Conformational Statistics of Semiflexible Macromolecular Chains with Internal Joints

Yu Zhou[†] and Gregory S. Chirikjian*,[‡]

*Department of Mechanical Engineering, State University of New York at Stony Brook,
Stony Brook, New York 11794, and Department of Mechanical Engineering, Johns Hopkins University,
Baltimore, Maryland 21218*

ABSTRACT: Fluctuations in the bending angles at internal irregularities of DNA and RNA (such as symmetric loops, bulges, and nicks/gaps) have been observed from various experiments. However, little effort has been made to computationally predict and explain the statistical behavior of semiflexible chains with internal defects. In this paper, we describe the general structure of these macromolecular chains as inextensible elastic chains with one or more internal joints which have limited ranges of rotation and propose a method to compute the probability density functions of the end-to-end pose of these macromolecular chains. Our method takes advantage of the operational properties of the noncommutative Fourier transform for the group of rigid-body motions in three-dimensional space, SE(3). Two representative types of joints, the hinge for planar rotation and the ball joint for spatial rotation, are discussed in detail. The proposed method applies to various stiffness models of semiflexible chainlike macromolecules. Examples are calculated using the Kratky−Porod model with specified stiffness, angular fluctuation, and joint locations. Entropic effects associated with internal angular fluctuations of semiflexible macromolecular chains with internal joints can be computed using this formulation. Our method also provides a potential tool to detect the existence of internal irregularities.

1. Introduction

The internal bending and flexibility of DNA has attracted considerable attention because of its important role in packaging, recombination, and transcription.

First discovered in the early 1980s,[1] intrinsically bent DNAs have been observed and studied experimentally.[2−8] Local structures, which are more complex than rigid bends, have also been observed from various experiments. In particular, errors during replication and recombination of double-stranded DNA may result in mismatched basepairs or the insertion/deletion of nucleotides. This results in internal structures such as symmetric loops, bulges, and gaps/nicks. These structures usually have different flexibility than regular double-helical DNA and may result in bends. Usually, these bends are not rigid but exhibit certain ranges of fluctuation. On the basis of gel electrophoresis experiments, Guo and Tullius showed that a single-nucleoside gap in a DNA duplex leads to an anisotropic, directional bend in the DNA helix axis.[9] Moreover, a gap is a site of anisotropic flexibility rather than a static, fixed bend. The bending angle caused by a single gap was estimated at around 12°. Roll et al. reported a fluctuating bending angle centered around 17° at the nick site of DNA based on nuclear magnetic resonance (NMR) experiments and molecular dynamics (MD) simulations.[10] Kahn et al. studied the bending and flexibility induced by symmetric internal loops by gel electrophoresis experiments and cyclization kinetics analysis and found that the internal loop induced high local isotropic flexibility.[11] They reported a root-mean-square (rms) bending angle of 43 ± 5° for 3-basepair loops, which reflects a fairly wide range of angular distribution. Moreover, the bending angle at a loop is essentially independent of the loop size. Zacharias and Hagerman quantified the bending angles caused by symmetric internal loops in RNA and reported an rms angular fluctuation of 16 ± 5° for each basepair in the absence of Mg^{2+} and 13 ± 4° in the presence of Mg^{2+}.[12] As for bulge-induced bending, fluorescence resonance energy transfer (FRET) measurements indicated a bending of 50°−70° for an A3 DNA bulge and 85°−105° for an A5 DNA bulge.[13] An NMR study also showed a bending angle of 90 ± 14° for an RNA bulge with five unpaired bases AAUAA[14] and 73 ± 11° for the A5 RNA bulge.[15] Furthermore, from the results of transient electric birefringence (TEB) measurements, Zacharias and Hagerman gave an estimation of a 10°−20° increment of bending angle induced per extra base.[16] They also stated that a given bulge-induced bend could be regarded as a relatively narrow distribution of angles with a nonzero mean angle. Other works have shown strong agreement with the above results regarding bulge-induced bending.[17−20]

Although much experimental attention has been given to the bending fluctuations caused by internal structures, so far no previous work has been published on how to computationally predict and explain the statistical behavior of macromolecular chains with those internal structures.

In contrast, many models have been put forth to describe the behavior of intrinsically straight semiflexible macromolecular chains and to explain experimental data. Widely used models include the freely jointed chain model,[21] the Kratky−Porod wormlike chain model,[22−24] the Yamakawa helical wormlike chain model,[25] and the Marko−Siggia double-helix chain model (describing the double-helix structure of DNA).[26,27] In our previous work, we proposed a general formulation for inextensible semiflexible macromolecular chains, which includes several above-mentioned polymer models as subcases.[28]

Meanwhile, various computational methods have been developed to estimate the conformational statistics of intrinsically straight macromolecular chains. Common techniques include Monte Carlo, matrix-generator, direct enumeration, and path

[†] State University of New York at Stony Brook.
[‡] Johns Hopkins University.
* Corresponding author. E-mail: gregc@jhu.edu.

integration.[29−37] For some basic polymer models, closed-form formulas have been derived for the probability distributions of loop length,[38] segmental orientations,[39] trajectories of a segment,[40] radius of gyration,[41] end-to-end position and distance,[42−49] and moments.[46,49,50] In our previous work, we showed that the probability density function (PDF) of the end-to-end pose for the most general model of the inextensible semiflexible macromolecular chain can be obtained by either solving a diffusion equation or convolving PDFs for short segments of the chain.[28,51−53]

Recently some effort has been given to describing macromolecular chains with internal rigid bends. Models have been proposed to describe the relationship between bending stiffness and DNA sequence.[54−56] Matrix-generator and Monte Carlo methods were the first tools used to study the influence of thermal fluctuation on the size and shape of bent DNA.[57] While classical semiflexible polymer theories are in general inadequate to describe the conformational statistics of these systems, Rivetti et al. made an effort to extend the applicability of the KP wormlike chain model to bent DNA and derived closed-form expressions for the mean-square end-to-end distance.[58] In our previous work, we proposed a general method to compute the full probability distribution of the end-to-end pose for the bent macromolecular chain with arbitrary stiffness and chirality parameters.[59] Ignoring the interactions between distal segments of the same macromolecular chain, our method applies to semiflexible inextensible chiral elastic macromolecular chains with internal rigid bends and twists.

Despite all these efforts, there is still a lack of effective tools to estimate the conformational statistics of semiflexible chains with a finite number of points of increased flexibility. In this paper, we propose a method capable of computing the PDF of the end-to-end pose for semiflexible macromolecular chains with internal flexible structures, such as loops, bulges, and nicks/gaps, ignoring the interactions between distal segments of the same macromolecular chain. In our work, the internal structures involving a small number of basepairs are treated as joints with limited ranges of rotation. Our method is based on our general statistical model of intrinsically straight inextensible semiflexible macromolecular chains[28] and uses the concept of the Fourier transform for the group of rigid-body motion in three-dimensional space, SE(3).

In section 2, we briefly review our general statistical model of intrinsically straight inextensible semiflexible macromolecular chains and relevant mathematical background on the Fourier transform for SE(3). In section 3, we present the idea of reducing the internal flexible structures into joints with limited ranges of motion, introduce a general method for computing the PDF of the end-to-end pose for a jointed macromolecular chain, and derive the formulas for two representative types of joints: the hinge and the ball joint. In section 4, we discuss the extreme case—the completely rigid chain with internal joints—and derive the closed form solution for the PDF of the end-to-end distance. In section 5, we demonstrate with examples how to implement our method to compute the desired PDFs and discuss the computational results.

2. Review of the General Inextensible Semiflexible Macromolecular Chain Model

The method presented in this paper is built on a general model of the intrinsically straight inextensible semiflexible macromolecular chain which was derived in our previous work.[28] We briefly review it in this section.
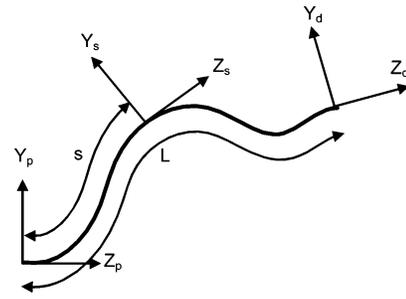
Figure 1. Frames of reference attached to a semiflexible macromolecular chain. $(X_pY_pZ_p)$ represents the frame of reference at the proximal end, $(X_sY_sZ_s)$ represents the frame of reference at arc length s, and $(X_dY_dZ_d)$ represents the frame of reference at the distal end. In the plot, X axes are not displayed but can be derived from Y and Z axes following the right-hand rule.

Given a macromolecular chain, one can define a local frame of reference attached to any point on the backbone of the macromolecular chain such that the Z axis of this frame points along the tangent of the backbone (Figure 1). The frame of reference at the proximal end of the chain is defined by $(\mathbf{R}(0),\mathbf{a}(0)) = (\mathbf{I},\mathbf{0})$, and the pose of the frame of reference at arc length s with respect to the frame of reference at the proximal end is defined by $(\mathbf{R}(s),\mathbf{a}(s))$. Here $\mathbf{R}(s)$ is a 3×3 rotation matrix defining the relative orientation at arc length s, $\mathbf{a}(s)$ is a 3×1 translation vector defining the relative position, $\mathbf{I}$ denotes the 3×3 identity matrix, and L denotes the total arc length of the chain.

Under the assumption that the semiflexible macromolecular chain is inextensible, and the variation of the conformation of the chain is influenced by Gaussian white noise (i.e., Brownian motion forcing), the PDF of the pose of the frame of reference at arc length s with respect to the frame of reference at the proximal end of the chain can be formulated as the path integral on the group of rotations in three-dimensional space, SO(3).[28]

$$f(\mathbf{a},\mathbf{R},s) = \int_{\mathbf{R}(0)}^{\mathbf{R}(s)} \delta\left(\mathbf{a}(s) - \int_0^s \mathbf{R}(\epsilon)\mathbf{e}_3 \, d\epsilon\right) e^{-\int_0^s U(t)dt} D[\mathbf{R}(s)]$$

(1)

where $\mathbf{e}_3 = [0,0,1]'$ and $\delta(.)$ is the Dirac delta function. Here $U(s)$ is the elastic energy per unit length measured in units of k_BT, where k_B denotes the Boltzmann constant and T denotes the temperature. It has the form

$$U(s) = \frac{1}{2}\omega(s)^T\mathbf{B}\omega(s) - \mathbf{b}^T\omega(s) + c$$

(2)

where $\omega(s)$ is called the spatial angular velocity at arc length s (which replaces time in the current context), $\mathbf{B}$ is a 3×3 positive semidefinite symmetric matrix called the stiffness matrix, $\mathbf{b}$ is a 3×1 vector describing the chirality, and c is a constant. Many elastic models fit into this general formulation, including the Kratky−Porod model, the Yamakawa model, and the Marko−Siggia model.[28] After some mathematical manipulation, one can obtain from (1) the following partial differential equation on SE(3) which describes the PDF of the pose of the frame of reference at arc length s with respect to the frame of reference at the proximal end of the chain[28]

$$\frac{\partial f(\mathbf{a},\mathbf{R},s)}{\partial s} = \left(\frac{1}{2}\sum_{k,l=1}^3 D_{lk}\tilde{X}_l^R \tilde{X}_k^R + \sum_{l=1}^3 d_l\tilde{X}_l^R - \tilde{X}_6^R\right)f(\mathbf{a},\mathbf{R},s)$$

(3)

with the initial condition $f(\mathbf{a},\mathbf{R},0) = \delta(\mathbf{a})\delta(\mathbf{R})$. In eq 3, $\mathbf{D} =$

$\mathbf{B}^{-1}$, $\mathbf{d} = -\mathbf{B}^{-1}\mathbf{b}$, and $\tilde{X}_i^R$ are the differential operators defined for SE(3) ($1 \leq i \leq 6$).

In our previous works,[28,51] a methodology for solving equations such as (3) was developed using the technique of the Fourier transform for SE(3). The matrix elements of this transform are[51,61]

$$\hat{f}_{l',m';l,m}^r(p,s) = \int_{SE(3)} f(\mathbf{a},\mathbf{R},s)\overline{U_{l,m;l',m'}^r(\mathbf{a},\mathbf{R};p)} \, d\mathbf{R} \, d\mathbf{a} \quad (4)$$

where for each $r \in Z$ and $p \in \mathscr{R}^+$, $U^r(\mathbf{a},\mathbf{R};p)$ is the irreducible unitary representation matrix of SE(3), p is the frequency factor introduced by the transform, and $d\mathbf{R}d\mathbf{a}$ is the invariant integration measure for SE(3). Applying the Fourier transform for SE-(3) on both sides of (3), one can obtain a system of linear ordinary differential equations in the form of

$$\frac{d\hat{\mathbf{f}}^r(p,s)}{ds} = \mathbf{B}^r(p)\,\hat{\mathbf{f}}^r(p,s) \quad (5)$$

where $\mathbf{B}^r$ is a matrix function that is independent of s under the assumption that the stiffness and chirality of the intrinsically straight macromolecular chain are constant along the whole chain. Solving (5), one obtains

$$\hat{\mathbf{f}}^r(p,s) = e^{s\mathbf{B}^r(p)} \quad (6)$$

Then, the desired PDF can be obtained by the inverse transform[51,61]

$$f(\mathbf{a},\mathbf{R},s) = \frac{1}{2\pi^2} \sum_{r=-\infty}^{\infty} \sum_{l'=|r|}^{\infty} \sum_{l=|r|}^{\infty} \sum_{m'=-l'}^{l'} \sum_{m=-l}^{l} \int_0^\infty \hat{f}_{l,m;l',m'}^r(p,s) U_{l',m';l,m}^r(\mathbf{a},\mathbf{R};p) p^2 \, dp \quad (7)$$

3. PDFs for the Semiflexible Macromolecular Chains with Internal Revolute Joints

The model represented by (3) applies to intrinsically straight inextensible elastic macromolecular chains, and the solution represented by (6) applies to macromolecular chains with constant stiffness and chirality. Some extensions need to be made to the method reviewed in section 2 so that it can be applied to computing the conformational statistics of chains with the internal structures mentioned in section 1.

3.1. General Idea. From the collection of works describing the internal structures of DNA and RNA which are of interest in this paper,[9−20] we notice that those internal structures usually introduce nonrigid local bending with angular fluctuation at those irregular sites on the macromolecular chains, with either larger ranges of rotation (e.g., symmetric loops) or smaller ranges of rotation (e.g., bulges). We also notice that usually only a small number of mismatched basepairs are involved in those local structures (usually <6 bps). Therefore, we model the conformational statistics of this category of macromolecular chains by representing internal structures as revolute (rotational) joints with limited ranges of motion where rotation angles follow certain probability distributions.

We consider the case of a macromolecular chain containing one internal joint (Figure 2), though the same methodology can be easily extended to fit the multijoint case. As a frame of reference traverses the backbone of the chain, one can simply divide the chain into two segments. Each segment has a PDF describing the ensemble of all possible motions of its distal end relative to its proximal end. The two segments are connected
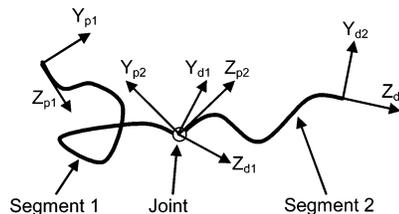
Figure 2. Diagram of a macromolecular chain with an internal joint. ($X_{pi}Y_{pi}Z_{pi}$) represents the frame of reference at the proximal end, and ($X_{di}Y_{di}Z_{di}$) represents the frame of reference at the distal end. In the plot, X axes are not displayed but can be derived from Y and Z axes following the right-hand rule.

by a joint which allows random rotation within a certain range and following a certain probability distribution. As a result, the PDF of the end-to-end pose for the whole chain can be obtained as the convolution of three PDFs

$$f(\mathbf{a},\mathbf{R},L) = (f_1 * f_j * f_2)(\mathbf{a},\mathbf{R},L) \quad (8)$$

where $f_1 = f_1(\mathbf{a},\mathbf{R},L_1)$ is the PDF of the pose of the frame of reference attached at the distal end of segment 1 relative to its own proximal end, $f_2 = f_2(\mathbf{a},\mathbf{R},L_2)$ is the PDF of the pose of the frame of reference attached at the distal end of segment 2 relative to its own proximal end, and f_j is the PDF of the pose of the frame of reference attached at the proximal end of segment 2 relative to the distal end of segment 1. f_j describes the junction between the two segments. Whereas $f_1(\mathbf{a},\mathbf{R},L_1)$ can be obtained by solving (3) with $s = L_1$ and $f_2(\mathbf{a},\mathbf{R},L_2)$ can be obtained by solving (3) with $s = L_2$, f_j has the following form

$$f_j(\mathbf{a},\mathbf{R}) = \delta(\mathbf{a})\,f(\mathbf{R}) \quad (9)$$

The Dirac delta function in translation means that the two segments meet at a point rather than being translated in space relative to each other. Meanwhile, the rotation at the joint follows a certain probability distribution. The convolution in (8) is a convolution defined on SE(3). Letting $\mathbf{g}$ and $\mathbf{h}$ denote any group elements in SE(3), and letting $f_a(\mathbf{g})$ and $f_b(\mathbf{g})$ be two arbitrary functions defined on SE(3), the convolution on SE(3) is defined as[51]

$$(f_a * f_b)(\mathbf{g}) = \int_{SE(3)} f_a(\mathbf{h})\,f_b(\mathbf{h}^{-1} \circ \mathbf{g})\,d(\mathbf{h}) \quad (10)$$

where the little circle is the group multiplication operator for rigid-body motions. Equation 8 can be extended further to solve the multijoint case by convolving more PDFs of involved segments and joints.

Since each PDF is a function of six pose variables (three position variables and three orientation variables), it is very complicated to calculate (8) directly, and the situation will become worse when more joints are involved. However, the Fourier transform for functions of SE(3)-valued argument provides a relatively easy way to compute (8). Applying the transform on both sides of (8), one obtains[51,60]

$$\hat{\mathbf{f}}^r(p) = \hat{\mathbf{f}}_2^r(p)\hat{\mathbf{f}}_j^r(p)\hat{\mathbf{f}}_1^r(p) \quad (11)$$

It is a general property that the Fourier transform of a convolution of functions is the product of the Fourier transforms of those functions multiplied in the reverse order. Since in this context the Fourier transforms are matrix-valued functions, the order of multiplication matters. Whereas $\hat{\mathbf{f}}_1^r$ and $\hat{\mathbf{f}}_2^r$ can be

computed from (6), $\hat{\mathbf{f}}_j^r$ can be obtained as the Fourier transform of f_j with matrix elements

$$\hat{f}_{jl',m';l,m}^r(p) = \int_{SE(3)} f_j(\mathbf{a},\mathbf{R})\overline{U_{l,m;l',m'}^r(\mathbf{a},\mathbf{R};p)}\ d\mathbf{R}\ d\mathbf{a}$$

$$= \int_{SE(3)} \delta(\mathbf{a})\,f(\mathbf{R})\overline{U_{l,m;l',m'}^r(\mathbf{a},\mathbf{R};p)}\ d\mathbf{R}\ d\mathbf{a}$$

$$= \int_{SO(3)} f(\mathbf{R})\overline{U_{l,m;l',m'}^r(\mathbf{0},\mathbf{R};p)}\ d\mathbf{R} \qquad (12)$$

Then, substituting $\hat{\mathbf{f}}^r$ into (7), one obtains the PDF of the end-to-end pose of the macromolecular chain with an internal revolute joint. Equation 11 can be easily extended to the multijoint case by multiplying the SE(3) Fourier transforms of more PDFs of involved segments and joints in the reverse order.

Besides the full PDF of the end-to-end pose, the PDF of the end-to-end distance is of particular interest because it is possible to be measured from experiments. Here, we parametrize the translation with spherical coordinates, with the letter "a" denoting the radial distance from the origin, θ denoting the polar angle, and ϕ denoting the azimuthal angle. We also parametrize the rotation with the *ZXZ* Euler angles, with α, β, and γ denoting the three Euler angles. Then one can compute the PDF of the end-to-end distance, $f(a)$, by integrating out other variables from (7)

$$f(a) = \frac{a^2}{2\pi^2}\int_0^\pi \int_0^{2\pi}\int_{SO(3)} f(\mathbf{a},\mathbf{R})\,\sin\theta\ d\mathbf{R}\ d\phi\ d\theta$$

$$= \frac{2a^2}{\pi}\int_0^\infty \hat{f}_{0,0;0,0}^0(p)\frac{\sin pa}{pa}\,p^2\ dp \qquad (13)$$

Whereas we have shown a general method to compute the PDF of the end-to-end pose of a macromolecular chain with internal joints, in the following we will focus on the PDF at the revolute joint, $f_j(\mathbf{a},\mathbf{R})$, and its Fourier transform. Here, two representative types of revolute joints, the hinge and the ball joint, will be discussed in detail.

3.2. The Hinge Case. When the angular fluctuation at an internal structure is constrained to move in a plane, the local structure can be modeled as a planar revolute joint, i.e., a hinge (Figure 3).
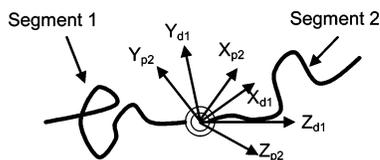
Figure 3. Diagram of a macromolecular chain with an internal hinge.

The workspace of a hinge with a limited range of rotation can be depicted as a sector with a unit radius (Figure 4). In Figure 4, X_h represents the orientation of the hinge axis which is determined by the *ZXZ* Euler angles α_0, β_0, and γ_0. The rigid-body motion at a hinge consists of only one rotation β_h about X_h.

We notice that an arbitrarily located sector with a unit radius, as shown in Figure 4, results from moving a sector of same area, which is inside the $Y-Z$ plane (Figure 5), to the new position determined by α_0, β_0, and γ_0 (Figure 4). Therefore, in general, denoting $f(\beta_h)$ as the probability distribution of the bending angle β_h at the hinge, the PDF of rigid-body motion on SE(3) at this hinge can be obtained as a convolution

$$f_j(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = (f_{j1}*f_{j2})(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) \qquad (14)$$
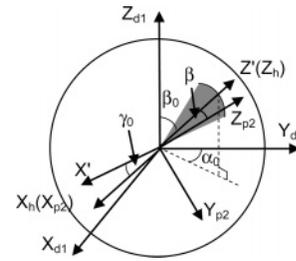
Figure 4. Workspace of a hinge with a limited range of rotation. $(X_{d1}Y_{d1}Z_{d1})$ represents the frame of reference at the distal end of segment 1, and $(X_{p2}Y_{p2}Z_{p2})$ represents the frame of reference at the proximal end of segment 2. $(X'Y'Z')$ is an intermediate frame of reference obtained by rotating $(X_{d1}Y_{d1}Z_{d1})$ by an angle of α_0 about its Z axis followed by rotating the current frame an angle of β_0 about its X axis. $(X_hY_hZ_h)$ is obtained by rotating $(X'Y'Z')$ an angle of γ_0 about its Z axis.
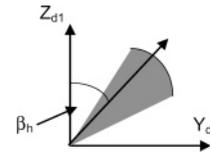
Figure 5. Workspace of a hinge constrained to move in the $Y-Z$ plane.

where

$$\alpha \in [0,2\pi], \quad \beta \in [0,\pi], \quad \text{and}\ \gamma \in [0,2\pi] \qquad (15)$$

and

$$f_{j1}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \delta(\mathbf{R}(\alpha_0,\beta_0,\gamma_0)^{-1}\mathbf{R}(\alpha,\beta,\gamma))\delta(\mathbf{a})$$

$$= \frac{\delta(\mathbf{a})\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma - \gamma_0)}{\sin\beta} \qquad (16)$$

$$f_{j2}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \frac{\delta(\mathbf{a})\delta(\alpha)f(\beta)\delta(\gamma)}{\sin\beta} \qquad (17)$$

Then the Fourier transform of $f_j(\mathbf{a},\mathbf{R})$ on SE(3) can be obtained as

$$\hat{\mathbf{f}}_j^r(p) = \hat{\mathbf{f}}_{j2}^r(p)\ \hat{\mathbf{f}}_{j1}^r(p) \qquad (18)$$

where

$$\hat{f}_{j1l',m';l,m}^r(p) = \int_{SE(3)} f_{j1}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma))\overline{U_{l,m;l',m'}^r(\mathbf{a},\mathbf{R};p)}\ d\mathbf{R}\ d\mathbf{a}$$

$$= \int_{SE(3)} \frac{\delta(\mathbf{a})\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma - \gamma_0)}{\sin\beta} \times$$

$$\overline{U_{l,m;l',m'}^r(\mathbf{a},\mathbf{R};p)}\ d\mathbf{R}\ d\mathbf{a}$$

$$= \int_{SO(3)} \frac{\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma - \gamma_0)}{\sin\beta}\overline{U_{l,m;l',m'}^r(\mathbf{0},\mathbf{R};p)}\ d\mathbf{R}$$

$$= \delta_{l,l'}\int_0^{2\pi}\int_0^\pi\int_0^{2\pi} \frac{\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma - \gamma_0)}{\sin\beta} \times$$

$$e^{im\alpha}(-1)^{m'-m}P_{m,m'}^l(\cos\beta)e^{im'\gamma}\,\sin\beta\ d\alpha\ d\beta\ d\gamma$$

$$= \delta_{l,l'}e^{im\alpha_0}P_{m',m}^l(\cos\beta_0)e^{im'\gamma_0} \qquad (19)$$

$$\hat{f}^{\mathrm{r}}_{j2l',m';l,m}(p) = \int_{SE(3)} f_{j2}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)} \, d\mathbf{R} \, d\mathbf{a}$$

$$= \int_{SE(3)} \frac{\delta(\mathbf{a})\delta(\alpha)f(\beta)\delta(\gamma)}{\sin\beta} \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)} \, d\mathbf{R} \, d\mathbf{a}$$

$$= \int_{SO(3)} \frac{\delta(\alpha)f(\beta)\delta(\gamma)}{\sin\beta} \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{0},\mathbf{R};p)} \, d\mathbf{R}$$

$$= \delta_{l,l'} \int_0^{2\pi} \int_\beta \int_0^{2\pi} \frac{\delta(\alpha)f(\beta)\delta(\gamma)}{\sin\beta} \times$$
$$e^{im\alpha}(-1)^{m'-m} P^l_{m,m'}(\cos\beta) e^{im'\gamma} \sin\beta \, d\alpha \, d\beta \, d\gamma$$

$$= \delta_{l,l'} \int_\beta f(\beta) P^l_{m',m}(\cos\beta) \, d\beta \qquad (20)$$

where $\delta_{i,j}$ is the Kronecker delta function and $P^l_{m,n}(x)$ is the generalized Legendre function.[51,62]

In particular, when $f(\beta_{\mathrm{h}})$ is a uniform distribution on the domain $[\beta_{\mathrm{h}1},\beta_{\mathrm{h}2}]$ with $\beta_{\mathrm{h}1} = \beta_{\mathrm{c}} - \Delta\beta$ and $\beta_{\mathrm{h}2} = \beta_{\mathrm{c}} + \Delta\beta$, one obtains

$$f(\beta_{\mathrm{h}}) = \frac{1}{2\Delta\beta} \qquad (21)$$

$$f_{j2}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \frac{\delta(\mathbf{a})\delta(\alpha)\delta(\gamma)}{2\Delta\beta \sin\beta} \qquad (22)$$

$$\hat{f}^{\mathrm{r}}_{j2l',m';l,m}(p) = \frac{\delta_{l,l'}}{2\Delta\beta} \int_{\beta_{\mathrm{c}}-\Delta\beta}^{\beta_{\mathrm{c}}+\Delta\beta} P^l_{m',m}(\cos\beta) \, d\beta \qquad (23)$$

3.3. The Ball Joint. If the angular fluctuation of the macromolecular chain at an internal structure is not constrained to a plane, the local structure can be modeled as a spatial revolute joint, the ball joint (Figure 6).
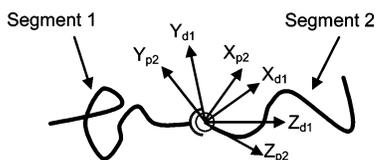
Figure 6. Diagram of a macromolecular chain with an internal ball joint.

With the rotation being parametrized using the *ZXZ* Euler angles, the rigid-body motion at the ball joint can be considered as consisting of a rotation α about the Z axis, followed by a rotation β about the current X axis, followed by a rotation γ about the current Z axis, with no translation in any direction. In fact, two Euler angles are enough to define the relative pose of the two segments connected by the ball joint (Figure 7). The
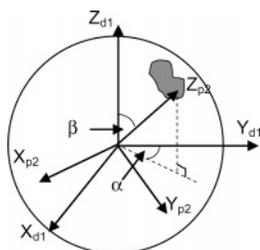
Figure 7. Workspace of a ball joint with a limited range of rotation.

ball joint model applies to the macromolecular chain which has not only bending fluctuations but also twisting fluctuation at

the local structure. The workspace of a ball joint with a limited range of rotation covers a certain area on the unit sphere (Figure 7).

In general, denoting the PDF of the bending angle β and the twisting angle α on the unit sphere as $f(\alpha,\beta)$, one has the corresponding PDF of rigid-body motion defined on SE(3) as

$$f_j(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \delta(\mathbf{a})f(\alpha,\beta)\delta(\gamma) \qquad (24)$$

Then the Fourier transform of $f_j(\mathbf{a},\mathbf{R})$ on SE(3) can be obtained as

$$\hat{f}^{\mathrm{r}}_{jl',m';l,m}(p) = \int_{SE(3)} f_j(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)} \, d\mathbf{R} \, d\mathbf{a}$$

$$= \int_{SE(3)} \delta(\mathbf{a})f(\alpha,\beta)\delta(\gamma) \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)} \, d\mathbf{R} \, d\mathbf{a}$$

$$= \int_{SO(3)} f(\alpha,\beta)\delta(\gamma) \overline{U^{\mathrm{r}}_{l,m;l',m'}(\mathbf{0},\mathbf{R};p)} \, d\mathbf{R}$$

$$= \delta_{l,l'} \int_0^{2\pi} \int_{\alpha,\beta} f(\alpha,\beta)\delta(\gamma) e^{im\alpha}(-1)^{m'-m} P^l_{m,m'}(\cos\beta) e^{im'\gamma} \times \\ \sin\beta \, d\alpha \, d\beta \, d\gamma$$

$$= \delta_{l,l'} \int_{\alpha,\beta} f(\alpha,\beta) e^{im\alpha} P^l_{m',m}(\cos\beta) \sin\beta \, d\alpha \, d\beta \qquad (25)$$

In particular, let us look at the case when $f(\alpha,\beta)$ is a uniform distribution on a circular region on the unit sphere (Figure 8).
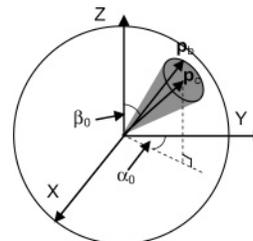
Figure 8. A ball joint with the uniform distribution on an arbitrary circular region on the unit sphere.

Here, we denote $\mathbf{p}_{\mathrm{c}}$ as the position vector of the center of the circular region on the unit sphere and denote $\mathbf{p}_{\mathrm{b}}$ as the position vector of an arbitrary point on the boundary of the circular region. α_0 and β_0 are the polar angle and the azimuthal angle of $\mathbf{p}_{\mathrm{c}}$. Clearly, $\mathbf{p}_{\mathrm{c}} = [\sin\beta_0 \sin\alpha_0, \sin\beta_0 \cos\alpha_0, \cos\beta_0]$. Since the circular region keeps a constant surface area when moving on the unit sphere, one can obtain from Figure 9 that inside the
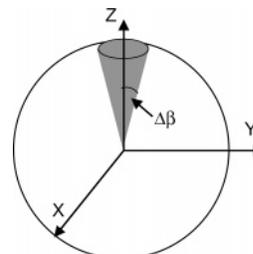
Figure 9. A ball joint with the uniform distribution on a circular region symmetric with respect to the Z axis on the unit sphere.

circular region

$$f(\alpha,\beta) = \frac{1}{2\pi(1 - \cos\Delta\beta)} \qquad (26)$$

and the corresponding PDF on SE(3) is

$$f_j(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \frac{\delta(\mathbf{a})\delta(\gamma)}{2\pi(1 - \cos\Delta\beta)} \qquad (27)$$

It is clear that when the circular region is symmetric with respect to the Z axis (Figure 9), the support of $f(\alpha,\beta)$ is defined as $\beta \in [0,\Delta\beta]$ and $\alpha \in [0,2\pi]$. However, when the circular region moves away from the Z axis, $\beta \in [\beta_0 - \Delta\beta, \beta_0 + \Delta\beta]$, but $\alpha \notin [0,2\pi]$. To find how α varies, one can obtain from the law of cosines

$$||\mathbf{p}_b - \mathbf{p}_c||^2 = 2 - 2\cos\Delta\beta \qquad (28)$$

Then letting $\mathbf{p}_b = [\sin\beta\sin\alpha, \sin\beta\cos\alpha, \cos\beta]$, one can obtain

$$\alpha = \alpha_0 \pm \cos^{-1}\left(\frac{\cos\Delta\beta - \cos\beta\cos\beta_0}{\sin\beta\sin\beta_0}\right) \qquad (29)$$

which gives the upper and lower bounds of α corresponding to any specific value of β.

Then the Fourier transform of $f_j(\mathbf{a},\mathbf{R})$ on SE(3) can be obtained as

$$\hat{f}^r_{jl',m';l,m}(p) = \frac{\delta_{l,l'}}{2\pi(1 - \cos\Delta\beta)} \times$$

$$\int_{\beta_0-\Delta\beta}^{\beta_0+\Delta\beta} \int_{\alpha_0-\cos^{-1}[(\cos\Delta\beta-\cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]}^{\alpha_0+\cos^{-1}[(\cos\Delta\beta-\cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]} e^{im\alpha} \times$$

$$P^l_{m',m}(\cos\beta)\sin\beta\,\mathrm{d}\alpha\,\mathrm{d}\beta$$

$$= \frac{\delta_{l,l'}}{2\pi m(1 - \cos\Delta\beta)i} \times$$

$$\int_{\beta_0-\Delta\beta}^{\beta_0+\Delta\beta} (e^{im[\alpha_0+\cos^{-1}(\cos\Delta\beta-\cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]} -$$

$$e^{im[\alpha_0-\cos^{-1}(\cos\Delta\beta-\cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]})P^l_{m',m}(\cos\beta)\sin\beta\,\mathrm{d}\beta \quad (30)$$

In fact, $f_j(\mathbf{a},\mathbf{R})$ for the uniform ball joint can also be obtained using the convolution on SE(3). We notice that a circular region arbitrarily located on the unit sphere results from moving a circular region with the same radius, which is symmetric to the Z axis (Figure 9), to the new position determined by α_0 and β_0 (Figure 8). Therefore, the PDF of rigid-body motion on SE(3) at this ball joint can be obtained as a convolution

$$f_j(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = (f_{j1}*f_{j2})(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) \qquad (31)$$

where

$$f_{j1}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \delta(\mathbf{a})\delta(\mathbf{R}(\alpha_0,\beta_0,0)^{-1}\mathbf{R}(\alpha,\beta,\gamma))$$

$$= \frac{\delta(\mathbf{a})\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma)}{\sin\beta} \qquad (32)$$

$$f_{j2}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma)) = \frac{\delta(\mathbf{a})\delta(\gamma)}{2\pi(1 - \cos\Delta\beta)} \qquad (33)$$

Then the Fourier transform of $f_j(\mathbf{a},\mathbf{R})$ on SE(3) can be obtained as

$$\hat{\mathbf{f}}^r_j(p) = \hat{\mathbf{f}}^r_{j2}(p)\hat{\mathbf{f}}^r_{j1}(p) \qquad (34)$$

where

$$\hat{f}^r_{j1l',m';l,m}(p) = \int_{\mathrm{SE}(3)} f_{j1}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma))\overline{U^r_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{a}$$

$$= \int_{\mathrm{SE}(3)} \frac{\delta(\mathbf{a})\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma)}{\sin\beta}\overline{U^r_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{a}$$

$$= \int_{\mathrm{SO}(3)} \frac{\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma)}{\sin\beta}\overline{U^r_{l,m;l',m'}(\mathbf{0},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}$$

$$= \delta_{l,l'}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} \frac{\delta(\alpha - \alpha_0)\delta(\beta - \beta_0)\delta(\gamma)}{\sin\beta} \times$$

$$e^{im\alpha}(-1)^{m'-m}P^l_{m,m'}(\cos\beta)e^{im'\gamma}\sin\beta\,\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\gamma$$

$$= \delta_{l,l'}e^{im\alpha_0}P^l_{m',m}(\cos\beta_0) \qquad (35)$$

$$\hat{f}^r_{2l',m';l,m}(p) = \int_{\mathrm{SE}(3)} f_{j2}(\mathbf{a},\mathbf{R}(\alpha,\beta,\gamma))\overline{U^r_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{a}$$

$$= \int_{\mathrm{SE}(3)} \frac{\delta(\mathbf{a})\delta(\gamma)}{2\pi(1 - \cos\Delta\beta)}\overline{U^r_{l,m;l',m'}(\mathbf{a},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{a}$$

$$= \int_{\mathrm{SO}(3)} \frac{\delta(\gamma)}{2\pi(1 - \cos\Delta\beta)}\overline{U^r_{l,m;l',m'}(\mathbf{0},\mathbf{R};p)}\,\mathrm{d}\mathbf{R}$$

$$= \delta_{l,l'}\int_0^{2\pi}\int_0^{\Delta\beta}\int_0^{2\pi} \frac{\delta(\gamma)}{2\pi(1 - \cos\Delta\beta)} \times$$

$$e^{im\alpha}(-1)^{m'-m}P^l_{m,m'}(\cos\beta)e^{im'\gamma}\sin\beta\,\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\gamma$$

$$= \frac{\delta_{l,l'}\delta_{m,0}}{1 - \cos\Delta\beta}\int_0^{\Delta\beta}P^l_{m',0}(\cos\beta)\sin\beta\,\mathrm{d}\beta \qquad (36)$$

4. PDF of the End-to-End Distance for the Rigid Chain with an Internal Joint

As mentioned before, the PDF of end-to-end distance is of particular interest and is a marginal density of the full PDF of end-to-end pose. Moreover, when the macromolecular chain is completely rigid, it is possible to derive the closed form solution for the PDF of the end-to-end distance. In this section, we take a look at the case of two rigid rods connected by a joint (Figure 10). Our method can easily be extended to the multijoint case
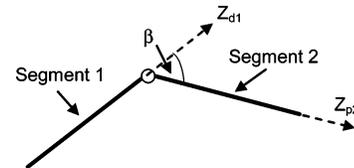
Figure 10. Diagram of two rigid rods connected by a joint.

by using geometric relationships.

In general, knowing the PDF of the bending angle, $p(\beta)$, one can derive the PDF of the end-to-end distance, $f(a)$, from the following relationship

$$f(a) = p(\beta)|\frac{\mathrm{d}\beta}{\mathrm{d}a}| \qquad (37)$$

where

$$\frac{\mathrm{d}\beta}{\mathrm{d}a} = -\frac{2a}{\sqrt{((L_1 + L_2)^2 - a^2)(a^2 - (L_1 - L_2)^2)}} \qquad (38)$$

Equation 38 can be derived from

$$a^2 = L_1^2 + L_2^2 - 2L_1L_2 \cos(\pi - \beta) \qquad (39)$$

which is obtained from the law of cosine.

When the joint is a hinge with a limited range of rotation, one can obtain from Figure 4

$$\cos \beta = -\sin \beta_0 \cos \gamma_0 \sin \beta_h + \cos \beta_0 \cos \beta \qquad (40)$$

from which one can further obtain

$$p(\beta) = f(\beta_h) \left| \frac{d\beta_h}{d\beta} \right| = \frac{k_1(\beta_h)}{k_2(\beta_h)} f(\beta_h) \qquad (41)$$

where

$$k_1(\beta_h) = \sqrt{1 - (-\sin \beta_0 \cos \gamma_0 \sin \beta_h + \cos \beta_0 \cos \beta_h)^2} \qquad (42)$$

$$k_2(\beta_h) = |\sin \beta_0 \cos \gamma_0 \cos \beta_h + \cos \beta_0 \sin \beta_h| \qquad (43)$$

Substituting (41) into (37), one obtains

$$f(a) = \frac{2a k_1(\beta_h(a))}{k_2(\beta_h(a))\sqrt{((L_1 + L_2)^2 - a^2)(a^2 - (L_1 - L_2)^2)}} f(\beta_h(a)) \qquad (44)$$

where

$$\beta_h(a) = \cos^{-1}$$

$$\left(\frac{k_3(a) \cos \beta_0 \pm \sqrt{k_3(a)^2 \cos^2 \beta_0 - (\cos^2 \beta_0 + \sin^2 \beta_0 \cos^2 \gamma_0)(k_3(a)^2 - \sin^2 \beta_0 \cos^2 \gamma_0)}}{\cos^2 \beta_0 + \sin^2 \beta_0 \cos^2 \gamma_0} \right) \qquad (45)$$

$$k_3(a) = \frac{a^2 - L_1^2 - L_2^2}{2L_1L_2} \qquad (46)$$

In particular, when $f(\beta_h)$ is a uniform distribution on the domain $[\beta_c - \Delta\beta, \beta_c + \Delta\beta]$, one obtains

$$f(a) = \frac{a k_1(\beta_h(a))}{\Delta\beta k_2(\beta_h(a))\sqrt{((L_1 + L_2)^2 - a^2)(a^2 - (L_1 - L_2)^2)}} \qquad (47)$$

When the joint is a ball joint with a limited range of rotation, and $f(\alpha,\beta)$ is a uniform distribution on a circular region of the unit sphere, one obtains from (33)

$$p(\beta) = \int_{\alpha_0 - \cos^{-1}[(\cos\Delta\beta - \cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]}^{\alpha_0 + \cos^{-1}[(\cos\Delta\beta - \cos\beta\cos\beta_0)/(\sin\beta\sin\beta_0)]} \frac{\sin\beta}{2\pi(1 - \cos\Delta\beta)} d\alpha$$

$$= \frac{\sin\beta \cos^{-1}\left(\dfrac{\cos\Delta\beta - \cos\beta\cos\beta_0}{\sin\beta\sin\beta_0}\right)}{\pi(1 - \cos\Delta\beta)} \qquad (48)$$

Substituting (48) into (37), one obtains

$$f(a) = \frac{a \cos^{-1}\left(\dfrac{2L_1L_2\cos\Delta\beta - (a^2 - L_1^2 - L_2^2)\cos\beta_0}{\sqrt{((L_1 + L_2)^2 - a^2)(a^2 - (L_1 - L_2)^2)}\sin\beta_0}\right)}{\pi L_1 L_2 (1 - \cos\Delta\beta)} \qquad (49)$$

5. Examples and Discussion

In this section, we will show how to implement our method derived in the last two sections by examples. We will present
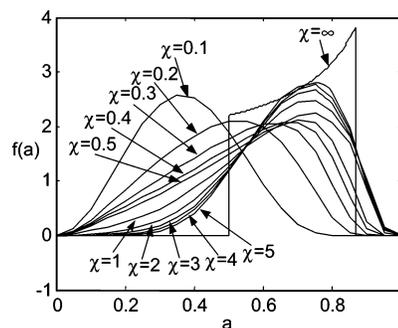
Figure 11. Variation of the PDF of the end-to-end distance of a macromolecular chain with a limited hinge in the middle with respect to the stiffness.

Table 1. Values of Computation Parameters Used in the Hinge Example

χ	l_b	p_b	χ	l_b	p_b
0.1	3	50	1.0	7	130
0.2	4	60	2.0	11	150
0.3	4	80	3.0	13	150
0.4	5	80	4.0	16	190
0.5	5	100	5.0	18	200

our results in the form of $f(a)$ instead of $f(\mathbf{a},\mathbf{R})$ for the convenience of display, though $f(\mathbf{a},\mathbf{R})$ contains more information than $f(a)$ and can be obtained using our methods.

The general formulation reviewed in section 2 applies to several stiffness models. Here we use the Kratky−Porod model in our examples. However, other models, such as the Yamakawa model and the Marko−Siggia model, also fit into our general formulation. For the Kratky−Porod wormlike chain model, the stiffness matrix, chirality vector, and constant term in (2) are defined as [28]

$$\mathbf{B} = \begin{pmatrix} \chi & 0 & 0 \\ 0 & \chi & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad c = 0 \qquad (50)$$

where χ is known as the stiffness parameter. The definitions of the stiffness matrix, chirality vector, and constants for the Yamakawa model and Marko−Siggia models can be found in ref 28. In our computation, the stiffness is measured in units of $k_B T$. Moreover, all the stiffness and length parameters are normalized by the total arc length of the macromolecular chain.

When implementing the Fourier transform for SE(3), by definition (as shown in eq 7), $r \in Z$, $p \in [0,\infty]$, and $\mathbf{B}^r$ is an infinite dimensional matrix. To do numerical computations, however, one must truncate r, l, l', and p at finite values and $\mathbf{B}^r$ at finite dimension. In particular, when $f(a)$ is of interest, we only need to consider $r = 0$, as suggested by (13).

As shown in Figure 11, we compute $f(a)$ of a macromolecular chain with a limited uniform hinge in the middle under different values of stiffness. In this example, we choose the range of rotation as $\beta_h \in [\pi/3, 2\pi/3]$ with $\alpha_0 = \beta_0 = \gamma_0 = 0$. The choices of the computation parameters are listed in Table 1, where $r \in [-l_b,l_b]$, $l \in [r,l_b]$, $l' \in [r,l_b]$, and $p \in [0,p_b]$.

As shown in Figure 12, we compute $f(a)$ for a macromolecular chain with a limited uniform ball joint in the middle under different values of stiffness. In this example, we choose the range of rotation as a circular region on the unit sphere centered at $\alpha_0 = \beta_0 = \pi/2$ with $\Delta\beta = \pi/6$. The choices of the computation parameters are listed in Table 2, where $r \in [-l_b,l_b]$, $l \in [r,l_b]$, $l' \in [r,l_b]$, and $p \in [0,p_b]$.

It is clear from our examples that, given the type of the joint and the range of rotation, the macromolecular chains with
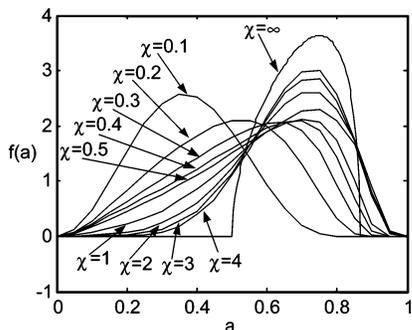
Figure 12. Variation of the PDF of the end-to-end distance of a macromolecular chain with a limited ball joint in the middle with respect to the stiffness.
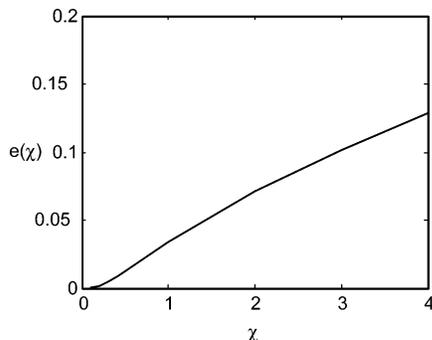
Figure 13. Root-mean-square difference between the PDF of the hinge example and that of the ball joint example. $e(\chi)$ denotes the root-mean-square difference between the PDF of the hinge case and that of the ball joint case. It is defined as $e(\chi) = \sqrt{\int_0^1 (f_{\text{hinge}}(a,\chi) - f_{\text{ball}}(a,\chi))^2 \, \mathrm{d}a}$. In the case of the rigid chain with a joint in the middle ($\chi = \infty$, which is not shown on the plot), $e(\chi = \infty) \approx 0.55$.

Table 2. Values of Computation Parameters Used in the Ball Joint Example

χ	l_b	p_b	χ	l_b	p_b
0.1	3	50	1.0	7	130
0.2	4	70	2.0	11	180
0.3	4	80	3.0	13	170
0.4	5	80	4.0	16	190
0.5	5	100			

different values of stiffness will have different PDFs of end-to-end distance. This means that one can determine the stiffness of a macromolecular chain with a known internal joint from the experimentally measured PDF. Meanwhile, knowing the stiffness of a macromolecular chain with an internal joint, it is also possible to distinguish between hinge-like and ball-like joints from experimentally measured PDFs. From our computation, we notice that the difference between the PDF of a macromolecular chain with a hinge and that of a chain with a ball joint is small when the stiffness is low and increases as the stiffness gets higher (Figure 13). Therefore, it would be easier to identify the type of the joint from experimentally obtained end-to-end distance distributions for a stiffer macromolecular chain than that for a more flexible chain.

A variety of experimental methods can be used to directly measure the end-to-end distance of a macromolecular chain and generate the PDF of the end-to-end distance. Single molecule FRET is the most widely adopted way to study the conformational distribution and dynamics of individual macromolecules, which efficiently measures the intramolecular distances from 3 to 10 nm.[63−66] At the low end, fluorescence quenching by TEMPO measures the sub-3 nm distances with a resolution of 0.5 nm,[63] and a single-molecule optical switch based on Cy5 and Cy3 can measure the intramolecular distances as short as 1

nm.[64] At the high end, a molecular ruler based on plasmon coupling of single gold and silver particles can measure intermolecular distances continuously up to 70 nm.[65] Meanwhile, single-molecule high-resolution colocalization (SHREC) of fluorescent probes can measure the intramolecular distances from 10 to 200 nm over time and has been used to generate the distribution of the end-to-end distance of DNA fragments.[66] Besides these distance-measuring methods, the distribution of the end-to-end distance of a free macromolecular chain can also be retrieve from the force−extension experiments.[67,68] By holding different end-to-end distances and measuring the average holding forces, one can establish a force−extension relation and retrieve the distribution of the end-to-end distance based on $\langle F \rangle = -[\partial \log(f(a))/\partial a]$, where $\langle F \rangle$ denotes the average holding force.[67] The holding experiment can be implemented by laser trap and atomic force microscopy.[68]

In addition to these direct measurement techniques, indirect methods that measure moments of the end-to-end distance distribution also exist. The following section discusses how the radius of gyration for semiflexible chains with internal defects (which can be computed from our model) relates to light scattering experiments.

6. Detecting Joints in Semiflexible Chains Based on Radius of Gyration

The proposed method also provides a potential tool for detecting the existence of joint-like internal structures in single macromolecular chains based on the radius of gyration measured from the laser light scattering experiment, which achieves a precision of 2−4%.[69−71]

In principle, the radius of gyration of a chain with internal joints will be different from that of an intrinsically straight chain. For a single macromolecular chain, one can measure the radius of gyration from the light scattering experiment. One can also compute the radius of gyration of a comparable intrinsically straight chain with same stiffness. By comparing the measured radius of gyration with the calculated one, one can determine whether there is a joint-like internal structure on the chain.

To study the feasibility of detecting the existence of internal joints on single macromolecular chains based on the radius of gyration, we compute and compare the radius of gyration of both intrinsically straight chain and jointed chain based on the PDFs generated by the proposed method. Here we use the macromolecular chain with a limited hinge in the middle as our example of the jointed chain, with $\alpha_0 = \beta_0 = \gamma_0 = 0$ and $\beta_h \in [\beta - \Delta\beta, \beta + \Delta\beta]$. β is the central bending angle at the hinge, and $\Delta\beta$ is the half range angle. Denoting the radius of gyration as R_G, we will study the impact of both β and $\Delta\beta$ on R_G.

The radius of gyration of a single macromolecular chain can be computed from the PDFs of its point-to-point distances. Discretizing a chain into n segments, one can calculate the radius of gyration as[72]

$$R_G = \frac{\sqrt{\sum_{j=1}^{n}\sum_{i=1}^{n}\langle r_{ij}^2 \rangle}}{\sqrt{2n}} \tag{51}$$

where

$$\langle r_{ij}^2 \rangle = \int_0^{l_{ij}} r_{ij}^2 f_{ij}(r_{ij}) \, \mathrm{d}r_{ij} \tag{52}$$
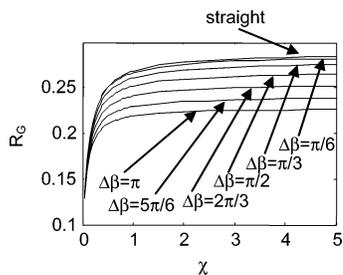
Figure 14. Comparison in R_G between the intrinsically straight and hinged macromolecular chains with $\beta = 0$ and different $\Delta\beta$.
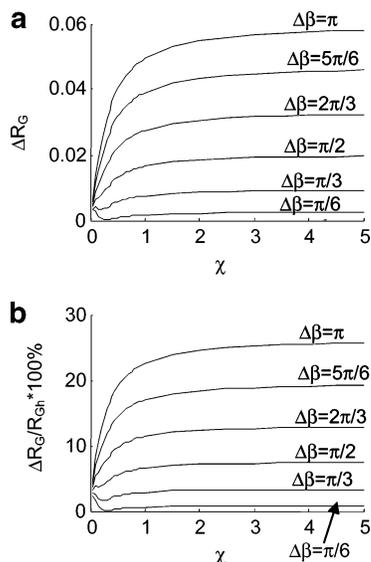
Figure 15. Difference in R_G between the intrinsically straight chain and the hinged chains with $\beta = 0$ and different $\Delta\beta$. Here, $\Delta R_G = R_{Gs} - R_{Gh}$, where R_{Gs} denotes the radius of gyration of the intrinsically straight chain and R_{Gh} denotes the radius of gyration of the hinged chain. Part a presents the difference in absolute value, and part b presents the difference in percentage.

Here, r_{ij} denotes the distance between two points i and j on the chain, l_{ij} denotes the arc length between i and j, and $f_{ij}(r_{ij})$ denotes the PDF of the distance between i and j which can be generated using the method presented in previous sections. Equation 51 applies to both intrinsically straight and jointed chains.

In Figure 14 and Figure 15, we study the impact of $\Delta\beta$ on R_G, with $\beta = 0$. In this case, the jointed chain fluctuates around a straight conformation. A DNA with a symmetric internal loop belongs to this category.

By changing χ, we obtain a class of $R_G - \chi$ curves for the intrinsically straight chain and the jointed chains with different $\Delta\beta$. Figure 14 shows that in general: (1) R_G of a jointed chain is smaller than that of an intrinsically straight chain with same stiffness because the joint causes the chain to fold back. (2) R_G of a flexible chain is small because the conformation of such a chain is usually an entangled coil. (3) R_G of the chain increases as χ increases because the conformation of a stiffer chain tends to stretch out. Moreover, a stiffer chain behaves more like a rigid chain. Therefore, R_G of a jointed chain converges to that of the rigid chain as χ increases. (4) R_G of a jointed chain with a larger $\Delta\beta$ tends to be smaller because such a chain is more likely to fold back.

The difference in R_G between the intrinsically straight chain and the jointed chains with $\beta = 0$ and different $\Delta\beta$ is presented in Figure 15.

Figure 15 shows that ΔR_G increases as $\Delta\beta$ increases. This means that an internal joint with a larger fluctuation range is
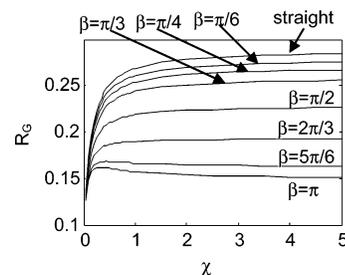
Figure 16. Comparison in R_G between the intrinsically straight and hinged macromolecular chains with $\Delta\beta = \pi/6$ and different β.
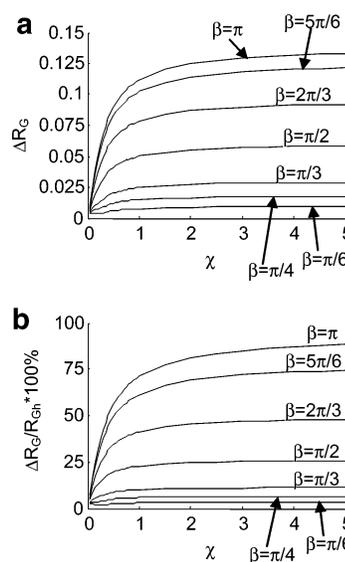
Figure 17. Difference in R_G between the intrinsically straight chain and the hinged chains with $\Delta\beta = \pi/6$ and different β. Part a presents the difference in absolute value, and part b presents the difference in percentage.

Table 3. Percentage Difference in R_G between the Straight Chain and the Jointed Chains with Zero β ($\chi = 1$)

$\Delta\beta$	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
$\Delta R_G/R_{Gh} \times 100\%$	0.64%	2.88%	6.55%	11.40%	16.98%	22.52%

more distinguishable from the intrinsically straight chain. Table 3 presents the percentage difference of R_G, when $\chi = 1$, between the intrinsically straight chain and the jointed chains with zero β. The table shows that a jointed chain with $\Delta\beta \geq \pi/2$ is highly detectable on the basis of experimentally measured R_G, according to the reported measurement precision.[69−71]

In Figure 16 and Figure 17, we study the impact of β on R_G, with $\Delta\beta = \pi/6$. In this case, the jointed chain fluctuates around a bent conformation. A DNA with a bulge or a gap belongs to this category.

By changing χ, we obtain a class of $R_G - \chi$ curves for the intrinsically straight chain and the jointed chains with different β. Figure 16 shows a similar relation between R_G and χ as Figure 14 does. Moreover, R_G of a jointed chain with a larger β tends to be smaller because β forces the chain to fold back.

The difference in R_G between the intrinsically straight chain and the jointed chains with $\Delta\beta = \pi/6$ and different β is presented in Figure 17.

Figure 17 shows that ΔR_G increases as β increases. This means that an internal joint with a larger bending is more distinguishable from the intrinsically straight chain. Table 4 presents the percentage difference of R_G, when $\chi = 1$, between the intrinsically straight chain and the jointed chains with $\Delta\beta = \pi/6$. The table shows that a jointed chain with $\beta \geq \pi/4$
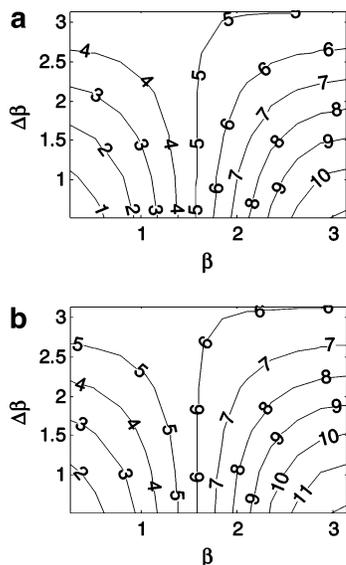
Figure 18. Difference in R_G between the intrinsically straight chain and the hinged chains with different β and $\Delta\beta$. Part a shows the contour plot of the percentage difference when $\chi = 1$, and part b shows the contour plot of the percentage difference when $\chi = 2$.

Table 4. Percentage Difference in R_G between the Straight Chain and the Jointed Chains with $\Delta\beta = \pi/6$ ($\chi = 1$)

β	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π
$\Delta R_G/R_{Gh} \times 100\%$	2.98%	5.93%	10.17%	22.89%	41.28%	61.45%	71.13%

is highly detectable on the basis of experimentally measured R_G, according to the reported measurement precision.[69−71]

Figure 18 shows the combined effect of both β and $\Delta\beta$ on R_G. In general, a jointed chain with larger β and/or $\Delta\beta$ is more distinguishable from the intrinsically straight chain.

7. Conclusion

In this paper, we proposed an effective computational method to estimate the conformational statistics of semiflexible macromolecular chains with internal structures which can be represented as joints with limited capability of rotation, such as DNAs containing symmetric loops, bulges, and nicks/gaps. Our method can compute the PDF of the end-to-end pose of a macromolecular chain with any stiffness and chirality in its minimum-energy states. Other marginal PDFs and statistical quantities can be further calculated from this full PDF. The proposed method extends our previous general formulation of classical polymer theory and can be used with different stiffness models, including the Krathy−Porod model, the Yamakawa model, and the Marko−Siggia model. This capability allows one to study the entropic effects of internal angular fluctuations on semiflexible macromolecular chains with internal joints. Our method also provides a potential tool to detect the existence of internal joints. With the complete spectrum of computational results obtained using this method, it is possible in principle to retrieve the stiffness, joint type, and joint location of a macromolecular chain with an internal defect from experimentally measured PDFs. The current method ignores the interactions among the segments of the same macromolecular chain. This will become a topic of our future work.

References and Notes

(1) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 7664−7668.

(2) Van der Vliet, P. C.; Verrijzer, C. P. *Bioessays* **1993**, *15*, 25−32.

(3) Perez-Martin, J.; Rojo, F.; De Lorenzo, V. *Microbiol. Rev.* **1994**, *58*, 268−290.

(4) Dickerson, R. E.; Goodsell, D.; Kopka, M. L. *J. Mol. Biol.* **1996**, *256*, 108−125.

(5) Hansma, H. G.; Browne, K. A.; Bezanilla, M.; Bruice, T. C. *Biochemistry* **1994**, *33*, 8436−8441.

(6) Griffith, J.; Bleyman, M.; Rauch, C. A.; Kitchin, P. A.; Englund, P. T. *Cell* **1986**, *46*, 717−724.

(7) Crothers, D. M.; Haran, T. E.; Nadeau, J. G. *J. Biol. Chem.* **1990**, *265*, 7093−7096.

(8) Benoff, B.; Yang, H.; Lawson, C. L.; Parkinson, G.; Liu, J.; Blatter, E.; Ebright, Y. W.; Berman, H. M.; Ebright, R. H. *Science* **2002**, *297*, 1562−1566.

(9) Guo, H.; Tullius, T. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3743−3747.

(10) Roll, C.; Ketterle, C.; Faibis, V.; Fazakerley, G. V.; Boulard, Y. *Biochemistry* **1998**, *37*, 4059−4070.

(11) Kahn, J. D.; Yun, E.; Crothers, D. M. *Nature (London)* **1994**, *368*, 163−166.

(12) Zacharias, M.; Hagerman, P. J. *J. Mol. Biol.* **1996**, *257*, 276−289.

(13) Gohlke, C.; Murchie, A. I. H.; Lilley, D. M. J.; Clegg, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11660−11664.

(14) Luebke, K. J.; Landry, S. M.; Tinoco, I. *Biochemistry* **1997**, *36*, 10246−10255.

(15) Dornberger, U.; Hillisch, A.; Gollmick, F. A.; Fritzsche, H.; Diekmann, S. *Biochemistry* **1999**, *38*, 12860−12868.

(16) Zacharias, M.; Hagerman, P. J. *J. Mol. Biol.* **1995**, *247*, 486-500.

(17) Luebke, K. L.; Tinoco, I. *Biochemistry* **1996**, *35*, 11677−11684.

(18) Gollmick, F. A.; Lorenz, M.; Dornberger, U.; Langen, J. von; Diekmann, S.; Fritzsche, H. *Nucleic Acids Res.* **2002**, *30*, 2669−2677.

(19) Feig, M.; Zacharias, M.; Pettitt, B. M. *Biophys. J.* **2001**, *81*, 352−370.

(20) Zacharias, M.; Sklenar, H. *Biophys. J.* **2000**, *78*, 2528−2542.

(21) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Interscience Publishers: New York, 1969.

(22) Kratky, O.; Porod, G. *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 1106−1122.

(23) Daniels, H. E. *Proc. R. Soc. Edinburgh* **1952**, *A63*, 290−311.

(24) Hermans, J. J.; Ullman, R. *Physica* **1952**, *18*, 951−971.

(25) Yamakawa, H. *Helical Wormlike Chains in Polymer Solutions*; Springer: New York, 1997.

(26) Marko, J. F.; Siggia, E. D. *Macromolecules* **1994**, *27*, 981−988.

(27) Zhou, H.; Ou-Yang, Z. *Phys. Rev. E* **1998**, *58*, 4816−4819.

(28) Chirikjian, G. S.; Wang, Y. *Phys. Rev. E* **2000**, *62*, 880−892.

(29) Fixman, M.; Alben, R. *J. Chem. Phys.* **1973**, *58*, 1553−1558.

(30) Maroun, R. C.; Olson, W. K. *Biopolymers* **1988**, *27*, 561−584.

(31) Hagerman, P. J. *Biopolymers* **1985**, *24*, 1881−1897.

(32) Lax, M.; Barrett, A. J.; Domb, C. *J. Phys. A: Math. Gen.* **1978**, *11*, 361−374.

(33) Croxton, C. A. *J. Phys. A: Math. Gen.* **1979**, *12*, 2475−2485.

(34) Nitta, K. H. *J. Chem. Phys.* **1994**, *101*, 4222−4228.

(35) Aguileragranja, F.; Kkuchi, R. *Physica A* **1992**, *182*, 331−345.

(36) Liao, Q.; Wu, D. C. *Acta Polym. Sin.* **2000**, *4*, 420−425.

(37) Wu, D. C.; Du, P.; Kang, J. *Sci. China Ser. B: Chem.* **1997**, *40*, 137−145.

(38) Liverpool, T. B.; Edwards, S. F. *J. Chem. Phys.* **1995**, *103*, 6716−6719.

(39) Walasek, J. *J. Polym. Sci., Part B: Polym. Phys.* **1988**, *26*, 1907−1922.

(40) Burlatskii, S. F.; Oshanin, G. S. *Theor. Math. Phys.* **1998**, *75*, 659−663.

(41) Coriell, S. R.; Jackson, J. L. *J. Math. Phys.* **1967**, *8*, 1276.

(42) Kosmas, M. K. *J. Phys. A: Math. Gen.* **1983**, *16*, L381-L384.

(43) Wilhelm, J.; Frey, E. *Phys. Rev. Lett.* **1996**, *77*, 2581−2584.

(44) Lagowski, J. B.; Noolandi, J.; Nickel, B. *J. Chem. Phys.* **1991**, *95*, 1266−1269.

(45) Ronca, G.; Yoon, D. Y. *J. Chem. Phys.* **1984**, *80*, 930−935.

(46) Mondescu, R. P.; Muthukumar, M. *Phys. Rev. E* **1998**, *57*, 4411−4419.

(47) Mondescu, R. P.; Muthukumar, M. *J. Chem. Phys.* **1999**, *110*, 12240−12249.

(48) Hoffman, G. G. *J. Phys. Chem. B* **1999**, *103*, 7167−7174.

(49) Papadopoulos, G. J.; Thomchick, J. *J. Phys. A: Math. Gen.* **1977**, *10*, 1115−1121.

(50) Dua, A.; Cherayil, B. J. *J. Chem. Phys.* **1998**, *109*, 7011−7016.

(51) Chirikjian, G. S.; Kyatkin, A. B. *Engineering Applications of Noncommutative Harmonic Analysis*; CRC Press: Boca Raton, FL, 2000.

(52) Chirikjian, G. S.; Kyatkin, A. B. *J. Fourier Anal. Appl.* **2000**, *6*, 583−606.

(53) Chirikjian, G. S. *Comput. Theor. Polym. Sci.* **2001**, *11*, 143−153.

(54) Hagerman, P. J. *Annu. Rev. Biochem.* **1990**, *29*, 755−781.

(55) Hagerman, P. J. *Biochim. Biophys. Acta* **1992**, *1131*, 125−132.

(56) Olson, W. K.; Zhurkin, V. B. *Biol. Struct. Dyn., 2, Proc. Ninth Conv. Discipline Biomol. Stereodyn.* **1996**, 341−370.

(57) Olson, W. K.; Marky, N. L.; Jernigan, R. L.; Zhurkin, V. B. *J. Mol. Biol.* **1993**, *232*, 530−551.

(58) Rivetti, C.; Walker, C.; Bustamante, C. *J. Mol. Biol.* **1998**, *280*, 41−59.

(59) Zhou, Y.; Chirikjian, G. S. *J. Chem. Phys.* **2003**, *119*, 4962−4970.

(60) Kyatkin, A. B.; Chirikjian, G. S. *Appl. Comput. Harmonic Anal.* **2000**, *9*, 220−241.

(61) Miller, W. *Commun. Pure Appl. Math.* **1964**, *17*, 527−540.

(62) Vilenkin, N. J.; Klimyk, A. U. *Representation of Lie Groups and Special Functions*; Kluwer Academic Publishers: Dordrecht, Holland, 1991.

(63) Zhu, P.; Clamme, J.; Deniz, A. A. *Biophys. J.: Biophys. Lett.* **2005**, L37−L39.

(64) Bates, M.; Blosser, T. R.; Zhuang, X. *Phys. Rev. Lett.* **2005**, *94*, 108101.1−4.

(65) Sonnichsen, C.; Reinhard, B. M.; Liphardt, J.; Alivisatos, A. P. *Nat. Biotechnol.* **2005**, *23*, 741−745.

(66) Churchman, L. S.; Okten, Z.; Rock, R. S.; Dawson, J. F.; Spudich, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1419−1423.

(67) Ranjith, P.; Kumar, P. B. S.; Menon, G. *Phys. Rev. Lett.* **2005**, *94*, 138102.1−4.

(68) Keller, D.; Swigon, D.; Bustamante, C. *Biophys. J.* **2003**, *84*, 733−738.

(69) Lai, E.; Zanten, J. H. V. *Biophys. J.* **2001**, *80*, 864−873.

(70) Zhou, H.; Miller, A. W.; Sosic, Z.; Buchholz, B.; Barron, A. E.; Kotler, L.; Karger, B. L. *Anal. Chem.* **2000**, *72*, 1045−1052.

(71) Jeng, L.; Balke, S. T.; Mourey, T. H.; Wheeler, L.; Romeo, P. *J. Appl. Polym. Sci.* **1993**, *49*, 1359−1374.

(72) Johnson, C. S.; Gabriel, D. A. *Laser Light Scattering*; Dover: Mineola, NY, 1994.