturally occurring DNAs and RNAs, consist of a combination of double-helical regions, stiff bends, and/or regions of increased flexibility (bulges and defects). The models and methods that we present here are general enough to address all of these scenarios.

Over the past few decades, experimental and theoretical studies of the mechanical behavior of double-stranded DNA have focused on long length scales (>100 nm). The most popular theoretical method is the elastic-filament model that treats the macromolecule as an inextensible elastic rod, and attributes to its bending deformations a classical elastic energy

Special Issue: Macromolecular Systems Understood through Multi-scale and Enhanced Sampling Techniques

Received: December 30, 2011
Revised: April 30, 2012
Published: June 7, 2012

   dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

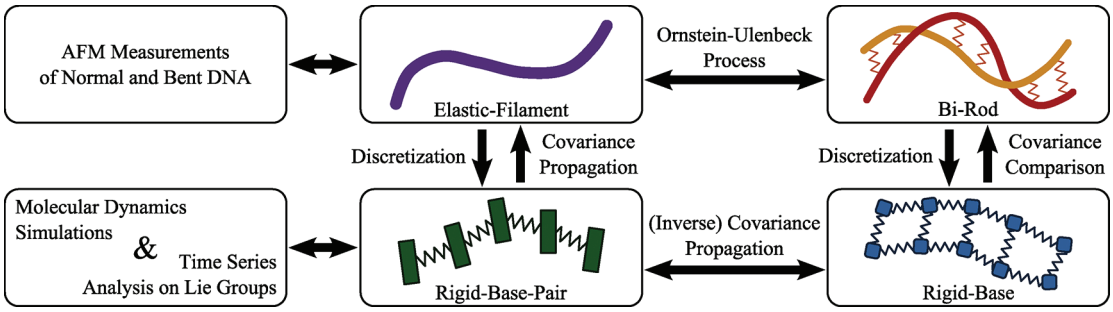


Figure 1. An overview of the models explored and some of the connections made between them.

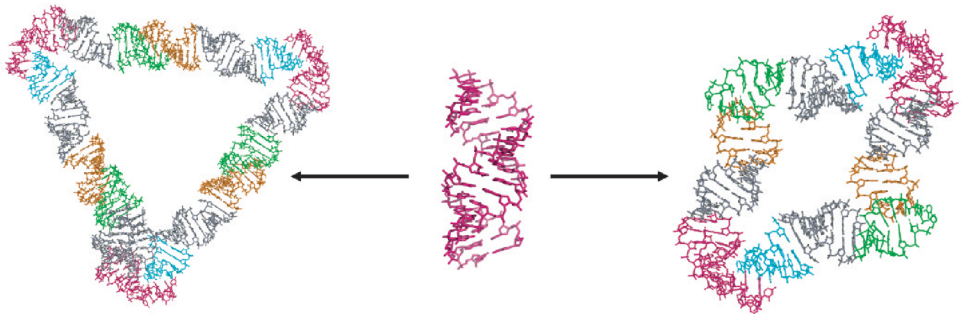


Figure 2. Two examples of how helical RNA can be used to create nanostructures.

cost (Hooke's law). Its success can be attributed to its simplicity and its successful description of experiments such as force spectroscopy on single DNA molecules.[3−5] However, for short and intermediate length scales, the elastic-filament model using DNA stretching, atomic force microscopy (AFM), and other methods has been largely insensitive to the details of mechanics crucial for cellular function, DNA packaging, transcription, gene regulation, and viral packaging.[6−10]

To further evaluate the elastic properties of DNA, DNA sequence-dependent flexibility has been examined.[11−13] Sequence-dependent local (base-pair step) force fields and chain properties have been inferred from molecular dynamics simulations and compared with results from other techniques. Accumulating evidence shows that there are preferential DNA binding sequences and significant variability depending on sequence. For example, it was shown that alternating AT repeats are 20% more flexible than control sequences.[14] Interestingly, studies on RNA elasticity models (of varying length scales) and RNA sequence-dependent flexibility are few, especially considering the greater wealth of structural features available.

In the emerging field of nanotechnology, it is becoming very important to know the properties of these biological motifs, especially when designing nanostructures.[15] For example, if the desired nanostructure is a very rigid cube, a natural building block for the sides of the cube would be the double helix. However, it would be useful to know the stiffness parameters for helices with different base compositions so that the stiffest helix could be selected for structure. This is one of many scenarios in which coarse-grained mechanical models can be used. Bridging different length scales is facilitated by methodologies such as those presented here.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the Lie-group methods that are common to all four models of DNA detailed in this paper. Section 3 reviews the generalized elastic filament or worm-like chain model in which local stiffness can be completely anisotropic, and the baseline backbone shape can be an arbitrary curve. The worm-like chain model is then applied to DNA with a drug induced bend as a special case. Section 4 describes the bi-rod model in which a double-helical nucleic acid structure is viewed as two intertwined elastic rods, and reconciles this model with the single elastic filament case using properties of the classical Ornstein−Uhlenbeck stochastic process. The rigid-base-pair model discussed in section 5 uses a rigid body to represent each base pair in double-stranded DNA or RNA. Predictions of this model with fine-grained molecular dynamics simulations are shown to be consistent. The final model detailed in section 6 represents each of the bases of DNA and RNA as rigid bodies connected by a network of elastic constraints to create a double-helical structure. Section 7 provides conclusions and discusses avenues for future work.

2. A LIE-GROUP TREATMENT OF RIGID-BODY MOTIONS

The set of rigid-body motions is denoted in this paper as $G = SE(3)$ (the special or proper Euclidean motion group in three space). Any $g \in G$ can be faithfully represented with a 4×4 *homogeneous transformation matrix* of the form

$$g = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} = \begin{pmatrix} \mathbb{I}_3 & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} R & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} = \mathrm{trans}(\mathbf{t})\mathrm{rot}(R) \tag{1}$$

where $R \in SO(3)$ (the special orthogonal group) is a 3×3 rotation matrix, $\mathbb{I}_n$ is the $n \times n$ identity matrix, $\mathbf{t}$ is a translation vector, and $\mathbf{0}^T = [0, 0, 0]$ is a row of zeros. The first matrix on the right-hand side of eq 1 is a pure translation, and the second is a pure rotation. The set of all such transformations together with the operation of matrix multiplication (which sometimes is denoted as $\circ$) forms a group, $(G, \circ)$. Often we will suppress the

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556−8572

∘ and denote the group product as the juxtaposition of group elements unless it is useful to emphasize it.

Every such rigid-body motion can be parametrized using the matrix exponential as

$$g = \exp(X) \quad \text{where} \quad X = \sum_{i=1}^{6} x_i E_i \quad \text{and}$$
$$\mathbf{x} = [x_1, ..., x_6]^T \tag{2}$$

The matrix exponential can be defined using the Taylor series such that

$$\exp(X) = \mathbb{I} + X + \frac{1}{2}X^2 + \frac{1}{3!}X^3 + ...$$

The matrix logarithm can be used as the inverse of the matrix exponential so that

$$\log(\exp(X)) = X$$

It should be noted that $\log(g)$ is not well-defined everywhere, but it can be used for nearly every $g \in SE(3)$. Here

$$E_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$
$$E_3 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$
$$E_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad E_6 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

are the infinitesimal generators of the group.

The linear vector space spanned by $\{E_i\}$, together with the matrix commutator operation $[X, Y] = XY - YX$, defines the Lie algebra $\mathcal{G} = se(3)$ associated with the Lie group $G = SE(3)$.

The explicit closed-form expression for the matrix exponential in eq 2 is known,[16] but the most relevant fact for our presentation is that when $\|\mathbf{x}\| \ll 1$, $\exp(X) \approx \mathbb{I}_4 + X$.

Let $g(s)$ be a reference frame that evolves with curve parameter s and $\dot{g}(s) \doteq dg/ds$ (throughout our formulation, the curve parameter, s, takes the place that "time" would normally have in formulations of rigid-body mechanics). The composite translational and angular "velocity" corresponding to $g(s) = \exp X(s)$ and described in the moving reference frame is extracted from

$$g^{-1}\dot{g} = \begin{pmatrix} R^T\dot{R} & R^T\dot{\mathbf{t}} \\ \mathbf{0}^T & 0 \end{pmatrix} = \sum_{i=1}^{6} \xi_i(s) E_i$$

using the "$\vee$" operator, which is the unique linear operator such that

$$E_i^{\vee} = \mathbf{e}_i$$

the ith natural unit basis vector for $\mathbb{R}^6$. Then,

$$(g^{-1}\dot{g})^{\vee} = \xi(s)$$

The opposite operation of $\vee$ is

$$\hat{\xi}(s) = g^{-1}\dot{g} = \sum_{i=1}^{6} \xi_i E_i$$

It is often convenient to decompose ξ into its angular ($\boldsymbol{\omega}$) and translational ($\mathbf{v}$) parts as

$$\xi = \begin{pmatrix} \boldsymbol{\omega} \\ \mathbf{v} \end{pmatrix}$$

Two important adjoint matrices are $\mathrm{Ad}(g)$ and $\mathrm{ad}(\hat{\xi})$ which for $SE(3)$ and $se(3)$ are defined as follows

$$\mathrm{Ad}(g) = \begin{pmatrix} R & O \\ TR & R \end{pmatrix} \quad \text{and} \quad \mathrm{ad}(\hat{\xi}) = \begin{pmatrix} \Omega & O \\ V & \Omega \end{pmatrix} \tag{3}$$

where $\Omega = -\Omega^T$ is the skew-symmetric matrix such that $\Omega\mathbf{y} = \boldsymbol{\omega} \times \mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^3$, and likewise, T and V are the skew-symmetric matrices such that $T\mathbf{y} = \mathbf{t} \times \mathbf{y}$ and $V\mathbf{y} = \mathbf{v} \times \mathbf{y}$.

The argument of $\mathrm{ad}(\cdot)$ need not be a velocity. It can be any linear combination of the matrices $\{E_i\}$, including X. The matrices $\mathrm{Ad}(g)$ and $\mathrm{ad}(X)$ are related by the equality

$$\mathrm{Ad}(\exp(X)) = \exp(\mathrm{ad}(X))$$

Their structure follows from the conditions

$$(gE_ig^{-1})^{\vee} = \mathrm{Ad}(g)\mathbf{e}_i \quad \text{and} \quad \widehat{[X,Y]} = \mathrm{ad}(X)\mathbf{y}$$

3. ELASTIC-FILAMENT (CHIRAL, ANISOTROPIC, AND SEQUENCE-DEPENDENT WORM-LIKE CHAIN) MODEL

The Fokker–Planck equation for the family of probability density functions (pdfs) $\{f(g; s) | s \in [0, L]\}$ describing the elastic filament model with referential (unperturbed) backbone shape $g_0(s)$ defined by $(g_0^{-1}(dg_0/ds))^{\vee} = \xi(s)$ and arc length/sequence-dependent diffusion matrix $D(s)$ is[17,18]

$$\frac{\partial f}{\partial s} = \frac{1}{2}\sum_{i,j=1}^{6} D_{ij}(s)\tilde{E}_i\tilde{E}_j f - \sum_{k=1}^{6} (\xi_0(s)\cdot\mathbf{e}_k)\tilde{E}_k f \tag{4}$$

where

$$\tilde{E}_i f(g) = \left(\frac{\mathrm{d}}{\mathrm{d}t} f(g \circ \exp(tE_i))\right)\Big|_{t=0} \tag{5}$$

is a Lie derivative (which can be thought of as a directional derivative in the direction defined by E_i). The collection of these derivatives can be written in a column vector as a gradient

$$\tilde{\mathbf{E}}f = [\tilde{E}_1 f, \tilde{E}_2 f, ..., \tilde{E}_6 f]^T$$

Equation 4 has diffusion and drift parameters, $D(s)$ and $\xi_0(s)$, which can be sequence dependent, consistent with observations in the literature.[19] Methods for generating marginal densities associated with $f(g; s)$ that differ from ours are known.[20] Throughout the text, $f(\cdot)$ will be used to describe a variety of probability density functions, the meaning of which should be clear based on the arguments of the function and the context in which it is used.

The translational part, $\mathbf{t}_0(s)$, of $g_0(s) = (R_0(s), \mathbf{t}_0(s))$ might be helical or straight, depending on the type of DNA or RNA. Here s is the arc length of this curve, so that

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

$$\left\| \frac{\mathrm{d}}{\mathrm{d}s}\,\mathbf{t}_0(s) \right\| = 1$$

However, for deformed and stretched versions of this curve, $g(s)$, s generally will not correspond to arc length, though the arc length of a deformed filament will always be a monotonically increasing function of s. Figure 3 shows an arbitrary curve filament and the relationship between $g(s)$, the curve, and s.

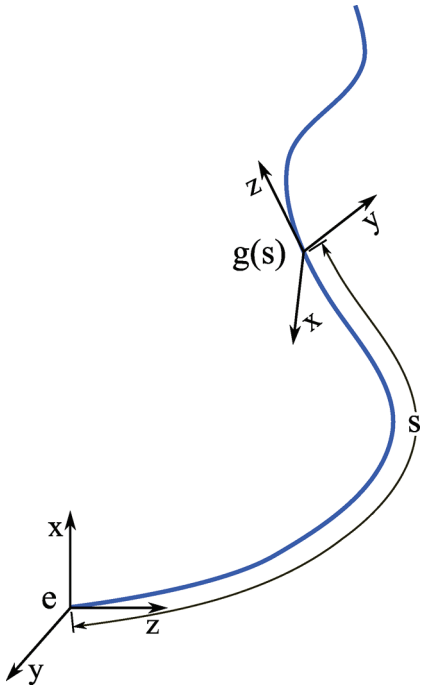


Figure 3. A parametrized set of reference frames defining an elastic filament.

Subject to Dirac delta initial conditions, $f(g;\,0) = \delta(g)$ (throughout the text, δ will be used as both the Dirac delta and the Kronecker delta; the Kronecker delta will be indicated using subscript arguments such as δ_{ab}), the solution to the Fokker–Planck equation in eq 4 gives the distribution of relative positions and orientations at a point s along the sequence relative to the proximal end. The solution is a probability density function for each fixed value of $s \in [0, L]$,

$$f(g;\,s) \geq 0 \quad \text{and} \quad \int_G f(g;\,s)\,\mathrm{d}g = 1 \tag{6}$$

Here, the integral is over the group G with respect to the Haar measure $\mathrm{d}g$. The Haar measure is used to provide a notion of consistent volume for the group allowing for a valid definition of an integral.[21]

The diffusion equation in eq 4 has been derived using several different analogies. First, the potential energy for a worm-like chain with flexibility in all six infinitesimal rigid-body motions is

$$\mathcal{V} = \frac{1}{2}\int_0^L [\xi - \xi_0]^T K[\xi - \xi_0]\,\mathrm{d}s$$

where K is the 6×6 stiffness matrix, L is the total length, and $0 \leq s \leq L$. In several works,[17,22−24] a path integral approach was used viewing this total potential energy as the "kinetic energy" of a rigid body

$$\mathcal{T} = \frac{1}{2}[\xi - \xi_0]^T K[\xi - \xi_0] \tag{7}$$

(with K interpreted as a 6×6 inertial matrix) integrated along a trajectory. The relationship between D and K is

$$D(s) = (2k_\mathrm{B}T)K^{-1}(s) \tag{8}$$

where k_B is the Boltzmann constant, T is temperature in degrees Kelvin, and K is measured in units of $k_\mathrm{B}T$. Therefore, the way to interpret eqs 4 and 8 is that diffusion is akin to flexibility, which is amplified by temperature.

Keeping with the analogy between s and time, the stochastic differential equation that is equivalent to eq 4 is

$$(g^{-1}\dot{g})^\vee\,\mathrm{d}s = \xi_0(s)\,\mathrm{d}s + B\,\mathrm{d}\mathbf{w} \tag{9}$$

where $\mathrm{d}\mathbf{w} \in \mathbb{R}^6$ is a vector of uncorrelated unit-strength white noises and B is any matrix such that $BB^T = D$. For $B = (2k_\mathrm{B}T)^{1/2}K^{-1/2}$, eq 9 can be written as

$$K(\xi - \xi_0)\,\mathrm{d}s = \sqrt{2k_\mathrm{B}T}\,K^{1/2}\,\mathrm{d}\mathbf{w} \tag{10}$$

Equation 10 can be interpreted in such a way that K is neither the stiffness nor the inertia matrix. Rather, it is a damping matrix as if there were a six-dimensional dashpot connecting an infinitesimal part of the filament at curve parameter s, which is also acted on by thermal agitation with strength $(2k_\mathrm{B}T)^{1/2}K^{1/2}$ consistent with the fluctuation–dissipation theorem.[25,26] When taking this point of view, $\mathcal{T}$ in eq 7 would not be the kinetic energy but rather this same quantity should be viewed as a Rayleigh dissipation function, and denoted as $\mathcal{R}$.

Two methods for solving eq 4 have been presented in the literature corresponding to two different scenarios. The first is for relatively short segments (i.e., those less than one persistence length). In this case, when using exponential coordinates, $g = \exp X$, the approximation

$$\tilde{E}_i f(g_0(s)\circ \mathrm{e}^X) \approx \left.\frac{\partial f}{\partial x_i}\right|_{g_0(s)}$$

is valid because the probability density is concentrated around the baseline value. In this case, eq 4 reduces to a classical diffusion equation with a Gaussian solution of the form

$$f(g;\,s) = \frac{1}{\alpha(\Sigma_{g_0}(s))}\exp\left\{-\frac{1}{2}[(\log(g_0^{-1}(s)\circ g))^\vee]^T \cdot [\Sigma_{g_0}(s)]^{-1}(\log(g_0^{-1}(s)\circ g))^\vee\right\} \tag{11}$$

where

$$\Sigma_{g_0}(s) = \int_0^s \mathrm{Ad}^{-1}(g_0(\tau))D(\tau)(\mathrm{Ad}^{-1}(g_0(\tau)))^T\,\mathrm{d}\tau \tag{12}$$

and $\alpha(\Sigma_{g_0}(s))$ is a normalizing factor to make the function a pdf for each fixed value of s. For sufficiently concentrated distributions,

$$\alpha(\Sigma) \approx (2\pi)^3|\Sigma|^{1/2} \tag{13}$$

Alternatively, given $f(g;\,s)$, the mean, g_0, and covariance, Σ_{g_0}, for any fixed value of s can be extracted so as to satisfy the conditions

$$\int_G (\log(g_0^{-1}(s)\circ g))^\vee f(g;\,s)\,\mathrm{d}g = \mathbf{0} \tag{14}$$

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

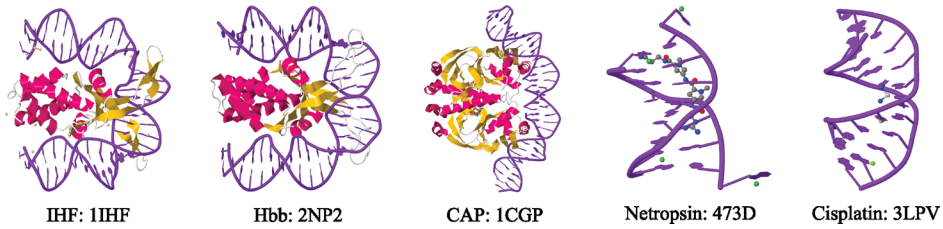


Figure 4. Five examples of DNA bending. Images generated from PDB data.[48−54]

and

$$\Sigma_{g_0}(s) = \int_G (\log(g_0^{-1}(s)\circ g))^\vee [(\log(g_0^{-1}(s)\circ g))^\vee]^T f(g;\, s)\, dg \tag{15}$$

For relatively long segments (from about half of a persistence length and longer), an alternative formulation based on the group Fourier transform is applicable. The only limitation on the filament lengths for which this model is applicable is that it is a phantom model, and does not account for excluded-volume effects. Therefore, in unconfined environments, this method (and the original governing equation, eq 4) is valid perhaps up to two or three persistence lengths. The group Fourier transform and corresponding inversion formula are written abstractly as

$$\hat{f}(\lambda) = \int_G f(g) U(g^{-1}, \lambda)\, dg \quad \text{and}$$
$$f(g) = \int_{\hat{G}} \mathrm{tr}[\hat{f}(\lambda) U(g, \lambda)]\, d(\lambda) \tag{16}$$

where λ is a "frequency" parameter drawn from the unitary dual of G, denoted as $\hat{G}$, and $\{U(g, \lambda)|\lambda \in \hat{G}\}$ is a complete set of irreducible unitary representation matrices (the "hat" operator ($\hat{\cdot}$) was previously defined for elements of the Lie algebra; in this context, it will be used to represent the Fourier transform of a function).[16,21,27]

The group Fourier transform has operational properties that are useful in solving eq 4. Namely, Fourier-space operator matrices, $\{W_i\}$, exist such that

$$\widehat{(\tilde{E}_i f)}(\lambda) = W_i(\lambda)\hat{f}(\lambda, s) \tag{17}$$

This allows us to write

$$\frac{\partial \hat{f}(\lambda, s)}{\partial s} = \mathcal{B}(s, \lambda)\hat{f}(\lambda, s) \quad \text{where}$$
$$\mathcal{B}(s, \lambda) = \frac{1}{2}\sum_{i,j=1}^{6} D_{ij}(s) W_i(\lambda) W_j(\lambda) - \sum_{k=1}^{6} (\xi_0(s)\cdot \mathbf{e}_k) W_k(\lambda) \tag{18}$$

In the case when $\xi_0(s)$ and $D(s)$ are not sequence-dependent (e.g., pristine B-form DNA of homogeneous composition), $\mathcal{B}(s, \lambda)$ becomes independent of s. In this case, $\mathcal{B}(s, \lambda) = \mathcal{B}_0(\lambda)$, and the Fourier-space solution simply becomes the matrix exponential $\hat{f}(\lambda, s) = \exp(s\mathcal{B}_0(\lambda))$. In the more general case, the linear system of ordinary differential equations can be solved by numerical integration.[16,21,27]

3.1. Bent DNA. One application of the elastic-filament model is in analyzing bent DNA. As shown in Figure 4, it is known that proteins bend DNA to activate or repress transcription;[28−34] DNA often forms loops or is wrapped around proteins,[7,35−42] and proteins scan (diffuse) along DNA to search for sites that are more flexible, or pre-bent.[30,43−47]

Moreover, many cancer drugs such as Cisplatin primarily target DNA and induce structural and mechanical changes like bending and cross-linking,[54] which can ultimately lead to apoptosis,[55,56] the desired outcome of cancer therapy. Cisplatin has a success rate of more than 90% against testicular and ovarian cancer but a low success rate for other cancers, such as lung cancer.[57] The underlying reason for these disparate success rates may be that DNA in different cancer cells responds differently to drug-induced DNA damage, such as DNA bends.[58−61] Cisplatin targets the major groove of DNA[62−64] and forms intrastrand biadducts between purine bases,[65,66] thus bending the major groove and subsequently opening up the minor groove. The major adduct is a 1,2-intrastrand adduct between adjacent guanine bases.[67] Despite the fact that this drug has been used for several decades, the structural changes induced by cisplatin in a DNA molecule, especially the local changes due to its interaction with a single molecule of cisplatin (e.g., bend angle, bend flexibility, binding geometry) are still debated, even though these properties may play a key role in how cellular proteins process this damage.[67,68] For example, the bend angle of a GG adduct in cisplatinated DNA, as determined by bulk experimental techniques and gel electrophoresis, NMR, and X-ray, ranges from 30 to 80°.[54,62,69−72] However, recent high resolution X-ray data point to a bend angle between 35 and 40°.[54] In the discussion below, we compare the techniques derived for the elastic-filament model with experimental data.

Numerical and Experimental Results. Since stiffnesses are measured in terms of k_BT, it follows that, under the extreme condition $T \to 0$, no diffusion would take place, and $f(g;\, 0,\, s) \to \delta(g_0^{-1}(s)\circ g)$. As T increases, these probability densities become less concentrated. For the biologically relevant case ($T \approx 300$), eq 4 can be solved using the harmonic analysis approach summarized by eqs 16−18.[21,73,74] If we make the shorthand notation $f_{s_1,s_2}(g) = f(g;\, s_1,\, s_2)$, then it will always be the case for $s_1 < s < s_2$ that (the notations $f_{s_1,s_2}(g)$ and $f(s;\, s_1,\, s_2)$ are used interchangeably)

$$f_{s_1,s_2}(g) = (f_{s_1,s} * f_{s,s_2})(g) = \int_G f_{s_1,s}(h) f_{s,s_2}(h^{-1}\circ g)\, dh \tag{19}$$

This is the convolution of two position and orientation distributions. While eq 19 will always hold for semiflexible phantom chains, for the homogeneous rod, there is the additional convenient properties that

$$f(g;\, s_1,\, s_2) = f(g;\, 0,\, s_2 - s_1) \quad \text{and}$$
$$f(g;\, s_2,\, s_1) = f(g^{-1};\, s_1,\, s_2) \tag{20}$$

The first of these says that for a uniform chain the position and orientation distribution only depend on the difference of arc length along the chain. The second provides a relationship between the position and orientation distribution for a uniform

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556−8572

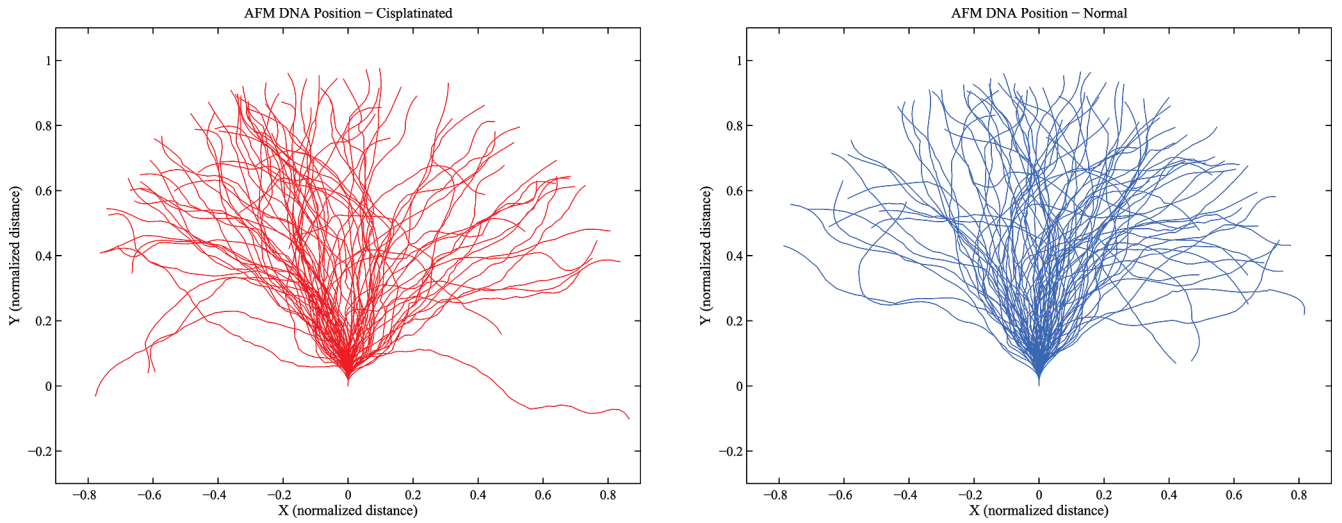


Figure 5. Two ensembles of DNA conformations as measured using an AFM: (left) cisplatinated DNA; (right) naked DNA.

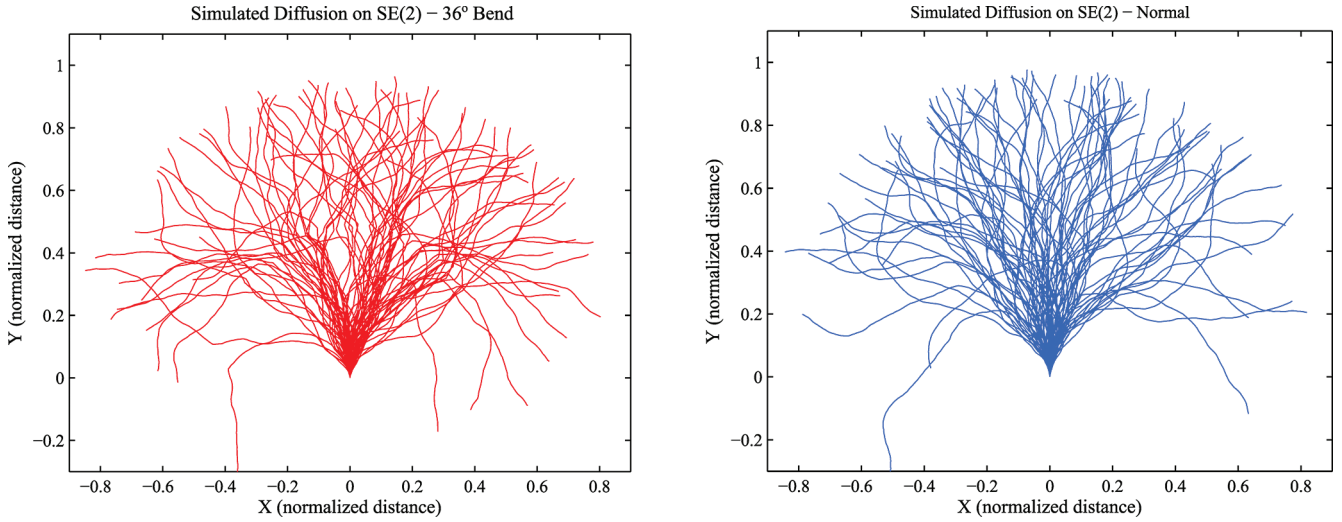


Figure 6. Two simulated ensembles of stochastic trajectories: (left) with a 36° bend; (right) without a bend.

chain resulting from taking the frame at s_1 to be fixed at the identity and recording the poses visited by s_2, and the distribution of frames that results when s_2 is fixed at the identity. When eq 20 holds, we can use the shorthand $f(g; s_1, s_2) = f(g; s_2 - s_1)$. Note that neither of these nor eq 19 will hold when excluded-volume interactions are taken into account, but our DNAs are short enough that excluded-volume effects are not important.

Since g describes fully the motion permissible in the group being considered, $f(g; s)$ is a joint density in all rigid-body degrees of freedom and any of the usual quantities of interest in polymer theory can be extracted: end-to-end distance, ring closure probabilities, end-to-end orientation distributions, etc. When a bend is present, the distribution of end-to-end position and orientation can be formulated as a convolution of the form $f(g; s_1) * b(g) * f(g; s_2)$, where $b(g)$ describes the kind of bend/twist between two segments and convolution is defined as in eq 19. Candidates for the form of $b(g)$, which are a product of delta functions in position and shifted delta functions in orientation, were given by Zhou and Chirikjian.[75] This convolution approach is particularly useful when coupled with the fact that for the group Fourier transform defined in eq 16

$$\widehat{(f_1 * f_2)}(\lambda) = \hat{f}_2(\lambda)\,\hat{f}_1(\lambda) \tag{21}$$

Figure 5 shows a series of position measurements along DNA conformations taken with an atomic force microscope (AFM) of an untreated and cisplatined 300 base pair DNA fragment. Conformations have been translocated in order that each proximal end has a common position and orientation. The DNA fragment had a single GG site in its center ($G_{150}G_{151}$) at which a GG intrastrand cisplatin cross-link was formed. The cisplatinated fragment had, thus, one central cisplatin-induced bend. The single GG sites in the DNA construct were selected because cisplatin has a high specificity of cross-linking GG sites in DNA.[76] The DNA substrate was treated with a 500-fold excess of cisplatin for 6 h and purified with a G-50 Sephadex spin column. Using inductively coupled plasma-mass spectrometry (ICP-MS), the platinum/DNA ratio was found to be 0.95, indicating nearly complete platination at the single GG site in the center of the fragment. The AFM images were obtained using deposition and imaging conditions in which the DNA molecules can equilibrate on the two-dimensional mica substrate.[77] In other words, the AFM-imaged DNA molecules behave like, and can be modeled as, worm-like chains

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

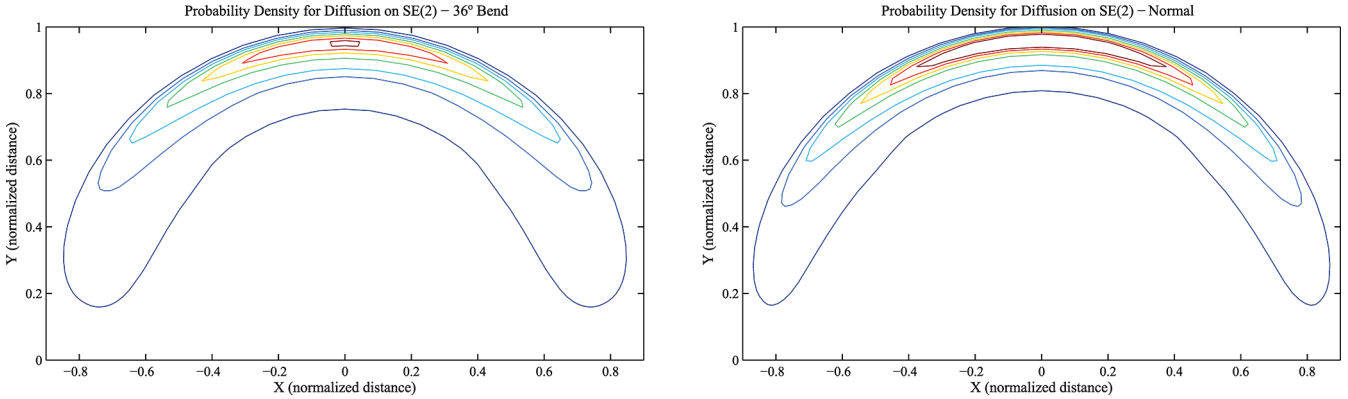


Figure 7. Distribution of the distal end position of the model relative to the proximal end: (left) with a 36° bend; (right) without a bend.

constrained to a plane. The AFM-imaged molecules can then be compared to DNA molecules that were simulated with a range of different parameters, such as different persistence lengths (or diffusion constants) and different bend angles. Figure 6 shows simulated trajectories for unperturbed or "naked" DNA and for the same DNA that has a bend induced in the middle. The comparisons of the R/L distributions of the AFM-imaged DNA molecules and the simulated DNA molecules were done using the chi-squared test, the Ansari–Bradley test,[78] and the bootstrap method using Matlab statistical subroutines. Here R represents the end-to-end distance and L, the total arc length. Comparing the normalized end-to-end distance distributions of simulated and real DNA molecules, a persistence length of 45 nm was obtained for the unmodified DNA fragment. Comparing the distributions of cisplatinated molecules with the same methods and assuming a persistence length of 45 nm for the DNA arms, a cisplatin-induced bend angle in the range 30–44° was obtained, with the best fit giving a value of 36°.[79,80] This is in very good agreement with the values obtained from gel electrophoresis and X-ray experiments[54,69,71,72,81] but disagrees with other published values.[70,82]

This model is confined to $SE(2)$, the planar motion group; however, most of what has been discussed to this point with regards to $SE(3)$ still holds for $SE(2)$, except for the explicit forms of the elements of $SE(2)$ and its associated Lie algebra, $se(2)$. These are naturally different, since elements of $SE(2)$ only have three degrees of freedom (as opposed to the six for $SE(3)$). These elements of $SE(2)$ can be represented with homogeneous matrices of the form

$$g(\theta, r, \phi) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & r\cos(\phi) \\ \sin(\theta) & \cos(\theta) & r\sin(\phi) \\ 0 & 0 & 1 \end{pmatrix} \tag{22}$$

using polar coordinates. Similar to eq 2, the infinitesimal generators of $SE(2)$ are

$$E_1 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad E_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

We can simulate a model of the worm-like chain to obtain statistics that can be compared with the AFM measurements. For example, the simulated strands in Figure 6 can be generated using a version of eq 9 of the form

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} ds = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} ds + \begin{pmatrix} \sqrt{2\eta} \\ 0 \\ 0 \end{pmatrix} dw_1 \tag{23}$$

Thus, the unperturbed path would have a constant "velocity" in the y direction. If we then look at the sampled means and covariances of the simulated and experimental data as defined in eqs 14 and 15 where the integrals are replaced with summations over the sampled values, we can determine whether or not our model accurately reflects the underlying phenomenon. The orientation of the distal end of the experimental data was approximated using the tanget of the curve at the end. For the data shown in Figures 5 and 6 where η = 0.7, the means of the four data sets were very similar. However, we can look at the sample covariances for the naked DNA

$$\Sigma_{\text{AFM}}^{\text{Naked}} = \begin{pmatrix} 0.323 & 0.032 & -0.587 \\ 0.032 & 0.057 & -0.073 \\ -0.587 & -0.073 & 1.384 \end{pmatrix} \quad \text{and}$$

$$\Sigma_{\text{Sim}}^{\text{Naked}} = \begin{pmatrix} 0.393 & 0.012 & -0.595 \\ 0.012 & 0.046 & -0.018 \\ -0.595 & -0.018 & 1.217 \end{pmatrix}$$

and the bent DNA

$$\Sigma_{\text{AFM}}^{\text{Bent}} = \begin{pmatrix} 0.480 & -0.058 & -0.829 \\ -0.058 & 0.076 & 0.132 \\ -0.829 & 0.132 & 1.781 \end{pmatrix} \quad \text{and}$$

$$\Sigma_{\text{Sim}}^{\text{Bent}} = \begin{pmatrix} 0.593 & -0.037 & -0.973 \\ -0.037 & 0.062 & 0.054 \\ -0.973 & 0.054 & 1.853 \end{pmatrix}$$

These numbers show that the simulated and measured data sets for naked DNA (or bent DNA) are more similar to each other than the two simulated (or measured) data sets are to each other. This can be expressed quantitatively by looking at the Frobenius norm of the difference between these matrices normalized as shown below. These normalized differences are calculated to be

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

$$\frac{\|\Sigma_{AFM}^{Naked} - \Sigma_{Sim}^{Naked}\|}{\|\Sigma_{AFM}^{Naked}\|} = 0.1210,$$

$$\frac{\|\Sigma_{AFM}^{Bent} - \Sigma_{Sim}^{Bent}\|}{\|\Sigma_{AFM}^{Bent}\|} = 0.1229,$$

$$\frac{\|\Sigma_{AFM}^{Naked} - \Sigma_{AFM}^{Bent}\|}{\|\Sigma_{AFM}^{Naked}\|} = 0.3828, \quad \text{and}$$

$$\frac{\|\Sigma_{Sim}^{Naked} - \Sigma_{Sim}^{Bent}\|}{\|\Sigma_{Sim}^{Naked}\|} = 0.5637$$

In addition to numerically simulating trajectories, the probability density function on the motion group can be obtained using Fourier techniques. Similar to the approach taken in previous work on $SE(3)$,[17] a probability density function on the motion group $SE(2)$ can be defined to describe the ensemble of conformations of an elastic filament. For the examples presented (contour plots shown in Figure 7), the following was used

$$\frac{\partial f}{\partial s} = (\eta(\tilde{E}_1)^2 + \tilde{E}_3)f \tag{24}$$

This is equivalent to the model given in eq 23. Using properties of the group Fourier transform from eqs 16 and 17, we can write

$$\frac{d\hat{f}}{ds} = (\eta(W_1(\lambda))^2 + W_3(\lambda))\hat{f} \tag{25}$$

where $W_i(\lambda)$'s are found using

$$W_i(\lambda) = \left(\frac{d}{dt}U(\exp(sE_i), \lambda)\right)\Big|_{s=0}$$

These matrices, which are infinite-dimensional, can be shown to be

$$[W_1]_{mk}(\lambda) = -jk\delta_{m,k} \tag{26}$$

$$[W_2]_{mk}(\lambda) = \frac{j\lambda}{2}[\delta_{k,m+1} + \delta_{n,m-1}] \tag{27}$$

$$[W_3]_{mk}(\lambda) = \frac{\lambda}{2}[\delta_{k,m-1} - \delta_{k,m+1}] \tag{28}$$

when the irreducible representation matrices, $U(g, \lambda)$, used for $SE(2)$ are

$$U_{mk}(g(\theta, r, \phi), \lambda) = j^{k-m}e^{-j[k\theta+(m-k)\phi]}J_{k-m}(\lambda r)$$

Here $J_p(\cdot)$ is the pth-order Bessel function, $j = (-1)^{1/2}$, and matrices are indexed such that $m, k \in \mathbb{Z}$. It should then be clear that

$$\hat{f}(\lambda; s) = \exp(s(\eta(W_1(\lambda))^2 + W_3(\lambda))) \tag{29}$$

and, using the inverse Fourier transform from eq 16,

$$f(g; s) = \sum_{m,k\in\mathbb{Z}} \int_0^\infty \hat{f}_{mk}(\lambda; s)U_{km}(g, \lambda)\lambda \, d\lambda \tag{30}$$

If we assume a static bend of the form[21]

$$b(\theta_0) = \frac{\delta(r)}{r}\delta(\theta - \theta_0) \tag{31}$$

where $\delta(\cdot)$ is the Dirac delta and θ_0 is the bend angle, we can use eqs 21, 29, and 30 to determine the pdf of two lengths of elastic filament connected by a static bend. Sample contour plots for such distributions are shown in Figure 7 for unit length and $\eta = 0.7$. In the bent example, the 36° bend occurs at the midpoint of the filament. Though the largest disparities may occur in the orientation, these contour plots have been marginalized over θ so that just the position remains. This was done because orientation is not as easily measured experimentally.

For the numerical integration, the infinite dimensional Fourier-space operator matrices, $\{W_i\}$, were truncated to 29 × 29 before exponentiation. Further truncation to 15 × 15 was performed prior to applying the Fourier inversion formula (eq 30) which was numerically integrated from $\lambda = 0$ to 100 using a step size of 0.1.

The contour plots in Figure 7 help to illustrate the differences inherent in the underlying distributions for bent and straight DNA. The broader shape and less severe peak in the contour plot associated with the bent case highlights the "smearing" effect that the addition of a bend has. This effect is more pronounced as η increases.

4. BI-ROD MODEL

In the bi-rod model due to Moakher and Maddocks,[83] double-helical DNA is considered to be two intertwined helical elastic rods connected with elastic contacts. On short length scales, this has different properties than the single elastic filament, whereas, on long length scales, it must converge to the single filament model, which has been verified experimentally.

Consider a bi-rod consisting of two elastic filaments described as trajectories $l(s), r(s) \in G$ for $s \in [0, L]$. Here we can think of these as a "left" and "right" filament, though they are intertwined and continually reverse order as s increases. In the referential configuration of standard B-form double-helical DNA, the backbone is given by

$$g_0(s) = \begin{pmatrix} R_3(2\pi ns/L) & s\mathbf{e}_3 \\ \mathbf{0}^T & 1 \end{pmatrix} = \text{screw}(\mathbf{e}_3, 2\pi ns/L, s) \tag{32}$$

where L/n is the helical repeat length (i.e., the helix makes n revolutions over a length L) and $R_3(\theta)$ is the rotation matrix describing counterclockwise rotation around the $\mathbf{e}_3$-axis. The referential shapes of the two elastic filaments in the bi-rod are helical and are defined by

$$r_0(s) = g_0(s)\circ\delta_0 \quad \text{and} \quad l_0(s) = g_0(s)\circ\delta_0^{-1}$$

for

$$\delta_0 = \text{trans}\left(\frac{1}{2}w\mathbf{e}_1\right) \tag{33}$$

where w is the distance between the centerlines of each filament in the bi-rod (here $\delta_0 \in G = SE(3)$ is a small rigid-body motion, and is unrelated to the Kronecker and Dirac delta functions used earlier).

Small deviations from this baseline shape can be described as

$$g(s) = g_0(s)\circ\exp(X_g) \quad \text{and} \quad \delta(s) = \delta_0\circ\exp(X_\delta)$$

where $\exp(X_{g,\delta}) \approx \mathbb{I}_4 + X_{g,\delta}$. The left and right filaments can then be written as

$$r(s) = g(s)\circ\delta(s) \quad \text{and} \quad l(s) = g(s)\circ\delta^{-1}(s)$$

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

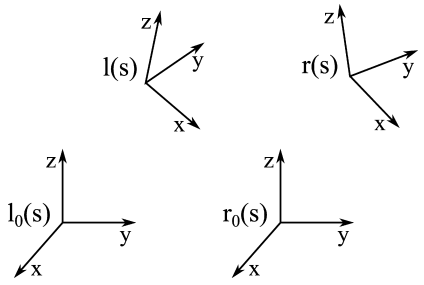


Figure 8. The relationship between reference frames in the bi-rod model.

Figure 8 illustrates the relationship between the left and right frames.

We now focus on the derivation of $f(r, l; s)$ or alternatively $f(g, \delta; s)$, with the ultimate goal of determining $D(s)$ or $K(s)$. The probability density $f(r, l; s)$ is one on the space $G \times G$, which as a group is the direct product of G with itself. As a result, the adjoint matrices in eq 3 extend to $G \times G$ as direct sums. For example, $\mathrm{Ad}(r, l) = \mathrm{Ad}(r) \oplus \mathrm{Ad}(l)$. Similar expressions hold for $\mathrm{ad}(\cdot)$, $\exp(\cdot)$, and $\log(\cdot)$.

The bi-rod version of eq 10 (again with s taking the place of time) can be derived by starting with the Rayleigh dissipation function

$$\mathcal{R} = \frac{1}{2}(\xi_r - \xi_{r_0})^T W_r (\xi_r - \xi_{r_0}) + \frac{1}{2}(\xi_l - \xi_{l_0})^T W_l (\xi_r - \xi_{l_0})$$

and the potential energy

$$\mathcal{V} = \frac{1}{2}\{\log([l^{-1}\circ r]^{-1}\circ l_0^{-1}\circ r_0)^\vee\}^T W \log([l^{-1}\circ r]^{-1}\circ l_0^{-1}\circ r_0)^\vee$$

Here W_r and W_l are symmetric weighting matrices for the right and left rods. As mentioned before in the context of worm-like chains, in some analogies, these matrices are stiffness matrices and in other analogies they are viewed as inertia matrices. At present, they are viewed as damping matrices. For the potential energy, W is a symmetric matrix that defines the spring force between the two rods. The resulting coordinate-free version of Lagrange's equations of motion with damping and external forcing, but without inertial terms, will be of the form

$$\frac{\partial \mathcal{R}}{\partial \xi_l} + \tilde{\mathbf{E}}_l \mathcal{V} = \mathbf{n}_l(s) \quad \text{and} \quad \frac{\partial \mathcal{R}}{\partial \xi_r} + \tilde{\mathbf{E}}_r \mathcal{V} = \mathbf{n}_r(s) \tag{34}$$

where the total noise vector is related to a vector of uncorrelated unit-strength Wiener processes as

$$\mathbf{n}(s)\, \mathrm{d}s = \begin{pmatrix} \mathbf{n}_l(s) \\ \mathbf{n}_r(s) \end{pmatrix} \mathrm{d}s = \sqrt{2k_B T}\, Z'\, \mathrm{d}\mathbf{w}$$

If, in analogy with eq 10 the diffusion and damping satisfy detailed balance conditions, then the coupling matrix Z' is

$$Z' = \begin{pmatrix} W_l & O \\ O & W_r \end{pmatrix}^{1/2}$$

The governing equations (eq 34) can be written explicitly for the case when $X_g^\vee$ and $X_\delta^\vee$ (and their derivatives with respect to s, $\dot{X}_g^\vee$ and $\dot{X}_\delta^\vee$) have small magnitudes. In this case, neglecting all but the linear terms in $X_g^\vee$ and $X_\delta^\vee$ and their derivatives, the following approximations can be made

$$\begin{aligned}
\log([l^{-1}r]^{-1} l_0^{-1} r_0)^\vee &\approx -\mathrm{Ad}(\delta_0^{-1}) X_\delta^\vee - X_\delta^\vee \\
\tilde{\mathbf{E}}_l \mathcal{V} &\approx -\mathrm{Ad}^T(\delta_0^{-1}\delta_0^{-1}) W (\mathbb{I} + \mathrm{Ad}(\delta_0^{-1})) X_\delta^\vee \\
\tilde{\mathbf{E}}_r \mathcal{V} &\approx W(\mathbb{I} + \mathrm{Ad}(\delta_0^{-1})) X_\delta^\vee \\
\frac{\partial \mathcal{R}}{\partial \xi_l} &\approx W_l \mathrm{Ad}(\delta_0)(\mathrm{ad}(g_0^{-1}\dot{g}_0) X_g^\vee \\
&\quad - \mathrm{ad}(g_0^{-1}\dot{g}_0) X_\delta^\vee + \dot{X}_g^\vee - \dot{X}_\delta^\vee) \\
\frac{\partial \mathcal{R}}{\partial \xi_r} &\approx W_r(\mathrm{ad}(\delta_0^{-1} g_0^{-1}\dot{g}_0 \delta_0) X_\delta^\vee \\
&\quad + \mathrm{Ad}(\delta_0^{-1}) \mathrm{ad}(g_0^{-1}\dot{g}_0) X_g^\vee \\
&\quad + \mathrm{Ad}(\delta_0^{-1}) \dot{X}_g^\vee + \dot{X}_\delta^\vee)
\end{aligned}$$

If we define

$$\mathbf{x} = \begin{pmatrix} X_g^\vee \\ X_\delta^\vee \end{pmatrix}$$

the stochastic differential equations describing the bi-rod can be written in the form

$$\mathrm{d}\mathbf{x} = -Q\mathbf{x}\, \mathrm{d}t + Z\, \mathrm{d}\mathbf{w} \tag{35}$$

where

$$Q = H^{-1} \begin{pmatrix} \mathrm{Ad}(\delta_0^{-1}) \mathrm{ad}(g_0^{-1}\dot{g}_0) & \mathrm{ad}(\delta_0^{-1} g_0^{-1}\dot{g}_0 \delta_0) \\ & \quad + W_r^{-1} W (\mathbb{I} + \mathrm{Ad}(\delta_0^{-1})) \\ \mathrm{ad}(g_0^{-1}\dot{g}_0) & -\mathrm{ad}(g_0^{-1}\dot{g}_0) - \mathrm{Ad}(\delta_0^{-1}) \cdot \\ & \quad W_l^{-1} \mathrm{Ad}^T(\delta_0^{-1}\delta_0^{-1}) \cdot \\ & \quad W(\mathbb{I} + \mathrm{Ad}(\delta_0^{-1})) \end{pmatrix}$$

$$Z = \sqrt{2k_B T} H^{-1} \begin{pmatrix} W_r^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathrm{Ad}(\delta_0^{-1}) W_l^{-1/2} \end{pmatrix}$$

and

$$H = \begin{pmatrix} \mathrm{Ad}(\delta_0^{-1}) & \mathbb{I} \\ \mathbb{I} & -\mathbb{I} \end{pmatrix}$$

These equations define a degenerate Ornstein–Uhlenbeck process. It is degenerate because the 12×12 matrix Q has rank ten with six eigenvalues with a zero real part. In other words, six degrees of freedom have elastic constraints on them from contact between the bi-rods, and six of the degrees of freedom feel only Brownian forcing.

Our goal is to obtain the corresponding probability density $f(g, \delta; s)$ and marginalize over δ. This will allow for the reconciliation of the bi-rod model with the elastic-filament model given in section 3. Using the fact that eq 35 is an Ornstein–Uhlenbeck process allows us to say that $f(g, \delta; s)$ is Gaussian. Details of the techniques used to demonstrate this can be found in the Supporting Information.

When Q has distinct eigenvalues, it can be expanded using the *spectral decomposition* as

$$Q = U \Lambda V^T = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$$

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

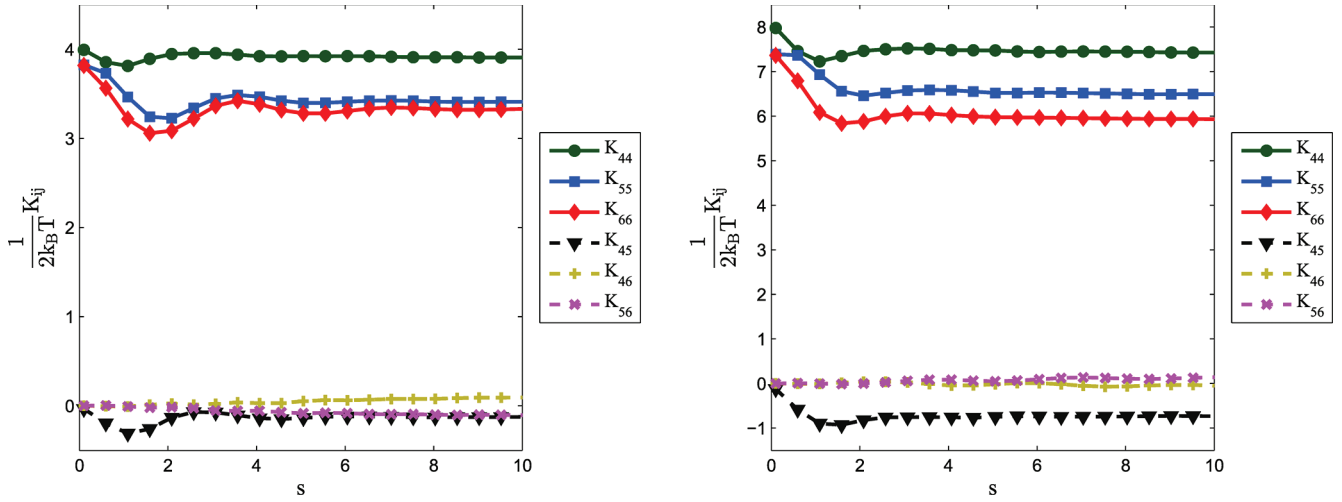


Figure 9. Examples of how the effective stiffness matrix, $K(s)$, varies for the values associated with rotation. Results are shown for two different sets of parameters: (left) $c_1 = 10$, $c_2 = 2$, $c_3 = 2$, $c_4 = 1$; (right) $c_1 = 10$, $c_2 = 4$, $c_3 = 5$, $c_3 = {}^1/_2$.

Here $U = [\mathbf{u}_1, ..., \mathbf{u}_n]$ and $V = [\mathbf{v}_1, ..., \mathbf{v}_n]$ are matrices such that the columns satisfy the following equations:

$$Q\mathbf{u}_i = \lambda_i \mathbf{u}_i \qquad Q^T\mathbf{v}_i = \lambda_i \mathbf{v}_i$$

If Q is independent of s then, as detailed in the Supporting Information, the derivative of the covariance Σ with respect to s is given by

$$\dot{\Sigma}(s) = \sum_{i,j} \exp(-(\lambda_i + \lambda_j)s)(\mathbf{v}_i^T Z Z^T \mathbf{v}_j)\mathbf{u}_i\mathbf{u}_j^T \tag{36}$$

Since we know that $f(g, \delta; s)$ is Gaussian, we can write

$$f(g, \delta; s) = \frac{1}{\alpha(\Sigma(s))} \exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma(s)^{-1}\mathbf{x}\right) \tag{37}$$

for an appropriate normalizing function $\alpha(\Sigma(s))$ and

$$\Sigma(s) = \begin{pmatrix} \Sigma_{g_0}(s) & \Sigma_{g_0,\delta_0}(s) \\ \Sigma^T_{g_0,\delta_0}(s) & \Sigma_{\delta_0}(s) \end{pmatrix}$$

Marginalizing over $X_\delta^\vee$ yields a probability density in $X_g^\vee$ that is also Gaussian with zero mean and covariance $\Sigma_{g_0}(s)$. We can relate this to the equivalent stiffness matrix defined in the context of the single elastic filament. Once $\Sigma_{g_0}(s)$ has been computed, the effective or equivalent diffusion matrix for the worm-like chain can be extracted by inverting eq 12 as

$$D(s) = \mathrm{Ad}(g_0(s))\dot{\Sigma}_{g_0}(s)\mathrm{Ad}^T(g_0(s))$$

where $\dot{\Sigma}_{g_0}(s) = (\mathrm{d}/\mathrm{d}s)\Sigma_{g_0}(s)$. The effective stiffness can then be computed using eq 8. This effective stiffness was determined as a function of s for B-form DNA with a width w of 2 nm and a pitch L/n of 3.4 nm. The damping and spring contact matrices used were

$$W_r = W_l = \begin{pmatrix} c_1\mathbb{I}_3 & O \\ O & c_2\mathbb{I}_3 \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} c_3\mathbb{I}_3 & O \\ O & c_4\mathbb{I}_3 \end{pmatrix}$$

Figures 9 and 10 demonstrate how these effective stiffnesses converge to stable values as s increases. Figure 9 illustrates how individual elements of the effective stiffness matrix converge to steady-state values. In Figure 10, the Frobenius norm of the difference between the equivalent stiffness matrix at s and the steady-state equivalent stiffness matrix is plotted as a function of s. These plots also demonstrate that, although the damping and spring contact matrices for the bi-rod model are constant with respect to s, the effective or equivalent stiffness varies as a function of s. However, since the effective stiffness converges to a steady-state value as s increases, the effective stiffness can be taken as a constant for sufficiently long rods. While the speed of convergence is a function of the parameter values, these examples provide numerical validation that the bi-rod and elastic-filament models indeed do converge as s (and thus length) increases.

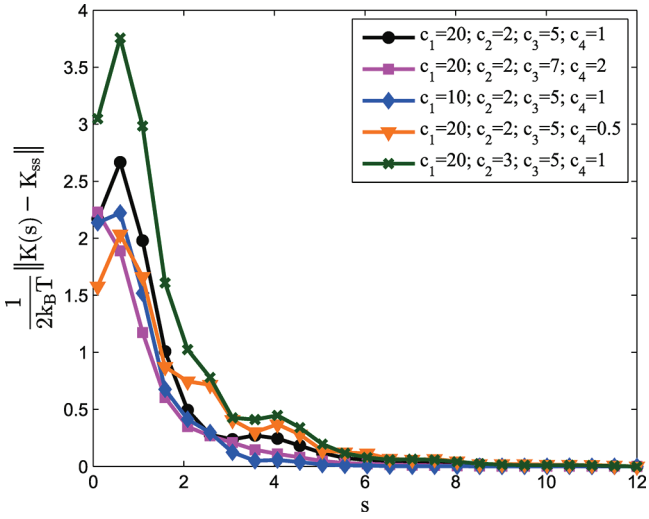


Figure 10. Examples of how the effective stiffness of the bi-rod model, $K(s)$, differs from its steady-state value, K_{ss}, for various values of $W_{l/r}$ and W. Here the Frobenius norm is used.

5. RIGID-BASE-PAIR MODELS

The rigid-base-pair model uses a rigid body to represent each base pair in a strand of DNA or RNA. These rigid bodies are attached to adjacent base pairs through elastic constraints. As such, the rigid-base-pair model can be viewed as a discretized

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

version of the single-elastic-filament model by assigning reference frames to each DNA or RNA base pair.

A number of approaches for attaching representative reference frames to base pairs have been used in the literature and in programs used to describe three-dimensional arrangements of DNA and RNA structures. A quantitative description of some of the discrepancies between the different approaches was presented by Lu et al.[84]

We will use the method in which the reference frames are positioned so that complementary bases form an ideal, planar Watson−Crick base-pair in the undistorted reference state with values of hydrogen bond donor−acceptor distances, C1′···C1′ virtual lengths, and purine N9−C1′···C1′ and pyrimidine N1−C1′···C1′ virtual angles consistent with those found in the crystal structures of small molecules.

Given this base frame assignment, we can say that, relative to the previous base pair, the kth base pair has a distribution $f_k(g)$ with mean g_k that satisfies

$$\mathbf{0} = \int_G \log^{\vee}(g_k^{-1} \circ g) f_k(g)\, \mathrm{d}g \tag{38}$$

and covariance Σ_k defined by

$$\Sigma_k = \int_G \log^{\vee}(g_k^{-1} \circ g)[\log^{\vee}(g_k^{-1} \circ g)]^T f_k(g)\, \mathrm{d}g \tag{39}$$

Then, for two base pairs with distributions $f_1(g)$ and $f_2(g)$, means g_1 and g_2, and covariances Σ_1 and Σ_2, the resulting distribution of stacking the two base pairs on top of each other can be represented as the convolution of the two distributions with

$$f_{1*2}(g) = (f_1 * f_2)(g) = \int_G f_1(h) f_2(h^{-1} \circ g)\, \mathrm{d}h \tag{40}$$

This new distribution will have mean g_{1*2} and covariance Σ_{1*2}. An exact analytical solution for g_{1*2} and Σ_{1*2} is not known in general for Lie groups; however, so-called "first-order" and "second-order" approximations do exist for these quantities in $SE(3)$. Such approximations were first introduced by Wang and Chirikjian in the context of serial robotic manipulators.[85−88] A similar concept was later applied to DNA by Becker and Everaers.[89,90]

A first-order approximation will be presented here for the convolution of two distributions. Using this recursively allows one to determine the mean and covariance of N base pairs. As will be shown in section 5.2, if N or the individual covariances are sufficiently large, the resulting distribution, $f_{1*2*\cdots*N}$, is described well with a Gaussian of the form

$$f_{1*2*\cdots*N}(g) = \frac{1}{\alpha(\Sigma_{1*2*\cdots*N})} \exp\left\{-\frac{1}{2}\mathbf{x}^T[\Sigma_{1*2*\cdots*N}]^{-1}\mathbf{x}\right\} \tag{41}$$

for

$$\mathbf{x} = (\log(g_{1*2*\cdots*N}^{-1} \circ g))^{\vee} \tag{42}$$

where $\alpha(\Sigma_{1*2*\ldots*N})$ is a normalizing factor as described in eq 13.

In the first-order theory of covariance propagation, we make the approximation $\log(g_2^{-1} \circ g_1) = X_1 - X_2$ or equivalently $[\log(g_2^{-1} \circ g_1)]^{\vee} = \mathbf{x}_1 - \mathbf{x}_2$, where $g_i = \exp(X_i)$ are elements of $SE(3)$. This converts the convolution integral over the group into one over the Lie algebra $se(3)$, which can be associated with $\mathbb{R}^6$ via the $\vee$ operator. The results are propagation formulas that produce the mean and covariance of the convolution of two functions in $SE(3)$ from the means and covariances of the two original distributions and the relative motion of the mean of the second distribution relative to the first.[86,87,89] The mean and covariance of this approximation are

$$g_{1*2} = g_1 \circ g_2 \tag{43}$$

and

$$\Sigma_{1*2} = \mathrm{Ad}(g_2^{-1})\Sigma_1 \mathrm{Ad}^{T}(g_2^{-1}) + \Sigma_2 \tag{44}$$

These formulas can be used to "piece together" serially connected distributions (each describing small motions) into a distribution that describes the overall distribution of the distal end of a semiflexible chain relative to its base. Using the fact that the adjoint is a homomorphism (i.e., $\mathrm{Ad}(g_1 \circ g_2) = \mathrm{Ad}(g_1)\mathrm{Ad}(g_2)$ and $\mathrm{Ad}(g^{-1}) = \mathrm{Ad}^{-1}(g)$), the covariance propagation formula generalizes to the concatenation of N reference frames with concentrated distributions as

$$\Sigma_{1*2*\cdots*N} = \sum_{k=1}^{N} \mathrm{Ad}^{-1}(g_{k+1} \circ \cdots \circ g_N)\Sigma_k \mathrm{Ad}^{-T}(g_{k+1} \circ \cdots \circ g_N) \tag{45}$$

To simplify the notation in the subsequent sections, the covariance propagation is simply denoted as

$$\Sigma_{1,N} = \sum_{k=1}^{N} \mathrm{Ad}_{g(k)}^{-1} \Sigma_k \mathrm{Ad}_{g(k)}^{-T} \tag{46}$$

It should be noted that the first-order covariance propagation detailed in eq 46 works relatively well for convolving distributions that are relatively concentrated (i.e., small values of $\|\Sigma_k\|$). This is typically the case for DNA and RNA, as they are relatively stiff. However, in cases where this assumption breaks down, a second-order approximation can be used. The details of this second-order propagation can be found in the literature.[88]

5.1. Inverse Propagation of Covariance. The recursive propagation scheme presented in eqs 44 and 46 assumes that values for Σ_k's are known. This may not always be the case. However, if $\Sigma_{1,N} = \Sigma_{\mathrm{data}}$ is known or can be obtained through dynamics simulations, then estimates of Σ_k can be obtained. These estimates assume that Σ_k is not dependent on k.

Since covariance matrices have a defined form, they can be represented as

$$\Sigma = \sum_{i=1}^{P} \sigma_i S_i \tag{47}$$

where S_i is the ith basis element for the set of all possible symmetric 6×6 matrices and P is the number of basis elements. Since a symmetric $n \times n$ matrix is defined by $n(n+1)/2$ independent entries, there are 21 independent entries in an $SE(3)$ covariance matrix. Since the choice of the basis $\{S_i\}$ is not unique, a natural choice for S_i is a matrix of the form $\mathbf{e}_j\mathbf{e}_j^T$ for the diagonal elements and $\mathbf{e}_j\mathbf{e}_k^T + \mathbf{e}_k\mathbf{e}_j^T$ for the off-diagonal elements, where $\mathbf{e}_j$ are the natural basis vectors for $\mathbb{R}^6$.

Expanding Σ_{data} and Σ_k from eq 46 in this way so that $\Sigma_{\mathrm{data}} = \sum_{i=1}^{P} \sigma_i^{\mathrm{data}} S_i$ and $\Sigma_k = \sum_{i=1}^{P} \sigma_i^k S_i$, we have

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556−8572

$$\sum_{i=1}^{P} \sigma_i^{\text{data}} S_i = \sum_{k=1}^{N} \text{Ad}_{g(k)}^{-1} \left(\sum_{i=1}^{P} \sigma_i^k S_i \right) \text{Ad}_{g(k)}^{-T} \quad (48)$$

$$= \sum_{i=1}^{P} \sigma_i^k \left(\sum_{k=1}^{N} \text{Ad}_{g(k)}^{-1}(S_i) \text{Ad}_{g(k)}^{-T} \right) \quad (49)$$

The symmetric matrix $\sum_{k=1}^{N} \text{Ad}_{g(k)}^{-1}(S_i)\text{Ad}_{g(k)}^{-T}$ can be expressed as

$$\sum_{k=1}^{N} \text{Ad}_{g(k)}^{-1}(S_i)\text{Ad}_{g(k)}^{-T} = \sum_{m=1}^{P} \psi_{im} S_m \quad (50)$$

and

$$\sum_{i=1}^{P} \sigma_i^{\text{data}} S_i = \sum_{i=1}^{P} \sigma_i^k \sum_{m=1}^{P} \psi_{im} S_m$$

$$= \sum_{m=1}^{P} \sigma_m^k \sum_{i=1}^{P} \psi_{mi} S_i$$

$$= \sum_{m=1}^{P} \sum_{i=1}^{P} \psi_{mi} \sigma_m^k S_i$$

Thus, it is clear that

$$\sum_{i=1}^{P} (\sigma_i^{\text{data}} - \sum_{m=1}^{P} \psi_{mi} \sigma_m^k) S_i = 0 \quad (51)$$

and

$$\sigma_i^{\text{data}} = \sum_{m=1}^{P} \psi_{mi} \sigma_m^k$$

for $i = 1, 2, ..., P$. Since S_i is a basis element, the resulting relation simplifies to the matrix expression

$$\sigma^{\text{data}} = \Psi^T \sigma^k \quad (52)$$

where

$$\sigma^{\text{data}} = [\sigma_1^{\text{data}}, \sigma_2^{\text{data}}, ..., \sigma_P^{\text{data}}]^T$$

$$\sigma^k = [\sigma_1^k, \sigma_2^k, ..., \sigma_P^k]^T$$

Then, if Ψ^T is invertible, we can solve for σ^k and then compute the helix stiffness parameters from $(1/2k_BT)K = \Sigma_k^{-1}$. This problem is similar to the estimation of model parameters for steerable needles described by Park et al.[91]

5.2. Statistics of Molecular Dynamics Simulations. Molecular dynamics simulations have been used to explore both rigid-base and rigid-base-pair models of DNA.[92,93] In this work, to determine whether the model proposed in eq 41 is valid, molecular dynamics simulations were performed on standard A-RNA helices which are stiffer than B-DNA. These models were built using Insight II (Accelrys, San Diego, CA, USA). Each helix consists of 14 base pairs with different compositions. The simulations were done for 16 ns using Amber 8.0.[94] The simulations used the ff99 Cornell force field for RNA,[95] which has proven to be a reliable force field for nucleic acids and the Amber molecular dynamics software. For the discussions below, we will say that simulations were simulated for n_t time steps and that the rigid-body transformation between the distal base pairs (g_1 and g_N) at time step i is $g^j = g_1^{-1} g_N$ for $i = 1, 2, ..., n_t$. The mean of all g^j's will be noted as g_μ.

Time Series Analysis and Stationarity of Molecular Dynamics Simulation. To analyze the stationarity of the helix during the molecular dynamics simulation, the transformation matrix g_μ is used to compute the cross-covariance matrices. For the Lie group of rigid-body motions, the cross-covariance matrices at lag l are computed as

$$\Gamma_{n_t}(l) = \frac{1}{n_t} \sum_{i=1}^{n_t} [\log(g_\mu^{-1} \circ g_i)^\vee][\log(g_\mu^{-1} \circ g_{i+l})^\vee]^T]$$

Frequently, autocorrelation plots are used to check randomness in a data set. Randomness can be ascertained by computing autocorrelations for data values at varying time lags. Random data sets have autocorrelations near zero for the time-lag separations. Conversely, if the data set is not random, then one or more of the autocorrelations will be significantly nonzero. To generate an autocorrelation plot of the cross-covariance elements, we first calculated the singular values of the cross-covariance matrix. We then plotted the singular values against

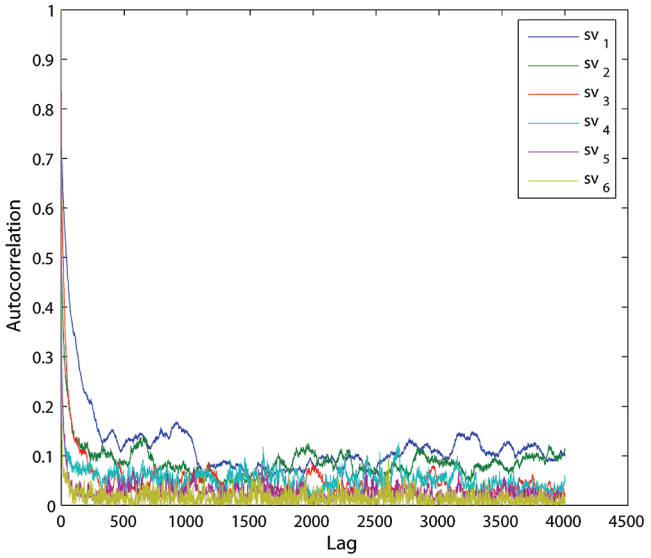


Figure 11. GCGC autocorrelation plot for the singular values of the cross-covariance matrix.

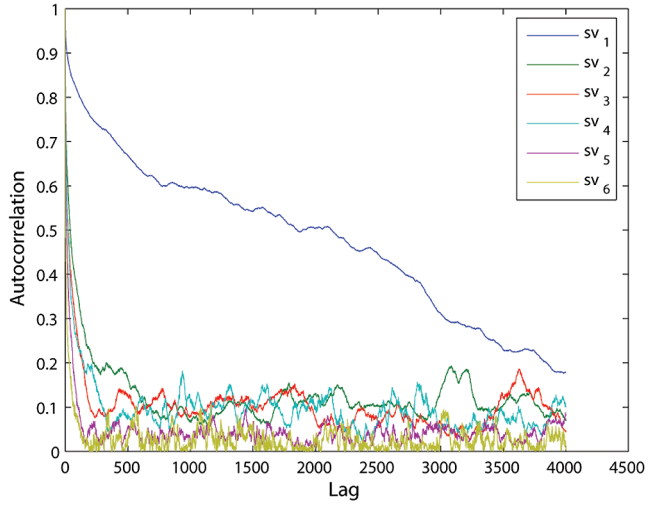


Figure 12. GCCG autocorrelation plot for the singular values of the cross-covariance matrix.

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

several different values for the lag, with the maximum lag equal to $n_t/4$. Autocorrelation plots are shown in Figures 11 and 12. For each sequence, g_i and g_μ represent the relative transformation from base-pair 3 to 12. The first and last two base-pairs were excluded from the calculations to prevent the inclusion of artifacts that might be due to end-effects during simulation.

As can be seen in Figure 11, the autocorrelation for a GCGC sequence has singular values quickly approaching zero and remaining near zero for different values of lag. This indicates that we can assume stationarity in our molecular dynamics simulations. Similar autocorrelation plots were obtained for sequences GCAU, AUAU, and AUUA. The only exception to this is the first singular value for the GCCG helix, as shown in Figure 12. This helix cannot be assumed to have stationarity and should definitely be examined further, as the equilibrium assumption may not hold.

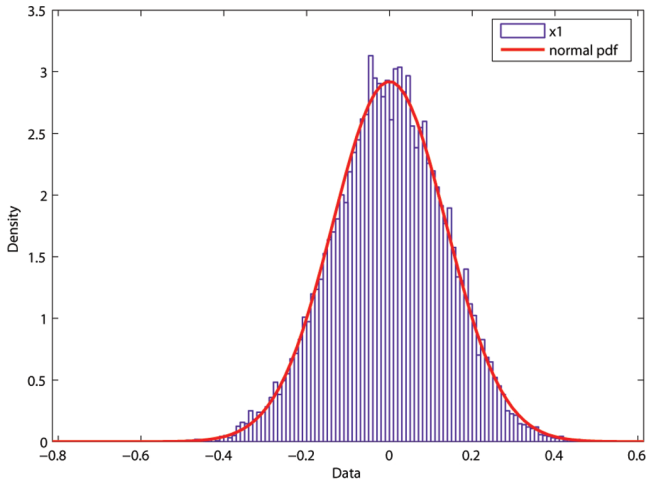


Figure 13. Probability density function with fit for the GCGC sequence for x_1.

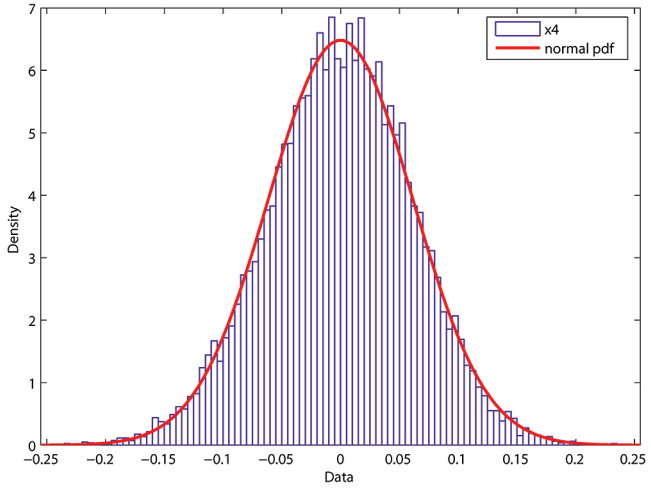


Figure 14. Probability density function with fit for the GCGC sequence for x_4.

Normality of Molecular Dynamics Simulation. To test for normality, first a few univariate tests are performed on the six components of $g_{\mu i} = \log(g_\mu^{-1} \circ g_i)^\vee$. The pdf and the corresponding normal fit curves for two of the components of $g_{\mu i}$ are plotted in Figures 13 and 14 for the GCGC helix. Similar normal probability plots are seen for the remaining four components of the GCGC helix as well as the other sequences. As in the time series calculations, $g_{\mu i}$ represents the relative transformation from base-pair 3 to 12. The first and last two base-pairs were again excluded.

To further test for normality, we use the multivariate normality test using the chi-squared plot shown in Figure 15 for

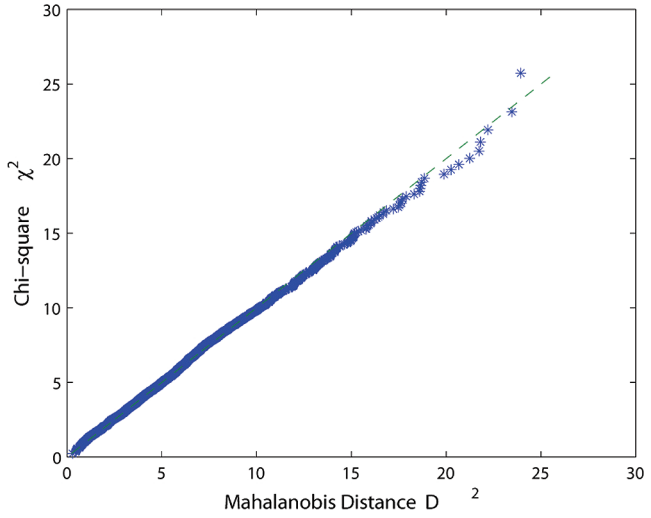
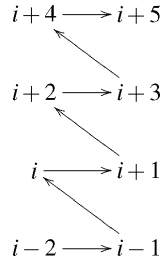


Figure 15. Chi-squared plot for the GCGC sequence parameters.

the six rigid-body parameters of $g_{\mu i}$. The straight dashed line indicates a multivariate normal fit, and our data closely resembles this fit. Furthermore, the p-value obtained for the Chi-squared test agrees with a hypothesis of multivariate normality. Comparable results for both the univariate and multivariate fitness tests are seen for the other helix sequences.

6. RIGID-BASE MODELS

Consider a double-helical DNA structure composed of N rigid bases, each of which can be connected to the others locally. Let the bases be numbered by the scheme shown below, where the horizontal direction indicates base pairs and vertical is the sequential direction.



Let $\bar{g}_i$ denote the position and orientation of the ith such body in a static minimal energy conformation, an unperturbed state. The relative transformation between body i and body j is then $\bar{g}_i^{-1} \circ \bar{g}_j$.

Let the ith rigid body move by the small amount $\exp(X_i) \approx \mathbb{I}_4 + X_i$, where $X_i = \sum_{l=1}^{6} x_{i,l} E_l$ for $X_i^\vee = \mathbf{x}_i = [x_{i,1}, x_{i,2}, ..., x_{i,6}]^T$. Then, the relative motion between bodies i and j after the small motion is

$$[\bar{g}_i(\mathbb{I}_4 + X_i)]^{-1}[\bar{g}_j(\mathbb{I}_4 + X_j)] = (\mathbb{I}_4 - X_i)(\bar{g}_i^{-1}\bar{g}_j)(\mathbb{I}_4 + X_j)$$

Retaining terms to first order, the result can be written as

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556–8572

$$(\bar{g}_i^{-1}\bar{g}_j)[\mathbb{I}_4 + X_j - (\bar{g}_i^{-1}\bar{g}_j^{-1})X_i(\bar{g}_i^{-1}\bar{g}_j)]$$

and the change in relative pose between body i and j is

$$\Delta g_{ij} = \mathbb{I}_4 + X_j - (\bar{g}_i^{-1}\bar{g}_j)^{-1}X_i(\bar{g}_i^{-1}\bar{g}_j)$$

The corresponding 6D vector of small motions is $\mathbf{x}_{ij} = (\Delta g_{ij} - \mathbb{I}_4)^\vee$ which can be written as

$$\mathbf{x}_{ij} = \mathbf{x}_j - \mathrm{Ad}(\bar{g}_j^{-1}\bar{g}_i)\mathbf{x}_i$$

Given a 6×6 stiffness K_{ij} connecting these two bodies, the corresponding potential energy is

$$\begin{aligned}\mathcal{V}_{ij} &= \frac{1}{2}\mathbf{x}_{ij}^T K_{ij}\mathbf{x}_{ij} \\ &= \frac{1}{2}[\mathbf{x}_i^T, \mathbf{x}_j^T]\begin{pmatrix} \mathrm{Ad}(\bar{g}_j^{-1}\bar{g}_i)^T K_{ij}\mathrm{Ad}(\bar{g}_j^{-1}\bar{g}_i) & -\mathrm{Ad}(\bar{g}_j^{-1}\bar{g}_i)^T K_{ij} \\ -K_{ij}\mathrm{Ad}(\bar{g}_j^{-1}\bar{g}_i) & K_{ij}\end{pmatrix}\begin{bmatrix}\mathbf{x}_i \\ \mathbf{x}_j\end{bmatrix}\end{aligned} \tag{53}$$

and the total potential energy will be

$$\mathcal{V} = \sum_{i=0}^{N-2}\sum_{j=i+1}^{N-1}\mathcal{V}_{ij} = \frac{1}{2}\mathbf{x}^T\tilde{K}\mathbf{x} \tag{54}$$

where $\mathbf{x} \in \mathbb{R}^{6N-6}$ is a composite vector of all small rigid-body motions of the structure (relative to the base 0) and $\tilde{K}$ is a composite $(6N-6) \times (6N-6)$ stiffness matrix.

Equation 54 shows that the potential energy of the system run at equilibrium is in quadratic form which leads to the probability density function that will be used, the Boltzmann distribution. This corresponding conformational Boltzmann distribution is

$$f(\mathbf{x}) = \frac{1}{\alpha(2k_BT\tilde{K}^{-1})}\exp\left(-\frac{1}{2(k_BT)}\mathbf{x}^T\tilde{K}\mathbf{x}\right) \tag{55}$$

Here $\alpha(2k_BT\tilde{K}^{-1})$ is the conformational partition function that normalizes $f(\mathbf{x})$ as a probability density. This is a Gaussian distribution, the covariances of which can be related to the stiffness matrix as

$$\Sigma = \int_{\mathbf{x}\in\mathbb{R}^{6N-6}}\mathbf{x}\mathbf{x}^T f(\mathbf{x})\,dx = 2k_BT\tilde{K}^{-1}$$

This model can be viewed as a discrete version of the bi-rod. If Δs is the step-length between stacked bases in the continuum model and W_l, W_r, and W in sectoin 4 are viewed as being functions of the curve parameter, s, then

$$W_l(p\Delta s) = K_{2p,2p+2}, \quad W_r(p\Delta s) = K_{2p+1,2p+3}, \quad \text{and} \quad W(p\Delta s) = K_{2p,2p+1} \tag{56}$$

for $p \in \{0, 1, 2, ...\}$.

The reference frame in the middle of g_0 and g_1 defines the proximal end of the chain, and the frame in the middle of g_{N-2} and g_{N-1} is the distal end. Let g_d represent the frame at the "midpoint" of the distal end of the chain. The reconciliation between this model and other models results from marginalizing over all degrees of freedom except those that are present in

$$g_d = (g_0^{-1}\circ g_1)^{-1/2}\circ g_0^{-1}\circ g_{N-2}\circ(g_{N-2}^{-1}\circ g_{N-1})^{1/2} \tag{57}$$

If we let $g_d = \bar{g}_d(\mathbb{I} + X_d)$, then we can look at the covariance matrix Σ_d associated with $\mathbf{x}_d$ by determining the relationship between $\mathbf{x}_d$ and $\mathbf{x}_1$, $\mathbf{x}_{N-2}$, and $\mathbf{x}_{N-1}$. We can start by letting $g_0\circ(g_0^{-1}\circ g_1)^{1/2} = e$ and use the referential configuration given by eqs 32 and 33. As shown in the Supporting Information, section S2, this leads to

$$\begin{aligned}g_d = \bar{g}_{N-2}\Bigg[&\mathbb{I}_4 + X_{N-2} + \frac{1}{2}\log(\mathrm{trans}(w\mathbf{e}_1)) \\ &- \frac{1}{2}X_{N-2}\mathrm{trans}(w\mathbf{e}_1) + \frac{1}{2}X_{N-1} \\ &+ \frac{5}{8}X_{N-2}\log(\mathrm{trans}(w\mathbf{e}_1)) - \frac{1}{8}X_{N-1}\log(\mathrm{trans}(w\mathbf{e}_1))\Bigg]\end{aligned} \tag{58}$$

and

$$\begin{aligned}\mathbf{x}_d = &\frac{1}{8}[3\mathbb{I}_4 + \mathrm{Ad}(\mathrm{trans}(-w\mathbf{e}_1))]\mathbf{x}_{N-2} \\ &+ \frac{1}{8}[5\mathbb{I}_4 - \mathrm{Ad}(\mathrm{trans}(-w\mathbf{e}_1))]\mathbf{x}_{N-1}\end{aligned} \tag{59}$$

Using the definition of the covariance, $\Sigma_d = \int\mathbf{x}_d\mathbf{x}_d^T f(\mathbf{x}_d)\,d\mathbf{x}_d$, allows us to compute Σ_d using blocks of $(2k_BT)\tilde{K}^{-1}$. Let

$$(2k_BT)\tilde{K}^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & & \\ \Sigma_{12}^T & \ddots & & & \\ \vdots & & & \Sigma_{(N-2)(N-2)} & \Sigma_{(N-2)(N-1)} \\ & & & \Sigma_{(N-2)(N-1)}^T & \Sigma_{(N-1)(N-1)}\end{pmatrix}$$

then

$$\begin{aligned}\Sigma_d = \frac{1}{64}(&A_1\Sigma_{(N-2)(N-2)}A_1^T + A_1\Sigma_{(N-2)(N-1)}A_2^T \\ &+ A_2\Sigma_{(N-2)(N-1)}^T A_1^T + A_2\Sigma_{(N-1)(N-1)}A_2^T)\end{aligned} \tag{60}$$

for $A_1 = 3\mathbb{I}_4 + \mathrm{Ad}(\mathrm{trans}(-w\mathbf{e}_1))$ and $A_2 = 5\mathbb{I}_4 - \mathrm{Ad}(\mathrm{trans}(-w\mathbf{e}_1))$. This can then be used to compare the rigid-base model to other models.

One way this can be used is to compare the rigid-base model to the rigid-base-pair model. Given stiffness values for the rigid-base model, we can determine the covariance associated with $\mathbf{x}_d$. If we assume that Σ_d equals Σ_{data} as discussed in section 5, the inverse propagation techniques in section 5.1 can be employed to determine the equivalent covariance between two adjacent base pairs for the rigid-base-pair model. This allows us to determine the stiffness between rigid-base pairs that would lead to equivalent end-to-end distributions for the rigid-base-pair model and rigid-base model.

For example, consider a rigid-base model where the "left" and "right" chain have the same stiffness such that

$$\frac{1}{2k_BT}K_{2p,2p+2} = \frac{1}{2k_BT}K_{2p+1,2p+3} = \begin{pmatrix} c_1\mathbb{I}_3 & O \\ O & c_2\mathbb{I}_3\end{pmatrix}$$

and the stiffness between chains is of the form

$$\frac{1}{2k_BT}K_{2p,2p+1} = \begin{pmatrix} c_3\mathbb{I}_3 & O \\ O & c_4\mathbb{I}_3\end{pmatrix}$$

Then, for 10 bases of B-form DNA with a width w of 2 nm, a pitch L/n of 3.4 nm, and steps δs of 0.34 nm, we obtain

dx.doi.org/10.1021/jp2126015 | J. Phys. Chem. B 2012, 116, 8556−8572

$$\Sigma_d = 10^{-2}\begin{pmatrix} 8.4641 & -0.4457 & -0.0072 & -0.0366 & -3.9053 & 0.1719 \\ -0.4457 & 9.1275 & 0.0376 & 4.9258 & 0.2085 & -0.0283 \\ -0.0072 & 0.0376 & 7.5779 & -0.3084 & 0.0456 & -0.3202 \\ -0.0366 & 4.9258 & -0.3084 & 25.0342 & -0.5587 & -0.0680 \\ -3.9053 & 0.2085 & 0.0456 & -0.5587 & 24.4314 & -0.2027 \\ 0.1719 & -0.0283 & -0.3202 & -0.0680 & -0.2027 & 22.6683 \end{pmatrix}$$

for $c_1 = 20$, $c_2 = 10$, $c_3 = 5$, and $c_4 = 5$. Using the inverse propagation technique in section 5.1 for a chain of five rigid base pairs, we calculate the covariance between base pairs to be

$$\Sigma_k = 10^{-2}\begin{pmatrix} 2.7254 & -0.1777 & -0.0113 & 0.2806 & -0.1961 & 0.0403 \\ -0.1777 & 1.6725 & 0.0053 & -0.2313 & -0.2376 & 0.0398 \\ -0.0113 & 0.0053 & 1.8945 & -0.0770 & -0.0765 & -0.0801 \\ 0.2806 & -0.2313 & -0.0770 & 6.4387 & -0.4306 & 0.0322 \\ -0.1961 & -0.2376 & -0.0765 & -0.4306 & 4.1842 & -0.0240 \\ 0.0403 & 0.0398 & -0.0801 & 0.0322 & -0.0240 & 5.6671 \end{pmatrix}$$

This leads to an equivalent stiffness for the rigid-base-pair model of

$$\frac{1}{2k_BT}K_{\text{equiv}} = \Sigma_k^{-1} = \begin{pmatrix} 37.2319 & 4.0365 & 0.2177 & -1.3506 & 1.8375 & -0.2744 \\ 4.0365 & 61.0976 & 0.0841 & 2.2820 & 3.8923 & -0.4528 \\ 0.2177 & 0.0841 & 52.8888 & 0.6930 & 1.0576 & 0.7456 \\ -1.3506 & 2.2820 & 0.6930 & 15.7946 & 1.7038 & -0.0791 \\ 1.8375 & 3.8923 & 1.0576 & 1.7038 & 24.4016 & 0.0684 \\ -0.2744 & -0.4528 & 0.7456 & -0.0791 & 0.0684 & 17.6622 \end{pmatrix}$$

Here we have illustrated a method for reconciling the rigid-base and rigid-base-pair models through the use of a frame located between the two distal bases in the rigid-base model. By comparing the statistics of this new frame with the frame attached to the distal end of a rigid-base-pair model of equal length, we can relate the stiffnesses between neighboring rigid bases to an equivalent 6D stiffness between rigid-base pairs. In this example, the equivalent stiffness was assumed to be homogeneous along the length of the equivalent rigid-base-pair model.

7. CONCLUSIONS

We have presented models of DNA and RNA at a number of scales. These models, both continuous and discrete, range from anisotropic elastic filaments (the coarsest) to rigid-base models (the finest). This spectrum of models has been unified through the use of standardized Lie group notation. Moreover, new analytical verification has been presented to reconcile and compare the different models. For example, we showed that, as expected, the bi-rod model does indeed converge to the single filament model as the filament length increases.

Additionally, molecular dynamics simulations and AFM measurements have been introduced to validate the use of some of these models. The molecular dynamics simulations and time series analysis were performed to validate the use of Gaussian models over exponential coordinates. The AFM measurements of naked and cisplatinated DNA have been shown to correspond well with the single elastic filament model in the plane. This example, which involved determining the bend angle, also highlights how these models can be used to determine physical characteristics of DNA or RNA.

Future investigations can be performed to further determine the lengths and stiffnesses for which each of these models is best suited. This may include determining how to appropriately transition from one model to another.

■ ASSOCIATED CONTENT

Supporting Information

Additional details regarding the classic multivariate Ornstein–Ulenbeck process used to derive eq 36. Also, a derivation is provided for the covariance matrix of the "midpoint" of the distal end of the rigid-base model presented in section 6. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: gregc@jhu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions. This work was funded in part by NSF grant IIS-0915542. This research was also supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

■ REFERENCES

(1) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499–512.
(2) Vologodskii, A. *Topology and Physics of Circular DNA*; CRC Press: Boca Raton, FL, 1992.
(3) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science* **1994**, *265*, 1599–1600.
(4) Bustamante, C.; Smith, S. B.; Liphardt, J.; Smith, D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279–285.
(5) Nelson, P. *Biological Physics: Energy, Information, Life*; W. H. Freeman: New York, 2004.
(6) Widom, J. *Q. Rev. Biophys.* **2001**, *34*, 269–324.
(7) Rippe, K.; von Hippel, P. H.; Langowski, J. *Trends Biochem. Sci.* **1995**, *20*, 500–506.
(8) Shore, D.; Langowski, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4833–4837.
(9) Wiggins, P. A.; van der Heijden, T.; Moreno-Herrero, F.; Spakowitz, A.; Phillips, R.; Widom, J.; Dekker, C.; Nelson, P. C. *Nat. Nanotechnol.* **2006**, *1*, 137–141.
(10) Cloutier, T. E.; Widom, J. *Mol. Cell* **2004**, *14*, 355–362.
(11) Lankas, F.; Sponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* **2003**, *85*, 2872–2883.
(12) Mazur, A. K. *Biophys. J.* **2006**, *91*, 4507–4518.
(13) Hagerman, P. J. *Annu. Rev. Biochem.* **1990**, *29*, 755–781.
(14) Okonogi, T. M.; Alley, S. C.; Reese, A. W.; Hopkins, P. B.; Robinson, B. H. *Biophys. J.* **2000**, *78*, 2560–2571.
(15) Afonin, K. A.; Bindewald, E.; Yaghoubian, A. J.; Voss, N.; Jacovetty, E.; Shapiro, B. A.; Jaeger, L. *Nat. Nanotechnol.* **2010**, *5*, 676–682.
(16) Chirikjian, G. S. *Stochastic Models, Information Theory, and Lie Groups: Analytic Methods and Modern Applications*; Birkhäuser: Boston, MA, 2011; Vol. 2.
(17) Chirikjian, G. S.; Wang, Y. *Phys. Rev. E* **2000**, *62*, 880–892.
(18) Chirikjian, G. S. *J. Phys.: Condens. Matter* **2010**, *22*.
(19) Coleman, B. D.; Olson, W. K.; Swigon, D. *J. Chem. Phys.* **2003**, *118*, 7127–7140.
(20) Mehraeen, S.; Sudhanshu, B.; Koslover, E. F.; Spakowitz, A. J. *Phys. Rev. E* **2008**, *77*.
(21) Chirikjian, G. S.; Kyatkin, A. B. *Engineering applications of noncommutative harmonic analysis: with emphasis on rotation and motion groups*; CRC Press: Boca Raton, FL, 2001.

(22) Kleinert, H. *Path integrals in quantum mechanics, statistics, and polymer physics*, 2nd ed.; World Scientific: River Edge, NJ, 1995.
(23) Odijk, T. *Macromolecules* **1995**, *28*, 7016−7018.
(24) Yamakawa, H. *Helical wormlike chains in polymer solutions*; Springer: Berlin, Germany, 1997.
(25) Nyquist, H. *Phys. Rev.* **1928**, *32*, 110−113.
(26) Callen, H. B.; Welton, T. A. *Phys. Rev.* **1951**, *83*, 34−40.
(27) Chirikjian, G. S.; Kyatkin, A. B. *J. Fourier Anal. Appl.* **2000**, *6*, 583−606.
(28) Carmona, M.; Magasanik, B. *J. Mol. Biol.* **1996**, *261*, 348−356.
(29) Crothers, D. M.; Gartenberg, M. R.; Shrader, T. E. *Methods Enzymol.* **1991**, *208*, 118−146.
(30) Erie, D. A.; Yang, G.; Schultz, H. C.; Bustamante, C. *Science* **1994**, *266*, 1562−1566.
(31) Griffith, J. D.; Makhov, A.; Zawel, L.; Reinberg, D. *J. Mol. Biol.* **1995**, *246*, 576−584.
(32) Pérez-Martin, J.; Espinosa, M. *Science* **1993**, *260*, 805−807.
(33) Rees, W. A.; Keller, R. W.; Vesenka, J. P.; Yang, G.; Bustamante, C. *Science* **1993**, *260*, 1646−1649.
(34) van der Vliet, P. C.; Verrijzer, C. P. *Bioessays* **1993**, *15*, 25−32.
(35) Allen, D. J.; Makhov, A.; Grilley, M.; Taylor, J.; Thresher, R.; Modrich, P.; Griffith, J. D. *EMBO J.* **1997**, *16*, 4467−4476.
(36) Lobell, R. B.; Schleif, R. F. *Science* **1990**, *250*, 528−533.
(37) Rippe, K.; Guthold, M.; von Hippel, P. H.; Bustamante, C. *J. Mol. Biol.* **1997**, *270*, 125−138.
(38) Su, W.; Porter, S.; Kustu, S.; Echols, H. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5504−5508.
(39) Wong, O. K.; Guthold, M.; Erie, D. A.; Gelles, J. *PLoS Biol.* **2008**, *6*, 2028−2042.
(40) Wyman, C.; Grotkopp, E.; Bustamante, C.; Nelson, H. C. *EMBO J.* **1995**, *14*, 117−123.
(41) Rivetti, C.; Guthold, M. In *RNA polymerases and associated factors*; Adhya, S. L., Garges, S., Eds.; Elsevier Academic Press: San Diego, CA, 2003; Vol. *370*, pp 34−50.
(42) Rivetti, C.; Guthold, M.; Bustamante, C. *EMBO J.* **1999**, *18*, 4464−4475.
(43) Guthold, M.; Zhu, X.; Rivetti, C.; Yang, G.; Thomson, N. H.; Kasas, S.; Hansma, H. G.; Smith, B.; Hansma, P. K.; Bustamante, C. *Biophys. J.* **1999**, *77*, 2284−2294.
(44) Bustamante, C.; Guthold, M.; Zhu, X.; Yang, G. *J. Biol. Chem.* **1999**, *274*, 16665−16668.
(45) Mirny, L.; Slutsky, M.; Wunderlich, Z.; Tafvizi, A.; Leith, J.; Kosmrlj, A. *J. Phys. A: Math. Theor.* **2009**, *42*.
(46) von Hippel, P. H.; Berg, O. G. *J. Biol. Chem.* **1989**, *264*, 675−678.
(47) Tafvizi, A.; Mirny, L. A.; van Oijen, A. M. *ChemPhysChem* **2011**, *12*, 1481−1489.
(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(49) *Jmol: an open-source Java viewer for chemical structures in 3D*, version 12.0.41; http://www.jmol.org.
(50) Rice, P. A.; Yang, S.; Mizuuchi, K.; Nash, H. A. *Cell* **1996**, *87*, 1295−1306.
(51) Mouw, K. W.; Rice, P. A. *Mol. Microbiol.* **2007**, *63*, 1319−1330.
(52) Schultz, S. C.; Shields, G. C.; Steitz, T. A. *Science* **1991**, *253*, 1001−1007.
(53) Abrescia, N. G. A.; Malinina, L.; Subirana, J. A. *J. Mol. Biol.* **1999**, *294*, 657−666.
(54) Todd, R. C.; Lippard, S. J. *J. Inorg. Biochem.* **2010**, *104*, 902−908.
(55) Zamble, D. B.; Mikata, Y.; Eng, C. H.; Sandman, K. E.; Lippard, S. *J. Inorg. Biochem.* **2002**, *91*, 451−462.
(56) He, Q.; Liang, C. H.; Lippard, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5768−5772.
(57) Hambley, T. W. *J. Chem. Soc., Dalton Trans.* **2001**, 2711−2718.
(58) Chaney, S. G.; Campbell, S. L.; Bassett, E.; Wu, Y. *Crit. Rev. Oncol. Hemat.* **2005**, *53*, 3−11.
(59) Lin, X.; Ramamurthi, K.; Mishima, M.; Kondo, A.; Christen, R. D.; Howell, S. B. *Cancer Res.* **2001**, *61*, 1508−1516.
(60) Raymond, E.; Faivre, S.; Chaney, S.; Woynarowski, J.; Cvitkovic, E. *Mol. Cancer Ther.* **2002**, *1*, 227−235.
(61) Vaisman, A.; Varchenko, M.; Umar, A.; Kunkel, T. A.; Risinger, J. I.; Barrett, J. C.; Hamilton, T. C.; Chaney, S. G. *Cancer Res.* **1998**, *58*, 3579−3585.
(62) Bellon, S. F.; Coleman, J. H.; Lippard, S. J. *Biochemistry* **1991**, *30*, 8026−8035.
(63) Jamieson, E. R.; Lippard, S. J. *Chem. Rev.* **1999**, *99*, 2467−2498.
(64) Pascoe, J. M.; Roberts, J. J. *Biochem. Pharmacol.* **1974**, *23*, 1345−1357.
(65) Bernal-Méndez, E.; Boudvillain, M.; González-Vlchez, F.; Leng, M. *Biochemistry* **1997**, *36*, 7281−7287.
(66) Kelland, L. *Nat. Rev. Cancer* **2007**, *7*, 573−584.
(67) Wang, D.; Lippard, S. J. *Nat. Rev. Drug Discovery* **2005**, *4*, 307−320.
(68) Wang, H.; Yang, Y.; Schofield, M. J.; Du, C.; Fridman, Y.; Lee, S. D.; Larson, E. D.; Drummond, J. T.; Alani, E.; Hsieh, P.; Erie, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14822−14827.
(69) Bellon, S. F.; Lippard, S. J. *Biophys. Chem.* **1990**, *35*, 179−188.
(70) Gelasco, A.; Lippard, S. J. *Biochemistry* **1998**, *37*, 9230−9239.
(71) Stehlikova, K.; Kostrhunova, H.; Brabec, V. *Nucleic Acids Res.* **2002**, *30*, 2894−2898.
(72) Takahara, P. M.; Frederick, C. A.; Lippard, S. J. *J. Am. Chem. Soc.* **1996**, *118*, 12309−12321.
(73) Chirikjian, G. S.; Wang, Y. F. *Phys. Rev. E* **2000**, *62*, 880−892.
(74) Chirikjian, G. S.; Kyatkin, A. B. *J. Fourier Anal. Appl.* **2000**, *6*, 583−606.
(75) Zhou, Y.; Chirikjian, G. S. *J. Chem. Phys.* **2003**, *119*, 4962−4970.
(76) Davies, M. S.; Berners-Price, S. J.; Hambley, T. W. *J. Am. Chem. Soc.* **1998**, *120*, 11380−11390.
(77) Rivetti, C.; Guthold, M.; Bustamante, C. *J. Mol. Biol.* **1996**, *264*, 919−932.
(78) Ansari, A. R.; Bradley, R. A. *Ann. Math. Stat.* **1960**, *31*, 1174−1189.
(79) Dutta, S. Studying the interaction of cancer and thrombosis therapeutics with protein and DNA. Ph.D. Thesis, Wake Forest University, Winston-Salem, NC, 2011.
(80) Dutta, S.; Rivetti, C.; Gassman, N. R.; Young, C. G.; Jones, B. T.; Scarpinato, K.; M., G. Analysis of single cisplatin-induced DNA bends by atomic force microscopy and simulations. *Microsc. Microanal.*, to be submitted for publication, **2012**.
(81) Takahara, P. M.; Rosenzweig, A. C.; Frederick, C. A.; Lippard, S. J. *Nature* **1995**, *377*, 649−652.
(82) Dunham, S. U.; Turner, C. J.; Lippard, S. J. *J. Am. Chem. Soc.* **1998**, *120*, 5395−5406.
(83) Moakher, M.; Maddocks, J. H. *Arch. Ration. Mech. Anal.* **2005**, *177*, 53−91.
(84) Lu, X. J.; Babcock, M. S.; Olson, W. K. *J. Biomol. Struct. Dyn.* **1999**, *16*, 833−843.
(85) Wang, Y.; Chirikjian, G. S. In *Advances in Robotic Kinematics*; Lenarcic, J., Roth, B., Eds.; Springer: Dordrecht, The Netherlands, 2006; pp 95−102.
(86) Wang, Y.; Chirikjian, G. S. *IEEE Trans. Robotics* **2006**, *22*, 591−602.
(87) Wang, Y.; Chirikjian, G. S. Proceedings of the IEEE International Conference on Robotics and Automation, Orlando, FL, May 15−19, 2006; pp 1848−1853.
(88) Wang, Y.; Chirikjian, G. S. *Int. J. Robotics Res.* **2008**, *27*, 1258−1273.
(89) Becker, N. B.; Everaers, R. *Phys. Rev. E* **2007**, *76*.
(90) Becker, N. B.; Everaers, R. *J. Chem. Phys.* **2009**, *130*.
(91) Park, W.; Reed, K. B.; Okamura, A. M.; Chirikjian, G. S. Proceedings of the IEEE International Conference on Robotics and Automation, Anchorage, AK, May 3−7, 2010; pp 3703−3708.
(92) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. *Nucleic Acids Res.* **2010**, *38*, 299−313.

(93) Lankas, F.; Gonzalez, O.; Heffler, L. M.; Stoll, G.; Moakher, M.; Maddocks, J. H. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10565−10588.
(94) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A. et al. *Amber 8*; University of California: San Francisco, CA, 2004.
(95) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*.