Available online at www.sciencedirect.com
SCIENCE DIRECT
Polymer 46 (2005) 11904–11917
polymer
www.elsevier.com/locate/polymer

A unified approach to conformational statistics of classical polymer and polypeptide models

Jin Seob Kim, G regory S. Chirikjian*

Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

Received 20 June 2005; received in revised form 1 September 2005; accepted 2 September 2005
Available online 10 October 2005

Abstract

We present a unified method to generate conformational statistics, which can be applied to any of the classical discrete-chain polymer models. The proposed method employs the concepts of Fourier transform and generalized convolution for the group of rigid-body motions in order to obtain probability density functions of chain end-to-end distance. In this paper, we demonstrate the proposed method with three different cases: the freely rotating model, independent torsion-angle energy model, and interdependent pair-wise energy model (the last two are also well-known as the rotational isomeric state model). As for numerical examples, for simplicity, we assume homogeneous polymer chains. For the freely rotating model, we verify the proposed method by comparing with well-known closed-form results for mean-squared end-to-end distance. In the interdependent pair-wise energy case, we take polypeptide chains such as polyalanine and polyvaline as examples.

Keywords: Conformational statistics; Rigid-body motion group; Non-commutative harmonic analysis

1. Introduction

Conformational studies on polymer chains have been applied to a number of areas, such as polymer science and biophysics, including protein folding [1,2]. An important quantity in conformational studies is the end-to-end distance distribution, or probability density function (PDF) of end-to-end distance.

From the ensemble average of end-to-end distance, or its distribution, many observable quantities can be predicted, including the radius of gyration, the viscosity of dilute polymer solutions, local concentration, scattering of radiation, etc. [3]. Another interesting quantity that depends on the end-to-end distance distribution is the reaction-limit rate, which is one of the crucial factors in loop formation in polypeptide chains [4]. It has also been shown that the end-to-end distance distribution is important in obtaining force-extension relations and elastic properties of semifexible polymers [5–7]. Well-known works by the Mark group have shown that elastic properties of polymer networks with and without filler particles can be derived from the end-to-end distance distribution of a polymer chain [8–12]. In order to determine this probability distribution, one needs a theoretical model for a polymer chain. Several phantom models of polymers have been developed to analyze their statistical behavior. These can be categorized into two main groups.

The first group consists of continuous chain models, which need mechanical properties such as bending/twist stiffness and persistence length, etc ([1,13,14]). Representative examples of this group are the Kratky–Porod model or worm-like chain (WLC) model, Yamakawa helical wormlike chain model, and Marko–Siggia model [14–17]. For example, attempts have been made to generate end-to-end distance distributions with the WLC model [18–20]. These works employ mathematical techniques from quantum physics to compute end-to-end distance distributions theoretically, and compare those with the results from Monte Carlo simulation. Also, Zhou has shown that loops in proteins can be modeled using the WLC model [21]. In another work, end-to-end distance distribution functions for polyelectrolyte chains have been derived using a charged WLC model, together with excluded volume effects [5]. Recently, one unified methodology has been reported by which one can describe probability density functions with respect to all continuous phantom models with quadratic energy function [22]. Also Zhou and Chirikjian have succeeded in generating probability density functions for bent semifexible polymers with this general approach [23].

* Corresponding author. Tel.: +1 410 516 6451; fax: +1 410 516 7254.
E-mail addresses: jkim115@jhu.edu (J.S. Kim), gregc@jhu.edu (G. S. Chirikjian).

0032-3861/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2005.09.012

The second group consists of discrete chain models ([1,13]). For example, the freely-jointed model, freely-rotating model, independent energy model, and interdependent pair-wise energy model (the last two are also called the rotational isomeric state (RIS) model) fall into this category. Among these models, the RIS model is treated as the most general one [3]. The end-to-end distance distribution for the freely-rotating model is known analytically [24]. As for the RIS model, several approximation methods have been developed to generate the end-to-end distance distribution [25–28]. These works use least squares inference together with the characteristic function (which is the classical Fourier transform of the spherically symmetric part of the end-to-end distance distribution) to obtain even moments of the end-to-end distance distribution. These works have been applied to symmetric and asymmetric chains [29]. Other widely used methods are numerical techniques considering full atomic detail in both polymer chains and solvents. One such technique is molecular dynamics (MD) simulation [30,31]. This method, however, has one big drawback that the computational cost is too high. Monte Carlo (MC) simulation has been preferred instead to obtain end-to-end distance distribution (e.g. [32]). Another method incorporates the RIS model (especially the interdependent pair-wise energy model) into MC [3]. In this work, one can generate many possible conformations of a polymer chain within the framework of the RIS model. Also, the largest eigenvalue method can be incorporated into this work to obtain the end-to-end distance distribution functions of long polymer chains ([33] and the references therein). A recent attempt combines MD and MC together such that MD is applied first to obtain an energy distribution with respect to torsional angle space and then MC and RIS models are applied for computation of mean end-to-end distance [34]. However, in general, MC also has some drawbacks, one of which is that it is not good for describing the 'tails' of some probability density functions [35]. A very general theoretical methodology has been published using the generalized convolution on the group of rigid-body motions [35].

In this paper, we present a unified method to analytically and exactly generate the probability density function (PDF) of any of the classical discrete-chain polymer models. The presented method is based on the generalized convolution concept [35], and combines it with ideas from non-commutative harmonic analysis [13]. Our proposed method can generate the full 6D PDF of relative end-to-end position and orientation of a polymer chain. Then the PDF of the end-to-end distance can be viewed as a 1D marginal PDF. Hence, our method can be applied to both symmetric and asymmetric chains. It will also be shown that, unlike other methods, the proposed method can be applied to any type of pair-wise potential energy in the RIS model. For more specific demonstrations, we apply the method to the case of the freely rotating model, independent energy model, and interdependent pair-wise energy model. For this reason, we describe the basic mathematics required to understand the formulation in the first section. In subsequent sections, we formulate the proposed method according to three different models, and demonstrate the efficient implementation of the method and its application to polypeptide chains. Finally, numerical examples follow thereafter.

2. Notation and terminology

In this section, we present the basic mathematics, which will be used in our entire paper.

2.1. Fourier transform for SE(3)

In this section, we give a brief review of the Fourier transform for the rigid-body motion group. For detailed definitions and explanations, see Ref. [13].

The special Euclidean group, $SE(3)$, is defined as a set which contains translations and rotations in 3D Euclidean space. Let g be an element of $SE(3)$, then $g=(\mathbf{r},R)$ can be written in matrix form as

$$g = \begin{pmatrix} R & \mathbf{r} \\ 0^{\mathrm{T}} & 1 \end{pmatrix}$$

Multiplication of any two such matrices results in a matrix of the same form. $SE(3)$ is a Lie group under matrix multiplication. Here, $R\in SO(3)$ is a rotation in three-dimensional space, and is parametrized using ZXZ Euler angles as

$$R(\alpha,\beta,\gamma) = \mathrm{ROT}[\mathbf{e}_3,\alpha]\mathrm{ROT}[\mathbf{e}_1,\beta]\mathrm{ROT}[\mathbf{e}_3,\gamma]$$

where $\mathrm{ROT}[\mathbf{e}_i,\varphi]$ denotes the rotation matrix describing the rotation by φ about the axis parallel to the unit vector $\mathbf{e}_i$. $\mathbf{r}\in\mathbb{R}^3$ represents translation in three-dimensional space, and is also parametrized by means of spherical coordinates as

$$r = \begin{pmatrix} r\cos\phi\sin\theta \\ r\sin\phi\sin\theta \\ r\cos\theta \end{pmatrix}$$

Matrix elements of the irreducible unitary representations of $SE(3)$, $U^s_{l',m';l,m}(\mathbf{r},R;p)$, are defined as [13,36]

$$U^s_{l',m';l,m}(\mathbf{r},R;p) = \sum_{j=-l}^{l}[l',m'|p,s|l,j](\mathbf{r})\tilde{U}^l_{j,m}(R) \tag{1}$$

In the above definition, the rotational part, $\tilde{U}^l_{m,n}$ are matrix elements of the irreducible unitary representations for $SO(3)$, which are defined as [37,38]

$$\tilde{U}^l_{m,n}(R(\alpha,\beta,\gamma)) = (-1)^{n-m}\mathrm{e}^{-\mathrm{i}(m\alpha+n\gamma)}P^l_{m,n}(\cos\beta) \tag{2}$$

where α, β, and γ are ZXZ Euler angles and $P^l_{m,n}(\cos\beta)$ is a generalized associated Legendre function, which can be calculated by the following integral

$$P^l_{m,n}(\cos\beta) = \frac{\mathrm{i}^{n-m}}{2\pi}\sqrt{\frac{(l-m)!(l+m)!}{(l-n)!(l+n)!}} \times \int_0^{2\pi}\left(\cos\frac{\beta}{2}\mathrm{e}^{\mathrm{i}\phi/2} + \mathrm{i}\sin\frac{\beta}{2}\mathrm{e}^{-\mathrm{i}\phi/2}\right)^{l-n} \times \left(\cos\frac{\beta}{2}\mathrm{e}^{-\mathrm{i}\phi/2} + \mathrm{i}\sin\frac{\beta}{2}\mathrm{e}^{\mathrm{i}\phi/2}\right)^{l+n}\mathrm{e}^{\mathrm{i}m\phi}\mathrm{d}\phi \quad (3)$$

or one can obtain by the following relation using the Jacobi polynomials

$$P^l_{m,n}(\cos\beta) = \left[\frac{(l-m)!(l+m)!}{(l-n)!(l+n)!}\right]^{1/2} \times \sin^{m-n}\frac{\beta}{2}\cos^{m+n}\frac{\beta}{2}P^{(m-n,m+n)}_{l-m}(\cos\beta) \quad (4)$$

The translational part in Eq. (1) is expressed as

$$[l',m'|p,s|l,m](\mathbf{r}) = \int_{\Theta=0}^{\pi}\int_{\Phi=0}^{2\pi} Q^{l'}_{s,m'}(\cos\Theta)\mathrm{e}^{-\mathrm{i}(m'+s)\Phi}\mathrm{e}^{-\mathrm{i}\mathbf{p}\cdot\mathbf{r}} \times Q^{l}_{s,m}(\cos\Theta)\mathrm{e}^{-\mathrm{i}(m+s)\Phi}\sin\Theta\mathrm{d}\Theta\mathrm{d}\Phi \quad (5)$$

where

$$\mathbf{p} = \begin{pmatrix} p\cos\Phi\sin\Theta \\ p\sin\Phi\sin\Theta \\ p\cos\Theta \end{pmatrix}$$

and

$$Q^l_{-s,m}(\cos\Theta) = (-1)^{l-s}\sqrt{\frac{2l+1}{4\pi}}P^l_{s,m}(\cos\Theta)$$

Here, we use $\mathrm{i}=\sqrt{-1}$ as an imaginary unit to distinguish it from the index i. One can also use the following series form to calculate the translational part of the matrix elements of IURs for $SE(3)$:

$$[l',m'|p,s|l,m](\mathbf{r}) = (4\pi)^{1/2}\sum_{k=|l'-l|}^{l'+l}\mathrm{i}^k\sqrt{\frac{(2l'+1)(2k+1)}{(2l+1)}}j_k(pr) \times C(k,0;l',s|l,s) \times C(k,m-m';l',m'|l,m)Y_k^{m-m'}(\theta,\phi) \quad (6)$$

where $C(k,0;l',\ s|l,\ s)$, $C(k,m-m';\ l',m'|l,m)$ are Clebsch–Gordan coefficients, $Y_k^{m-m'}(\theta,\phi)$ are spherical harmonic functions, and $j_k(pr)$ is the kth spherical Bessel function. According to [37], Clebsch–Gordan coefficients are defined as

where

$$\Delta(l_1,l_2,l) = \sqrt{\frac{(l_1+l_2-l)!(l_1-l_2+l)!(l_2-l_1+l)!}{(l_1+l_2+l+1)!}}$$

Finally, we are at the stage of defining the Fourier transform for $SE(3)$. Based on the above formulae, the matrix elements of the Fourier transform of a function $F(g)$, wherein $g=(\mathbf{r},R)\in SE(3)$, is obtained by the following relation

$$\hat{F}^s_{l',m';l,m}(p) = \int_{SE(3)} F(g)\overline{U^s_{l,m;l',m'}(g;p)}\mathrm{d}g \quad (8)$$

where $\mathrm{d}g=\mathrm{d}R\mathrm{d}\mathbf{r}$ with $\mathrm{d}R=(1/8\pi^2)\sin\beta\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\gamma$ and $\mathrm{d}\mathbf{r}=r^2\sin\theta\mathrm{d}\mathbf{r}\mathrm{d}\theta\mathrm{d}\phi$.

The inverse Fourier transform is defined as

$$F(g) = \frac{1}{2\pi^2}\sum_{s=-\infty}^{\infty}\int_0^{\infty}p^2\mathrm{d}p\ \mathrm{trace}(\hat{F}^sU^s) \quad (9)$$

or in component form as

$$F(g) = \frac{1}{2\pi^2}\sum_{s=-\infty}^{\infty}\sum_{l'=|s|}^{\infty}\sum_{l=|s|}^{\infty}\sum_{m'=l'}^{l'}\sum_{m=l}^{l}\int_0^{\infty}p^2\mathrm{d}p\,\hat{F}^s_{l,m;l',m'}(p)U^s_{l',m';l,m}(g;p) \quad (10)$$

The convolution of two functions on rigid-body motion group $F_1(g)$, $F_2(g)$ is defined as

$$(F_1 * F_2)(g) = \int_{SE(3)} F_1(h)F_2(h^{-1}\circ g)\mathrm{d}h \quad (11)$$

where $h,\ g\in SE(3)$. The geometric meaning of this convolution is that the second function is swept and weighted by the first. For example, if the full distribution of positions and orientations of two adjacent segments of a polymer chain are known (Fig. 1), then the concatenation of the segments yields a chain with distribution $F_1{*}F_2$. Note that generally the order of concatenation matters for inhomogeneous chains and $F_1{*}F_2\neq F_2{*}F_1$ if $F_1\neq F_2$. This convolution of functions on the group can be calculated by direct sequential products of Fourier transform of each function as

$$\widehat{(F_1 * F_2)} = \hat{F}_2\hat{F}_1 \quad (12)$$

Note that unlike the case of the classical convolution theorem the order of multiplication matters.

2.2. PDF of end-to-end distance

In this section, we derive the probability density function of end-to-end distance for discrete-link polymer models using the Fourier transform obtained in the previous section. The PDF of end-to-end distance, denoted as $f(r)$, is, in fact, a marginal 1D

$$C(l_1,m_1;l_2,m_2|l,m) = (-1)^{l_1+l_2-l}\sqrt{2l+1}\Delta(l_1,l_2,l)\sqrt{(l_1-m_1)!(l_1+m_1)!(l_2-m_2)!(l_2+m_2)!(l-m_1-m_2)!(l+m_1+m_2)!} \times \sum_s(-1)^s[s!(l_1+l_2-l-s)!(l_2-m_2-s)!(l-l_2-m_1+s)!(l_1+m_1-s)!(l-l_1+m_2+s)!]^{-1} \quad (7)$$

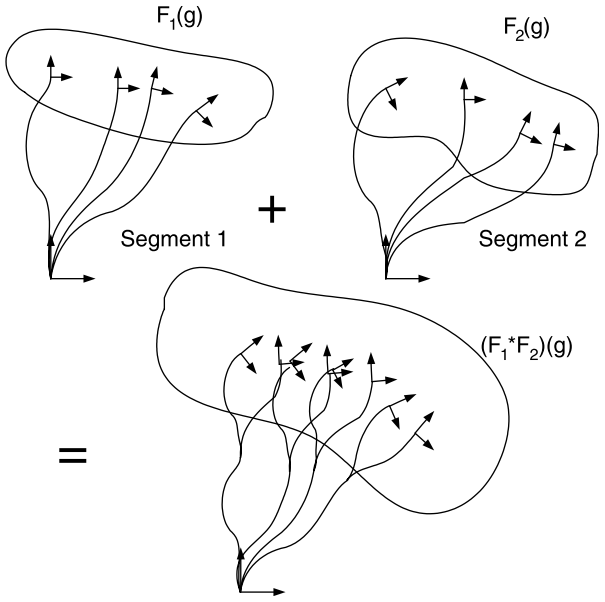

Fig. 1. Pictorial explanation of the concept of convolution. The second function is swept and weighted by the first. If the full distribution of positions and orientations of two adjacent segments of a polymer chain, F_1 and F_2, are known, then the concatenation of the segments yields a chain with distribution $F_1{*}F_2$.

PDF of the 6D PDF of relative end-to-end position and orientation, denoted as $F(g)$ where $g=(\mathbf{r}, R)\in SE(3)$. The final form of the result to be derived can be found in the literature [13,22,23]. However, since those do not contain detailed derivations, we derive it in this section.

The inverse Fourier transform of $\hat{F}$ can be obtained using Eqs. (9) or (10). To obtain the probability density function of end-to-end distance, let us first consider the integral over $SO(3)$ of $F(g)$

$$\int_{SO(3)} F(\mathbf{r},R)\mathrm{d}R = \frac{1}{2\pi^2}\sum_{s,l',l,m',m}\int_0^\infty p^2\mathrm{d}p\hat{F}^s_{l,m;l'm'}(p) \times\sum_{j=-l}^{l}[l',m'|p,s|l,j](\mathbf{r})\int_{SO(3)}\tilde{U}^l_{j,m}(R)\mathrm{d}R \tag{13}$$

If we separate and write for the last integral, then it becomes

$$\int_{SO(3)}\tilde{U}^l_{j,m}(R)\mathrm{d}R = \int_{\gamma=0}^{2\pi}\int_{\beta=0}^{\pi}\int_{\alpha=0}^{2\pi}(-1)^{m-j}\mathrm{e}^{-\mathrm{i}(j\alpha+m\gamma)}P^l_{j,m}(\cos\beta)\mathrm{d}R$$

where $\mathrm{d}R=(1/8\pi^2)\sin\beta\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\gamma$. In order for this integral to have non-zero value, one can easily find that $j=0$, and $m=0$. The integral on β, then simply becomes

$$\int_{\beta=0}^{\pi}P^l_{0,0}(\cos\beta)\sin\beta\mathrm{d}\beta = \int_{-1}^{1}P_l(x)\mathrm{d}x$$

Here, we use the relation $P^l_{0,0}(\cos\beta)=P_l(\cos\beta)$, where P_l is the lth Legendre polynomial, and substitute $\cos\beta$ into x. The integral of each Legendre polynomial becomes zero when $l\neq 0$. Hence, we find the condition that $l=0$. Looking at the range of summation, one can also find that $s=0$ should be satisfied. Therefore, Eq. (13) can be expressed in a compact form as

$$\int_{SO(3)}F(\mathbf{r},R)\mathrm{d}R = \frac{1}{2\pi^2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \times\int_0^\infty p^2\mathrm{d}p\hat{F}^0_{0,0;l',m'}(p)[l',m'|p,0|0,0](\mathbf{r}) \tag{14}$$

If we integrate Eq. (14) over the surface of a unit sphere and multiply by r^2, then the result will be the probability density function of end-to-end distance. Let that probability density function be denoted as $f(r)$, then it is of the form

$$f(r) = \frac{r^2}{2\pi^2}\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\left\{\int_{SO(3)}F(\mathbf{r},R)\mathrm{d}R\right\}\sin\theta\mathrm{d}\theta\mathrm{d}\phi \tag{15}$$

Since, $[l',m'|p,0|0,0](\mathbf{r})$ consists of such functions as $\mathrm{e}^{\mathrm{i}m\phi}$ and $P_{l'}(\cos\theta)$ due to the fact that $l=m=s=0$, one can easily find that $l'=m'=0$ by the similar reasoning with that given previously. Finally, we can get the end-to-end distribution as

$$f(r) = \frac{2r^2}{\pi}\int_0^\infty\hat{F}^0_{0,0;0,0}\frac{\sin(pr)}{pr}p^2\mathrm{d}p = \frac{2r}{\pi}\int_0^\infty\hat{F}^0_{0,0;0,0}\sin(pr)p\mathrm{d}p \tag{16}$$

Here, we use the equality $j_0(pr)=(\sin(pr))/pr$ [13].

3. Probability density function of end-to-end distance

As mentioned earlier, we examine discrete-chain polymer models. In this section, we derive the probability density function of end-to-end distance for three different discrete-chain polymer models.

3.1. The freely-rotating model

Let us assume the geometry shown in Fig. 2. Fig. 2 shows a schematic diagram of the ith link. In that figure, L_i corresponds to the length of the ith bond, α_i is the ith dihedral or torsional angle, and β_0 is the ith bond angle. Here, we use the convention that the local z-axis coincides with each bond. Then the position of the distal end with respect to the proximal end can be described by means of spherical coordinates, $\mathbf{r}=[r\cos\phi\sin\theta,\ r\sin\phi\sin\theta,\ r\cos\theta]^{\mathrm{T}}$.

In this case, $r=L_i$, and $\theta=0$. Torsional and bond angles are related to the rotation matrix of the distal end of link i with respect to its own proximal end, and can be described using by ZXZ Euler angles $R_{ZXZ}(\alpha,\beta,\gamma)$, in which case $\alpha=\alpha_i$, $\beta=\beta_0$, and $\gamma=0$.

According to the geometry shown, one can find that the appropriate form of the probability density function for this

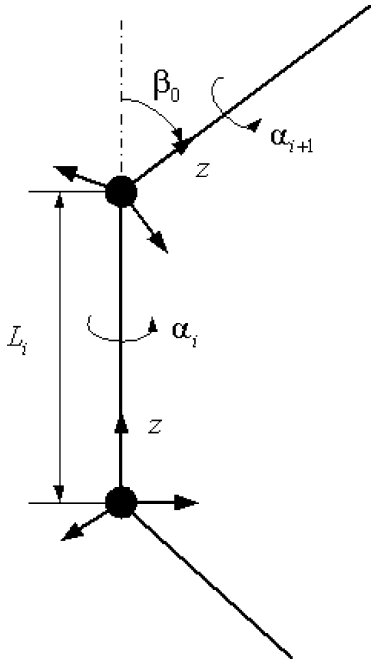

Fig. 2. Schematic diagram of one link. Reference frame is attached so that the local z axis coincides with each link. α_i is the ith torsional angle, and β_0 is the ith bond angle. These are described using ZXZ Euler angles.

single link can be described using Dirac delta functions. If we express it explicitly, it is of the form

$$^{i}F(\mathbf{r}, R(\alpha,\beta,\gamma)) = \frac{4\pi}{r^2 \sin\theta \sin\beta}\delta(\theta)\delta(r-L_i)\frac{1}{2\pi}\delta(\beta-\beta_0)\delta(\gamma) \tag{17}$$

Here, since the position of the distal end has the singularity associated with spherical coordinates ($\theta=0$), the ϕ value does not appear in the above equation. Instead, the effect of integration with respect to ϕ is included in the constant, so that consequently, the constant term contains $4\pi/r^2 \sin\theta$ [13]. As for the α angle, which corresponds to the ith torsional angle α_i of a polymer, we assume that rotation around this angle to be uniformly distributed, so that the probability density of those angles is $1/2\pi$. This assumption makes it the freely-rotating chain model. Then we can take the $SE(3)$-Fourier transform for the function in Eq. (17) by using Eq. (8). From Ref. [13], one sees that

$$\overline{U^s_{l,m;l',m'}(g;p)} = (-1)^{(l-l')}(-1)^{(m-m')}U^s_{l,-m;l',-m'}(g;p)$$

With the above expression, Eq. (1), the properties of Dirac delta function, and the fact that $\int_{\phi=0}^{2\pi} e^{im\phi}d\phi$ has non-zero value only when $m=0$, we can derive the probability density function of the ith link as

$$^{i}\hat{F}^s_{l',m';l,m}(p) = (-1)^{(l-l')}[l,0|p,s|l',0](\mathbf{r}_0)P^{l'}_{m',0}(\cos\beta_0), \quad m=0 \tag{18}$$

where $\mathbf{r}_0$ means the position vector of distal end with respect to the reference frame attached to the proximal end. In component form, it can be written as

$$^{i}\hat{F}^s_{l',m';l,m}(p) = (-1)^{(l-l')}\delta_{m,0} \times\sqrt{\frac{2l+1}{2l'+1}}\sum_{k=|l-l'|}^{l+l'}\left\{i^k(2k+1)j_k(pL_i) \times C(k,0;l,s|l',s)C(k,0;l,0|l',0)P^{l'}_{m',0}(\cos\beta_0)\right\} \tag{19}$$

where $\delta_{m,0}$ is a Kronecker delta.

Since, we get the Fourier transform of the probability density function of the ith link derived above, now we can obtain the Fourier transform of an N-link polymer by utilizing the generalized convolution on $SE(3)$, which is simply expressed in Eq. (12). Let us denote the Fourier transform of an N-link polymer as $\hat{F}$. This can be obtained simply by multiplying each Fourier transform in reversed order as

$$\hat{F} = {}^{N}\hat{F}\ldots{}^{1}\hat{F} \tag{20}$$

when we apply the above arguments to obtain the end-to-end distance distribution. Since, we only need to consider the case when $s=0$, Eq. (19) can be further simplified to the following form

$$^{i}\hat{F}^0_{l',m';l,m}(p) = (-1)^{(l-l')}\delta_{m,0}\sqrt{\frac{2l+1}{2l'+1}} \times\sum_{k=|l-l'|}^{l+l'}\left\{i^k(2k+1)j_k(pL_i)(C(k,0;l,0|l',0))^2 \times P^{l'}_{m',0}(\cos\beta_0)\right\} \tag{21}$$

where $\delta_{m,0}$ is a Kronecker delta. In the above equation, the Clebsch–Gordan coefficient can be calculated by the following simple formula [39]

$$C(a,0;b,0|c,0) = \frac{(-1)^{g-c}\sqrt{2c+1}\,g!}{(g-a)!(g-b)!(g-c)!}\sqrt{\frac{(2g-2a)!(2g-2b)!(2g-2c)!}{(2g+1)!}}$$

when $a+b+c=2g$, where g is a positive integer. When $a+b+c=2g+1$, the corresponding Clebsch–Gordan coefficients have zero values. Then by utilizing Eqs. (20) and (16), we can obtain the probability density function of end-to-end distance for the N-link polymer chain.

3.2. The independent energy model

In this model, the potential energy is expressed as

$$E(\alpha_1,\ldots,\alpha_N) = \sum_{i=1}^{N}E_i(\alpha_i)$$

With this in mind, we define the PDF of the ith link as

$$ {}^{i}F(g) = \frac{4\pi}{r^2 \sin\theta \sin\beta}\delta(\theta)\delta(r-L_i)\frac{\mathrm{e}^{-E_i(\alpha)/k_{\mathrm{B}}T}}{Z}\delta(\beta-\beta_0)\delta(\gamma) \tag{22} $$

where $g=(\mathbf{r},R(\alpha,\beta,\gamma))\in SE(3)$ and the partition function Z is defined as $Z=\int_0^{2\pi}\mathrm{e}^{-E_i(\alpha)/k_{\mathrm{B}}T}\mathrm{d}\alpha$. The Fourier transform of this function becomes

$$ \begin{aligned} {}^{i}\hat{F}^{s}_{l',m';l,m}(p) &= \int_{SE(3)} {}^{i}F(g)\overline{U^{s}_{l,m;l',m'}(g;p)}\mathrm{d}(g) \\ &= \int_{SE(3)} \frac{1}{2\pi}\delta(\theta)\delta(r-L_i)\frac{\mathrm{e}^{-E_i(\alpha)/k_{\mathrm{B}}T}}{Z} \\ &\times \delta(\beta-\beta_0)\delta(\gamma)(-1)^{l-l'}(-1)^{m-m'} \\ &\times \sum_{j=-l'}^{l'}[l,-m|p,s|l',j](\mathbf{r})\tilde{U}^{l'}_{j,-m'}(R)\mathrm{d}r\mathrm{d}\phi\mathrm{d}\theta\mathrm{d}\alpha\mathrm{d}\beta\mathrm{d}\gamma \end{aligned} \tag{23} $$

Using the fact that $\int_0^{2\pi}\mathrm{e}^{\mathrm{i}(j+m)\phi}\mathrm{d}\phi$ vanishes except when $j=-m$, $|m|\le l'$ Eq. (23) is further simplified to the following form

$$ \begin{aligned} {}^{i}\hat{F}^{s}_{l',m';l,m}(p) =& (-1)^{l-l'}P^{l'}_{m',m}(\cos\beta_0)\int_0^{2\pi}\frac{\mathrm{e}^{-E_i(\alpha)/k_{\mathrm{B}}T}\mathrm{e}^{\mathrm{i}m\alpha}}{Z}\mathrm{d}\alpha \\ &\times\sqrt{\frac{2l+1}{2l'+1}}\sum_{k=|l-l'|}^{l+l'}\mathrm{i}^{k}(2k+1)j_k(pL_i) \\ &\times C(k,0;l,s|l',s)C(k,0;l,-m|l',m) \end{aligned} \tag{24} $$

only when $|m|\le l'$. If not, the above Fourier transform becomes zero. After that, by using Eqs. (20) and (16) we can obtain the end-to-end distance distribution.

3.3. The interdependent pairwise energy model

As we did in the freely-rotating model and in the independent energy model, let us assume the geometry shown in Fig. 2. The difference now is that

$$ E(\alpha_1,\ldots,\alpha_N) = \sum_{i=1}^{N-1}E_{i,i+1}(\alpha_i,\alpha_{i+1}) \tag{25} $$

First, we define the following. Let $\mathrm{d}(g_{-\alpha})$ be $((r^2\sin\theta\ \sin\beta)/8\pi^2)\mathrm{d}r\mathrm{d}\phi\mathrm{d}\theta\mathrm{d}\beta\mathrm{d}\gamma$, and $\int_{SE(3)-\alpha}=\int_\gamma\int_\beta\int_\theta\ \int_\phi\int_r$, i.e. only $\int_\alpha(\cdot)\mathrm{d}\alpha$ is missing in the integration and measure. We also define ${}^{i}F(g)=(4\pi/(r^2\sin\theta\ \sin\beta))\delta(r-L_i)\delta(\theta)(\beta-\beta_0)\delta(\gamma)$. Then referring to [13], we can define

$$ \begin{aligned} {}^{i}\hat{\hat{F}}^{s}_{l',m';l,m}(p,\alpha_i) &= \int_{SE(3)-\alpha} {}^{i}F(g)\overline{U^{s}_{l,m;l',m'}(g;p)}\mathrm{d}(g_{-\alpha}) \\ &= \int_{SE(3)-\alpha}\frac{1}{2\pi}\delta(\theta)\delta(r-L_i)\delta(\beta-\beta_0)\delta(\gamma)(-1)^{l-l'} \\ &\times(-1)^{m-m'}\sum_{j=-l'}^{l'}[l,-m|p,s|l',j](\mathbf{r}) \\ &\times\tilde{U}^{l'}_{j,-m'}(R)\mathrm{d}r\mathrm{d}\phi\mathrm{d}\theta\mathrm{d}\beta\mathrm{d}\gamma \end{aligned} \tag{26} $$

By the same analogy as in the previous case, due to the fact that the integral $\int_{\phi=0}^{2\pi}\mathrm{e}^{\mathrm{i}(j+m)\phi}\mathrm{d}\phi$ has non-zero value only when $j=-m$ only for $|m|\le l'$, it becomes

$$ \begin{aligned} &{}^{i}\hat{\hat{F}}^{s}_{l',m';l,m}(p,\alpha_i) \\ &= (-1)^{l-l'}\sqrt{\frac{2l'+1}{2l+1}}P^{l'}_{m',m}(\cos\beta_0)\mathrm{e}^{\mathrm{i}m\alpha_i}\sum_{k=|l-l'|}^{l+l'}\mathrm{i}^{k} \\ &\times(2k+1)j_k(pL_i)C(k,0;l,s|l',s)C(k,0;l,-m|l',-m) \end{aligned} \tag{27} $$

Now, let us consider the Boltzmann-weighted convolution of the ith and $i+1$st links of the form

$$ {}^{i,i+1}F = \int_{SE(3)} {}^{i}F(h)^{i+1}F(h^{-1}\circ g)\mathrm{e}^{-E_{i,i+1}/k_{\mathrm{B}}T}\mathrm{d}h \tag{28} $$

If we take the $SE(3)$-Fourier Transform of the convolution, then it is written as

$$ \begin{aligned} {}^{i,i+1}\hat{F}^{s}(p) &= \int_{SE(3)}\int_{SE(3)} {}^{i}F(h)^{i+1}F(h^{-1}\circ g)\mathrm{e}^{-E_{i,i+1}/k_{\mathrm{B}}T}\mathrm{d}hU^{s}(g^{-1};p)\mathrm{d}g \\ &= \int_{SE(3)}\int_{SE(3)} {}^{i}F(h)^{i+1}F(g')\mathrm{e}^{-E_{i,i+1}/k_{\mathrm{B}}T}U^{s}(g'^{-1};p) \\ &\times U^{s}(h^{-1};p)\mathrm{d}h\mathrm{d}g' \\ &= \int_{\alpha_i}\int_{\alpha_{i+1}}\left(\int_{SE(3)-\alpha} {}^{i+1}F(g')\overline{U^{s}(g';p)}^{T}\mathrm{d}(g'_{-\alpha})\right) \\ &\times\left(\int_{SE(3)-\alpha} {}^{i}F(h)\overline{U^{s}(h;p)}^{\mathrm{T}}\mathrm{d}(h_{-\alpha})\right)\mathrm{e}^{-E_{i,i+1}/k_{\mathrm{B}}T}\mathrm{d}\alpha_{i+1}\mathrm{d}\alpha_i \end{aligned} \tag{29} $$

Here we use $g'=h^{-1}\circ g$, $\mathrm{d}g'=\mathrm{d}g$, $U^{s}(g'^{-1}\circ h^{-1})=U^{s}(g'^{-1})U^{s}(h^{-1})$, and $U^{s}(g^{-1})=\overline{U^{s}(g)}^{\mathrm{T}}$. By means of Eqs. (26) and (27),

Eq. (29) can be written in the compact form

$$ {}^{i,i+1}\hat{F}^s(p) = \int_{\alpha_i}\int_{\alpha_{i+1}} {}^{i+1}\hat{\hat{F}}^s(p,\alpha_{i+1})\,{}^{i}\hat{\hat{F}}^s(p,\alpha_i)\mathrm{e}^{-E_{i,i+1}/k_\mathrm{B}T}\mathrm{d}\alpha_{i+1}\mathrm{d}\alpha_i \tag{30} $$

Suppose that there are N links, and let the number density of the distal end be ${}^{\mathrm{num}}F$, then the Fourier Transform gives

$$ {}^{num}\hat{F}^s(p) = \int_{\boldsymbol{\alpha}} {}^{N}\hat{\hat{F}}(p,\alpha_N)\,{}^{N-1}\hat{\hat{F}}(p,\alpha_{N-1})\ldots{}^{1}\hat{\hat{F}}(p,\alpha_1)\mathrm{e}^{-E(\boldsymbol{\alpha})/k_\mathrm{B}T}\mathrm{d}\boldsymbol{\alpha} \tag{31} $$

where $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_N)$ and $E(\boldsymbol{\alpha})=\sum_{i=1}^{N-1}E_{i,i+1}(\alpha_i,\alpha_{i+1})$. Then the Fourier Transform of the PDF can be obtained by normalization as

$$ \hat{F}^s = \frac{{}^{num}\hat{F}^s}{Z} \tag{32} $$

where the partition function Z is defined as

$$ Z = \int_{\boldsymbol{\alpha}} \mathrm{e}^{-E(\boldsymbol{\alpha})/k_\mathrm{B}T}\mathrm{d}\boldsymbol{\alpha} = \int_{\alpha_1}\ldots\int_{\alpha_N}\mathrm{e}^{-E(\boldsymbol{\alpha})/k_\mathrm{B}T}\mathrm{d}\alpha_N\ldots\mathrm{d}\alpha_1 \tag{33} $$

Now, we employ a similar method as in Refs. [13,35]. First, let us define for $i+1<l-1$

$$ \begin{aligned} \hat{\hat{F}}^s(p;\alpha_i,\alpha_l) = \int_{\alpha_{i+1}}\ldots\int_{\alpha_{l-1}} {}^{l}\hat{\hat{F}}^s(p,\alpha_l)\,{}^{l-1}\hat{\hat{F}}^s(p,\alpha_{l-1})\ldots{}^{i+1}\hat{\hat{F}}^s(p,\alpha_{i+1}) \\ \times{}^{i}\hat{\hat{F}}^s(p,\alpha_i)\mathrm{e}^{-E_{i,l}(\alpha_i,\ldots,\alpha_l)/k_\mathrm{B}T}\mathrm{d}\alpha_{l-1}\ldots\mathrm{d}\alpha_{i+1} \end{aligned} \tag{34} $$

Then, for example assuming that we divide the total chain into two segments $(\alpha_0,\ldots,\alpha_i)$ and $(\alpha_{i+1},\ldots,\alpha_N)$, the Fourier Transform of the number density ${}^{\mathrm{num}}F$ becomes

$$ \begin{aligned} {}^{num}\hat{F}^s(p) = \int_{\alpha_0}\int_{\alpha_N}\int_{\alpha_i}\int_{\alpha_{i+1}} \hat{\hat{F}}^s(p;\alpha_{i+1},\alpha_N)\hat{\hat{F}}^s(p;\alpha_0,\alpha_i) \\ \times\mathrm{e}^{-E_{i,i+1}(\alpha_i,\alpha_{i+1})/k_\mathrm{B}T}\mathrm{d}\alpha_{i+1}\mathrm{d}\alpha_i\mathrm{d}\alpha_N\mathrm{d}\alpha_0 \end{aligned} \tag{35} $$

In practice, we can break the whole chain apart into segments with 2 or 3 monomers. More specifically, let us define

$$ \hat{\hat{F}}^s(p;\alpha_i,\alpha_{i+1}) = {}^{i+1}\hat{\hat{F}}^s(p,\alpha_{i+1})\,{}^{i}\hat{\hat{F}}^s(p,\alpha_i)\mathrm{e}^{-E_{i,i+1}/k_\mathrm{B}T} \tag{36} $$

Then Eq. (34) becomes

$$ \begin{aligned} \hat{\hat{F}}^s(p;\alpha_i,\alpha_{i+3}) = \int_{\alpha_{i+1}}\int_{\alpha_{i+2}} \hat{\hat{F}}^s(p;\alpha_{i+2},\alpha_{i+3})\hat{\hat{F}}^s(p;\alpha_i,\alpha_{i+1}) \\ \times\mathrm{e}^{-E_{i+1,i+2}(\alpha_{i+1},\alpha_{i+2})/k_\mathrm{B}T}\mathrm{d}\alpha_{i+2}\mathrm{d}\alpha_{i+1} \end{aligned} \tag{37} $$

and we can apply this equation sequentially to reach Eq. (35). After that, we can apply the same normalization as in Eq. (32). Then the end-to-end distance distribution $f(r)$ can be calculated from Eq. (16).

Note that, in order to obtain the partition function Z, we can apply the same method described above. Specifically, if we use 1 instead of ${}^{i}\hat{\hat{F}}(p;\alpha_i)$ and ${}^{i+1}\hat{\hat{F}}(p;\alpha_{i+1})$ in Eq. (36), then the convolution-like function eventually gives the partition function.

4. Efficient implementation

In numerical work, computational cost is often a critical issue. In this section, we mention efficient methods for calculating the end-to-end distribution with the proposed method.

Among the three different models presented in this paper, the computational speed for the freely rotating and the independent energy model cases is faster than that of the interdependent energy model. We can store the Fourier transform of each link in advance as functions of the frequency factor p, then can apply Eq. (12) to obtain the end-to-end distribution of a given polymer. This process is nothing more than matrix multiplication.

However, when it comes to the interdependent energy model case, the situation is not as simple as the two other cases. As one can see from Eq. (37), one needs double integration at each step. For example, assume that we divide one torsion angle into n cells. Then the number of points in each α_i becomes $n+1$. Each process of matrix multiplication and summation in Eq. (37) needs $O(n^4)$ computations. The main problem is that if we use $n=50$ or greater than that ($n=100$, for example), which is required in most cases because one needs a large value of n to avoid aliasing effects in the Fourier transform, then $O(n^4)$ is really a huge number, which means that the direct implementation of Eq. (37) is not an efficient way to implement the method for the interdependent energy model case. For this reason, we present more efficient ways of implementing the interdependent energy model.

First, we compute the Fourier series of $u_{i,i+1}(\alpha_i,\alpha_{i+1})=\mathrm{e}^{-E_{i,i+1}(\alpha_i,\alpha_{i+1})/k_\mathrm{B}T}$. Then it becomes

$$ u_{i,i+1}(\alpha_i,\alpha_{i+1}) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} {}^{i,i+1}\hat{u}_{m,n}\mathrm{e}^{\mathrm{i}m\alpha_i}\mathrm{e}^{\mathrm{i}n\alpha_{i+1}} \tag{38} $$

In practice, if we use a band width, B, for approximating the exponential of the energy function, then we can utilize the exponential of the energy function as

$$ u_{i,i+1}(\alpha_i,\alpha_{i+1}) \approx \sum_{m=-B}^{B}\sum_{n=-B}^{B} {}^{i,i+1}\hat{u}_{m,n}\mathrm{e}^{\mathrm{i}m\alpha_i}\mathrm{e}^{\mathrm{i}n\alpha_{i+1}} \tag{39} $$

Here the Fourier coefficient $\hat{u}$ is defined as

$$ {}^{i,i+1}\hat{u}_{m,n} = \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi} u_{i,i+1}(\alpha_i,\alpha_{i+1})\mathrm{e}^{-\mathrm{i}m\alpha_i}\mathrm{e}^{-\mathrm{i}n\alpha_{i+1}}\mathrm{d}\alpha_i\mathrm{d}\alpha_{i+1} $$

Also a closer investigation of Eq. (27) gives

$$ {}^{i}\hat{\hat{F}}^s_{l',m';l,m}(p,\alpha_i) = {}^{i}D_{l',m';l,m}(p)\mathrm{e}^{\mathrm{i}m\alpha_i} \tag{40} $$

where

$$
{}^{i}D_{l',m';l,m}(p) = (-1)^{l-l'}\sqrt{\frac{2l'+1}{2l+1}}P^{l'}_{m',m}(\cos\beta_0) \times \sum_{k=|l-l'|}^{l+l'} \mathrm{i}^k(2k+1)j_k(pL_i)C(k,0;l,s|l',s) \times C(k,0;l,-m|l',-m) \tag{41}
$$

for $|m| \le l'$. Otherwise it becomes zero. If we express it in matrix form, it becomes

$$
{}^{i}\hat{\hat{F}}^{s} = {}^{i}D\left[\backslash \mathrm{e}^{\mathrm{i}m\alpha_i} \backslash\right] \tag{42}
$$

where $[\backslash \mathrm{e}^{\mathrm{i}m\alpha_i} \backslash]$ means the diagonal matrix whose diagonal elements are of the form $\mathrm{e}^{\mathrm{i}m\alpha_i}$. Let us assume that there are 8 links in a given segment of polymer. We can construct the following by means of Eqs. (36) and (39)

$$
{}^{1,2}\hat{\hat{F}}^{s} = {}^{2}\hat{\hat{F}}^{s}{}^{1}\hat{\hat{F}}^{s}u_{1,2}(\alpha_1,\alpha_2) = \sum_{j_1=-B}^{B}\sum_{k_1=-B}^{B} {}^{2}\hat{\hat{F}}^{s}{}^{1}\hat{\hat{F}}^{s}{}^{1,2}\hat{u}_{j_1,k_1}\mathrm{e}^{\mathrm{i}j_1\alpha_1}\mathrm{e}^{\mathrm{i}k_1\alpha_2} \tag{43}
$$

Similarly, we can construct for the 3rd and 4th links as

$$
{}^{3,4}\hat{\hat{F}}^{s} = \sum_{j_2=-B}^{B}\sum_{k_2=-B}^{B} {}^{4}\hat{\hat{F}}^{s}{}^{3}\hat{\hat{F}}^{s}{}^{3,4}\hat{u}_{j_2,k_2}\mathrm{e}^{\mathrm{i}j_2\alpha_3}\mathrm{e}^{\mathrm{i}k_2\alpha_4} \tag{44}
$$

Then ${}^{1,4}\hat{\hat{F}}^{s}$ can be expressed, by Eq. (42), as

$$
{}^{1,4}\hat{\hat{F}}^{s} = \sum_{j_1,k_1}\sum_{j_2,k_2}\sum_{j,k} {}^{4}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_4+k_2)\alpha_4} \backslash\right] \times \int_{\alpha_3}\int_{\alpha_2} {}^{3}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_3+j_2)\alpha_3} \backslash\right]^{3,4}\hat{u}_{j_2,k_2}{}^{2}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_2+k_1)\alpha_2} \backslash\right] \times {}^{1,2}\hat{u}_{j_1,k_1}{}^{2,3}\hat{u}_{j,k}\mathrm{e}^{\mathrm{i}j\alpha_2}\mathrm{e}^{\mathrm{i}k\alpha_3}\mathrm{d}\alpha_2\mathrm{d}\alpha_3{}^{1}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_1+j_1)\alpha_1} \backslash\right] \tag{45}
$$

where each m_i represents the index in Eq. (42) for the ith link. Due to the fact that the integral $\int_{\alpha=0}^{2\pi} \mathrm{e}^{\mathrm{i}m\alpha}\mathrm{d}\alpha$ has non-zero value only when $m=0$, this can be simplified further to the following form

$$
{}^{1,4}\hat{\hat{F}}^{s} = \sum_{j_1=-B}^{B}\sum_{k_2=-B}^{B} {}^{4}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_4+k_2)\alpha_4} \backslash\right]^{2,3}C_{j_1,k_2}{}^{1}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_1+k_1)\alpha_1} \backslash\right] \tag{46}
$$

where

$$
{}^{2,3}C_{j_1,k_2} = \sum_{j_2,k}\sum_{j,k_1} {}^{3}D\left[\backslash 2\pi\delta_{j_2+k,-m_3} \backslash\right]^{2}D\left[\backslash 2\pi\delta_{j+k_1,-m_2} \backslash\right] \times {}^{3,4}\hat{u}_{j_2,k_2}{}^{1,2}\hat{u}_{j_1,k_1}{}^{3,4}\hat{u}_{j,k} \tag{47}
$$

where $\delta_{j,k}$ is a Kronecker delta. Here ${}^{1,2}\hat{u}_{j_1,k_1}$, ${}^{3,4}\hat{u}_{j_2,k_2}$, and ${}^{2,3}\hat{u}_{j,k}$ represent the Fourier coefficients for $E_{1,2}(\alpha_1,\alpha_2)$, $E_{3,4}(\alpha_3,\alpha_4)$, and $E_{2,3}(\alpha_2,\alpha_3)$, respectively. Similarly,

$$
{}^{5,8}\hat{\hat{F}}^{s} = \sum_{j_1'=-B}^{B}\sum_{k_2'=-B}^{B} {}^{8}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_8+k_2')\alpha_8} \backslash\right]^{6,7}C_{j_1'k_2'}{}^{5}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_5+j_1')\alpha_5} \backslash\right] \tag{48}
$$

With Eqs. (46) and (48), we can obtain the following

$$
{}^{1,8}\hat{\hat{F}}^{s} \sum_{j_1=-B}^{B}\sum_{k_2'=-B}^{B} {}^{8}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_8+k_2')\alpha_8} \backslash\right]^{2,7}C_{j_1k_2'}{}^{1}D\left[\backslash \mathrm{e}^{\mathrm{i}(m_1+j_1)\alpha_1} \backslash\right] \tag{49}
$$

where

$$
{}^{2,7}C_{j_1k_2'} = \sum_{j_1',k}\sum_{j,k_2} {}^{6,7}C_{j_1k_2'}{}^{5}D\left[\backslash 2\pi\delta_{j_1'+k,-m_5} \backslash\right] \times {}^{4}D\left[\backslash 2\pi\delta_{j+k_2,-m_4} \backslash\right]^{4,5}\hat{u}_{j,k}{}^{2,3}C_{j_1k_2} \tag{50}
$$

Here ${}^{4,5}\hat{u}_{j,k}$ is the Fourier coefficient for $E_{4,5}(\alpha_4,\alpha_5)$. If there are more than eight links, then one can repeat Eqs. (49) and (50) to obtain ${}^{1,N}\hat{\hat{F}}^{s}$ where N is the number of links. At the final step, we can obtain ${}^{num}\hat{\hat{F}}^{s}$ by

$$
{}^{num}\hat{\hat{F}}^{s} = \sum_{k_2'}\sum_{j_1} {}^{N}D\left[\backslash 2\pi\delta_{k_2',-m_N} \backslash\right]^{2,N-1}C_{j_1k_2'}{}^{1}D\left[\backslash 2\pi\delta_{j_1,-m_1} \backslash\right] \tag{51}
$$

The partition function Z can be calculated similarly with all ${}^{i}\hat{\hat{F}}^{s}$'s replaced by 1 in the above procedure.

The above approach has an advantage compared to the direct double integration in that we do not need to perform integration. Instead, we can only select the set of indices which makes the integration of $e^{\mathrm{i}m\alpha}$ part nonzero. Then, the total computational cost for each summation and matrix multiplication processes becomes $O(B^4 \times a^2)$, where the maximum value of a is $O(B)$. In practice, the number of B as 4–7 can give a good approximation of the exponential of energy function. Another issue is that, since $u_{i,i+1}(\alpha_i,\alpha_{i+1})$ is already expressed in terms of harmonics, we do not need a large value of N_b, which is the band width in the Fourier transform for $SE(3)$, compared with the case where the original energy function is used. Since the computational cost also depends on the size of Fourier transform matrix for $SE(3)$, one can find that the approach presented in this section is much faster than direct double integration approach.

5. Application to polypeptide chains

We can also apply our proposed method to polypeptide chains. Polypeptide chains have interesting features compared with other general chains, such as polyethylene, etc. We depict the diagram of polypeptide chain structure in Fig. 3. First, the torsion angle around the C–N bond is fixed to be 180°. The torsion angle between C_α and C is called ψ angle, and that between N and C_α is called ϕ angle. The allowable range of values for the angle pair (ψ,ϕ) is obtained from the Ramachandran plot [40]. This also shows that the behavior

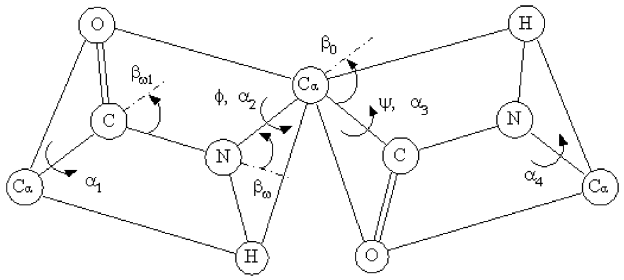

Fig. 3. Schematic diagram of a polypeptide chain. The torsion angle of the C–N bond is fixed at 180°. In numerical examples, β_ω and $\beta_{\omega 1}$ are assumed 63.8° and 58.3°, respectively.

of a polypeptide chain can be described using interdependent pairwise energy model. In fact, this is already known as 'the Flory isolated-pair hypothesis' [1,41]. Although it turns out that this isolated-pair hypothesis requires some modification [41], it still serves as a good approximation to describe some features of polypeptide chains. In this section, we apply our proposed method to obtain end-to-end distance distribution of polypeptide chain models defined by the Ramachandran plot.

Looking at Fig. 3, as mentioned earlier, unlike other polymers such as polyethylene, etc. the C–N bond does not have energy interaction with two adjacent bonds, C_α–C and N–C_α. Let the probability density function of the C–N bond be ${}^{\omega}F$. If the torsional angle along the C–N bond is 180°, or π radians, then ${}^{\omega}F$ becomes

$$ {}^{\omega}F(g) = \frac{4\pi}{r^2 \sin\theta \sin\beta}\delta(\alpha-\pi)\delta(\beta-\beta_\omega)\delta(\gamma)\delta(r-L_\omega)\delta(\theta) \tag{52} $$

where β_ω and L_ω are bond angle and bond length of C–N bond, respectively. Here the ω angle is assumed to be 180° or π radians. The Fourier Transform for $SE(3)$ gives

$$ \begin{aligned} {}^{\omega}\hat{F}^{s}_{l',m';l,m} &= \mathrm{e}^{\mathrm{i}\pi m} P^{l'}_{m',m}(\cos\beta_\omega)(-1)^{l-l'}\sqrt{\frac{2l+1}{2l'+1}} \\ &\quad \times \sum_{k=|l-l'|}^{l+l'} \mathrm{i}^k (2k+1) C(k,0;l,-m|l',-m) \\ &\quad \times C(k,0;l,s|l',s) j_k(pL_\omega) \end{aligned} \tag{53} $$

for $|m| \le l'$. Otherwise, this becomes zero. One can, then, define the following together with Eq. (26)

$$ \hat{\hat{F}}^{s}(p;\alpha_i,\alpha_{i+1}) = \hat{\hat{F}}^{s}(p;\alpha_{i+1})\,{}^{\omega}\hat{F}^{s}\hat{\hat{F}}^{s}(p;\alpha_i) \tag{54} $$

Then we can apply the method in the previous section to this polypeptide chain model. Note that the polypeptide chain case has much simpler form than general polymer chains due to the fact that

$$ E(\boldsymbol{\alpha}) = E_{1,2}(\alpha_1,\alpha_2) + E_{3,4}(\alpha_3,\alpha_4) + \ldots + E_{N-1,N}(\alpha_{N-1},\alpha_N) \tag{55} $$

That is, each pairwise energy becomes independent. Hence, we can use an energy function with the following form

$$ u_{i,i+1}(\alpha_i,\alpha_{i+1}) = \frac{\mathrm{e}^{-E_{i,i+1}(\alpha_i,\alpha_{i+1})/k_\mathrm{B}T}}{Z} \tag{56} $$

where

$$ Z = \int_{\alpha_{i+1}}\int_{\alpha_i} \mathrm{e}^{-E_{i,i+1}(\alpha_i,\alpha_{i+1})/k_\mathrm{B}T}\,\mathrm{d}\alpha_i\mathrm{d}\alpha_{i+1} \tag{57} $$

First we consider 8 residues among the polypeptide chain consisting of N residues. For convenience, let us assume that N has the form of 2^k where k is a positive integer greater than 3. We can construct the pairwise Fourier Transform-like matrices as

$$ {}^{i,i+1}\hat{\hat{F}}^{s} = {}^{i+1}\hat{\hat{F}}^{s}\,{}^{\omega}\hat{\hat{F}}^{s}\,{}^{i}\hat{\hat{F}}^{s} \tag{58} $$

for $i=1$, 3. Then

$$ {}^{1,4}\hat{\hat{F}}^{s} = {}^{4}D\left[\backslash \mathrm{e}^{\mathrm{i}m_4\alpha_4}\backslash\right]^{2,3}C^{1}D\left[\backslash \mathrm{e}^{\mathrm{i}m_1\alpha_1}\backslash\right] \tag{59} $$

where

$$ {}^{2,3}C = \sum_{j=-B}^{B}\sum_{k=-B}^{B} {}^{\omega}\hat{F}^{s\,3}D\left[\backslash 2\pi\delta_{k,-m_3}\backslash\right]^{2}D\left[\backslash 2\pi\delta_{j,-m_2}\backslash\right]{}^{\omega}\hat{F}^{s\,2,3}\hat{u}_{j,k} \tag{60} $$

We can do similarly with the subchain consisting of residues 5–8 as

$$ {}^{5,8}\hat{\hat{F}}^{s} = {}^{8}D\left[\backslash \mathrm{e}^{\mathrm{i}m_8\alpha_8}\backslash\right]^{6,7}C^{5}D\left[\backslash \mathrm{e}^{\mathrm{i}m_5\alpha_5}\backslash\right] \tag{61} $$

Then

$$ {}^{1,8}\hat{\hat{F}}^{s} = {}^{8}D\left[\backslash \mathrm{e}^{\mathrm{i}m_8\alpha_8}\backslash\right]^{2,7}C^{1}D\left[\backslash \mathrm{e}^{\mathrm{i}m_1\alpha_1}\backslash\right] \tag{62} $$

where

$$ {}^{2,7}C = \sum_{j=-B}^{B}\sum_{k=-B}^{B} {}^{6,7}C^{5}D\left[\backslash 2\pi\delta_{k,-m_5}\backslash\right]^{4}D\left[\backslash 2\pi\delta_{j,-m_4}\backslash\right]^{2,3}C^{4,5}\hat{u}_{j,k} \tag{63} $$

After that, we can repeat until N links are reached. Finally we can obtain the Fourier Transform of end-to-end distance probability density function $F(g)$ as

$$ \begin{aligned} \hat{F}^{s}(p) &= \frac{1}{4\pi^2}{}^{N}D\left[\backslash 2\pi\delta_{m_N,0}\backslash\right]^{2,N-1}C^{1}D\left[\backslash 2\pi\delta_{m_1,0}\backslash\right] \\ &= {}^{N}D\left[\backslash \delta_{m_N,0}\backslash\right]^{2,N-1}C^{1}D\left[\backslash \delta_{m_1,0}\backslash\right] \end{aligned} \tag{64} $$

Here $4\pi^2$ in the first line is a normalization factor. Because all $u(\alpha_i,\alpha_{i+1})$ are normalized according to Eq. (56), the integration about the final two angles α_N and α_1 requires a normalization factor, which becomes $\int_{\alpha_N}\int_{\alpha_1}\mathrm{d}\alpha_1\mathrm{d}\alpha_N = 4\pi^2$.

6. Numerical examples

In this section three kinds of examples are demonstrated: the freely-rotating chain, the independent energy chain, and the interdependent pairwise energy chain.

6.1. The freely-rotating and the independent energy chain model

First let us take an example for the freely-rotating model. In order to verify our model, we utilize well-known formulas for the freely rotating chain model [1]. According to the Flory's theory [1], when all the link lengths are the same and denoted as L, and all the bond angles have the same values as θ, then the average of the square of the end-to-end distance for the N-link polymer chain can be calculated as [1,13]

$$\langle r^2\rangle = NL^2(1+\alpha)(1-\alpha)^{-1} - (2\alpha L^2)(1-\alpha^N)(1-\alpha)^{-2} \quad (65)$$

where $\alpha = \cos\theta$. If we further normalize the end-to-end distance with the total chain length (divide Eq. (65) by N^2L^2), the equation becomes

$$\langle r^2\rangle_{\text{normalized}} = \frac{(1+\alpha)(1-\alpha)^{-1}}{N} - \left(\frac{2\alpha}{N^2}\right)(1-\alpha^N)(1-\alpha)^{-2} \quad (66)$$

In Fig. 4 are shown the resulting end-to-end distance probability density functions for two different cases. Here the number of links in the polymer chain is fixed to be 20. If we calculate the area under the curves, it gives 1.0000, which means that the obtained curves truly represent the probability density functions. In Table 1 are shown the squares of the end-to-end distance. One can see that the corresponding results are in excellent agreement with those from Eq. (66). In the simulations, as the band width for Fourier transform for $SE(3)$ and upper limit of integration with respect to frequency factor p, denoted as N_b and L_p, respectively, $N_b = 12$ and $L_p = 50$ are used in the case of $\theta = \pi/4$, and $N_b = 16$ and $L_p = 60$ are used in the case of $\theta = \pi/6$.

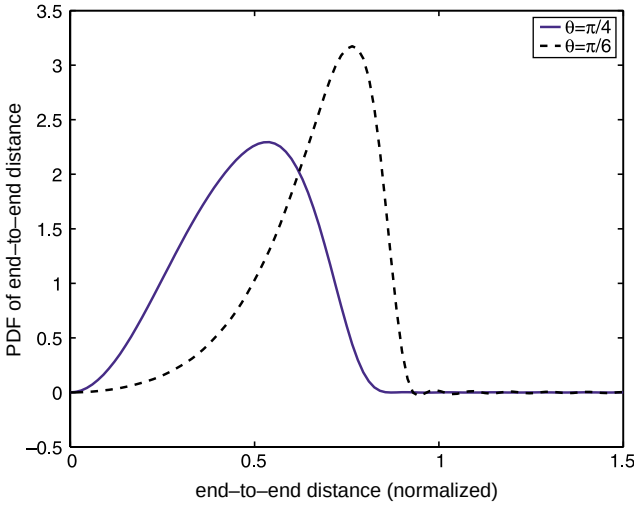

Fig. 4. Plot of the resulting end-to-end distance probability density functions for the freely-rotating model. The number of links is set to be 20. θ denotes the bond angle of the polymer chain. The continuous line corresponds to the case when $\theta = \pi/4$, and the dotted line to the case when $\theta = \pi/6$. In the simulations, the band width of the Fourier transform for $SE(3)$ and upper limit of integration with respect to frequency factor p are denoted as N_b and L_p, respectively, and $N_b = 12$ and $L_p = 50$ are used in the case of $\theta = \pi/4$, and $N_b = 16$ and $L_p = 60$ are used in the case of $\theta = \pi/6$.

Table 1
The comparison of the mean squares of the end-to-end distance

θ	From our model	From Eq. (66)
$\pi/4$	0.2502	0.2502
$\pi/6$	0.4687	0.4688

As for the next example, we consider the following potential energy

$$E(\alpha) = 11.8 + 7.66\cos\alpha + 4.64\cos 2\alpha + 8.8\cos 3\alpha \quad (67)$$

in units of kcal/mol, which is the torsional potential energy for n-butane.[3] In Fig. 5 is shown the torsional potential energy and exponential of the potential energy. This potential energy has three minima, gauche$^+$ (near $\alpha = 60°$), trans (near $\alpha = 180°$), and gauche$^-$ (near $\alpha = 300°$). This expression is one general form of torsional potential energy, which appears in many of polymer chains. In Fig. 6 is shown the resulting probability density function of end-to-end distance for this independent energy model. Looking at part (a), which is the case of the number of links being 16, one can see that the PDF is described more accurately as the band width for Fourier transform for $SE(3)$, N_b, and the upper limit for the integration with respect to the frequency factor p, L_p, get larger. Especially, $N_b = 7$ appears to be sufficient for describing the 'mountain' part of the PDF, but in order for the better 'tail' description, we need larger L_p such as 40 in that figure. In part (b), we show the PDFs for two different numbers of links. As one can expect, when the number of links is 128, the resulting end-to-end distance distribution becomes more concentrated toward the left side.

6.2. The interdependent pairwise energy chain model

We now demonstrate the interdependent pairwise energy model case. Among various polymers, from natural to artificial ones, polypeptide chains are of special interest these days. Hence we take polypeptides as examples of the interdependent

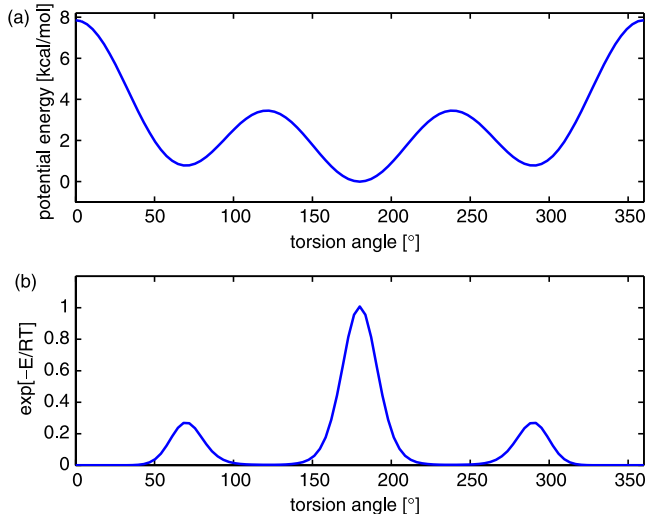

Fig. 5. The plot of torsional potential energy and its exponential used in the example of the independent energy model. (a) the torsional potential energy, (b) exponential of the torsional potential energy.

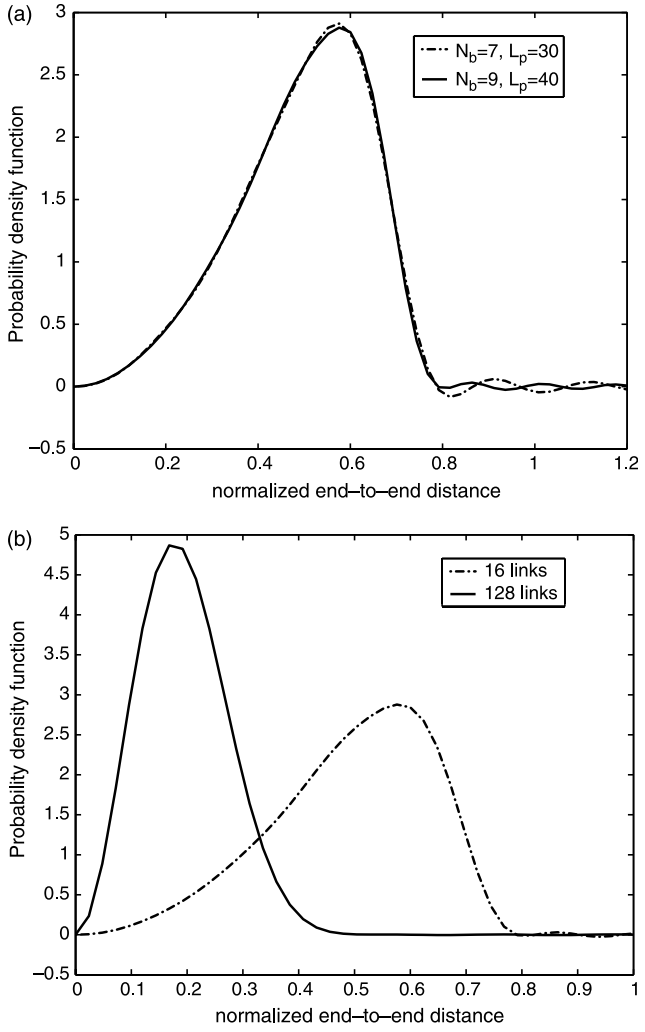

Fig. 6. Plot of the resulting end-to-end distance probability density functions for the independent energy model. All the monomer chains are assumed to be identical, with $\beta_0 = 75.1°$. (a) The plot of PDFs with different N_b and L_p. It shows that as N_b and L_p get larger, one can describe the corresponding PDF better. (b) The plot of PDFs with different number of links. From this figure, one can see that the PDF for chains with a large number of links gets more concentrated on the left side.

pairwise energy model. In order to generate end-to-end distance distributions, we need a 2D energy map. One can find a way to compute the energy map from first principles using Lennard–Jones potentials [1], or using MD simulations [42]. In general, MD simulation can be trusted more than just using Lennard–Jones potentials. However, in this paper, we generate Ramachandran-like plots as a probability distribution of torsional energy. The purpose of this example is to show that the proposed method can generate end-to-end distance distributions for any type of pairwise energy function, which justifies the usage of Ramachandran-like plots. We also utilize the simplified geometric model for a polypeptide chain based on the same reason above. All the information about bond angles and bond lengths of peptide units are borrowed from Ref. [41]. The hard-sphere contact distances are also borrowed from Ref. [41]. As for hard-sphere radius for residues (alanine and valine), we treat CH_3 to be approximately of the same size as C, from the fact that, according to [1], the radius of CH_2 are greater than that of C by only 0.15 Å. Fig. 7 shows the Ramachandran-like plot for each dipeptide, alanine and valine. We treat these maps as the exponential of the torsional potential energy, $e^{E_{i,i+1}(\alpha_i,\alpha_{i+1})/k_B T}$ and the height of each allowed (gray) region is set at 2. In Fig. 8 is shown the resulting end-to-end distance PDF for a polyalanine chain with 16 residues.

As mentioned earlier, the small value of bandwidth B, which is for the classical Fourier series approximation of the exponential of the torsional potential energy, is good enough for describing the corresponding PDF, and so is the bandwidth N_b for Fourier transform matrix for $SE(3)$. In order to verify our method, we performed MC simulation for polyalanine with 16 residues. In the MC sampling, since the probabilities within the allowable region are the same, we randomly select a set of pairs of ϕ, ψ angles to generate as many conformations of a polyalanine chain as possible given computing/time constraints. In practice, we generate 10^6 conformations to generate the histogram of the end-to-end distance. In Fig. 9 is shown

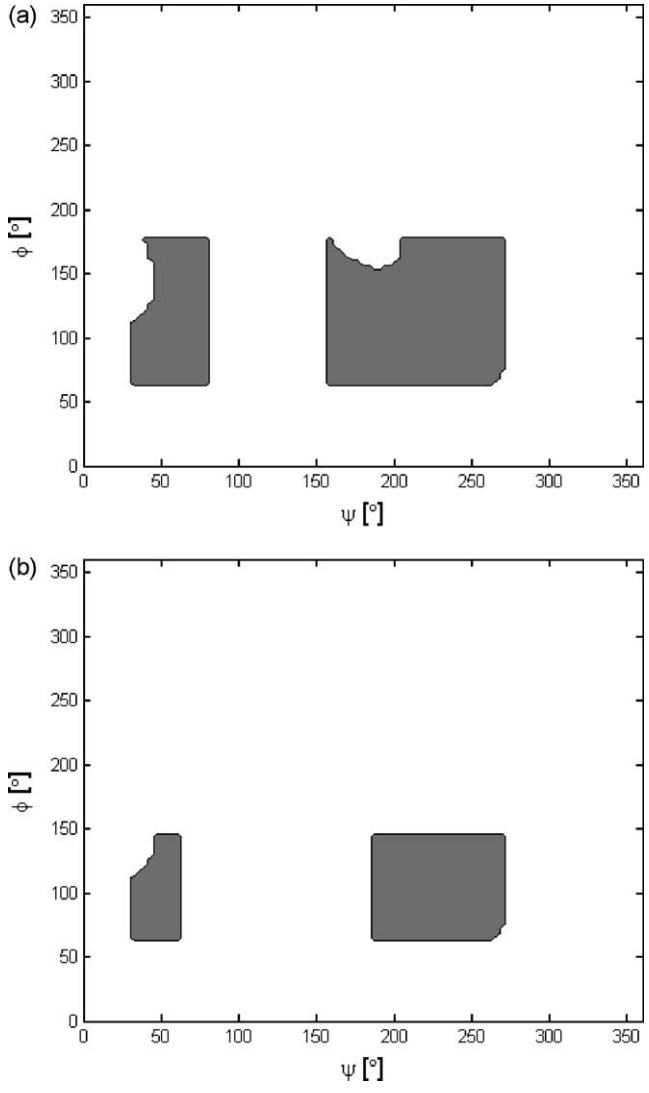

Fig. 7. The Ramachandran-like plots of approximate models for dipeptide used in the numerical examples. (a) Alanine dipeptide. (b) Valine dipeptide. Gray represents allowed regions.

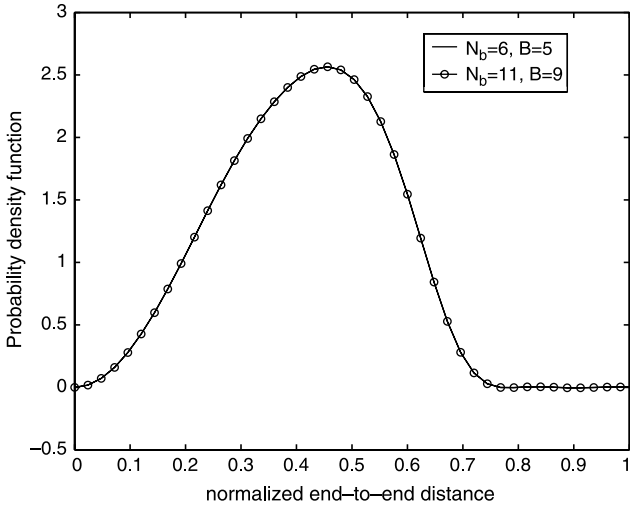

Fig. 8. Plot of the resulting end-to-end distance probability density functions for the interdependent energy model. This corresponds to the case of a polyalanine chain. The number of residues is 16 in both cases. It shows that lower value of N_b and B is good enough to describe the corresponding PDF of the end-to-end distance.

the comparison between the result from MC simulation and that from our method. One can observe good agreement between both the results, though MC is not able to give the exact PDF in this situation. We believe this happens for two reasons: first, the conformational space is too large to allow for sufficient sampling; and second, random number generators in software packages (in this case Matlab) are not truly random, and this lack of true randomness becomes noticable as the number of samples becomes very large. Also in Fig. 10 one can see that the PDFs corresponding to different energy becomes different. Looking at Fig. 11, which shows the resulting end-to-end distance PDF for polyalanine chains with different number of links, one can also see that when the number of links is 128, the resulting end-to-end distance distribution becomes more concentrated toward the left side.

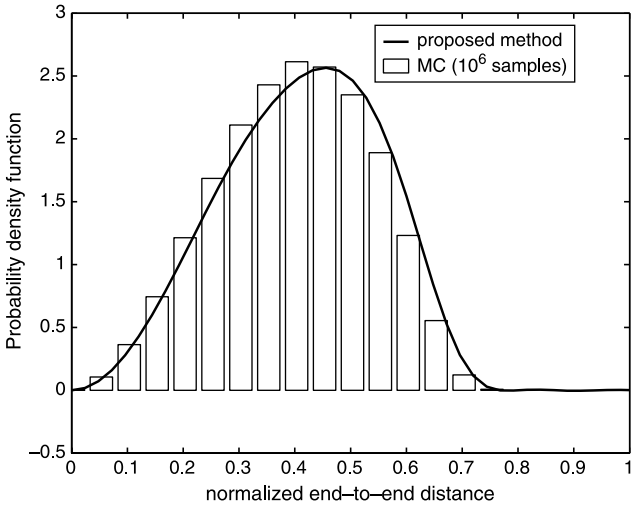

Fig. 9. Comparison of the result from the proposed method with MC sampling. In MC sampling, 10^6 conformations are randomly selected to generate histogram which is equivalent to PDF of the end-to-end distance. Both the cases correspond to polyalanine with 16 residues.

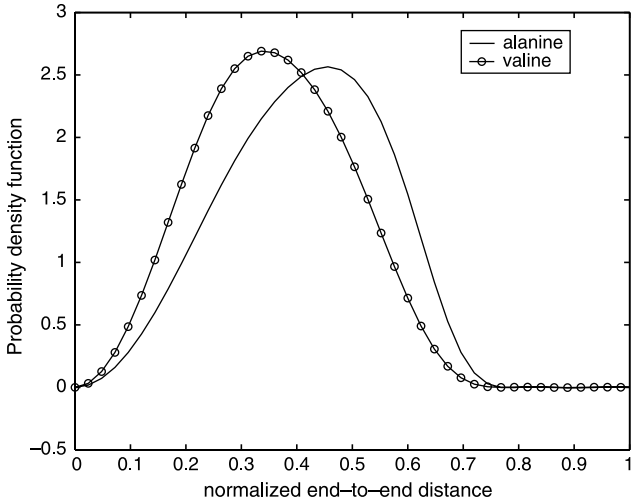

Fig. 10. Comparison of two different polypeptides with the same number of links. In this figure, the number of residues is set to be 16.

6.3. The comparison with other chain models

In this section, we demonstrate the comparison between the end-to-end distance distribution from other chain models distribution and that from the proposed method.

One can find the formula of the spatial distribution for the Gaussian chain model and the freely-jointed chain model in Refs. [1,3]. Following the notation in Ref. [1], the end-to-end distribution for the Gaussian chain is expressed as

$$W(\mathbf{r}) = \left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2} \exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right) \tag{68}$$

and for the N-link freely-jointed chain model

$$W(\mathbf{r}) = \frac{1}{2\pi^2 r}\int_0^\infty \sin(qr)\left[\frac{\sin(ql)}{ql}\right]^N q\mathrm{d}q \tag{69}$$

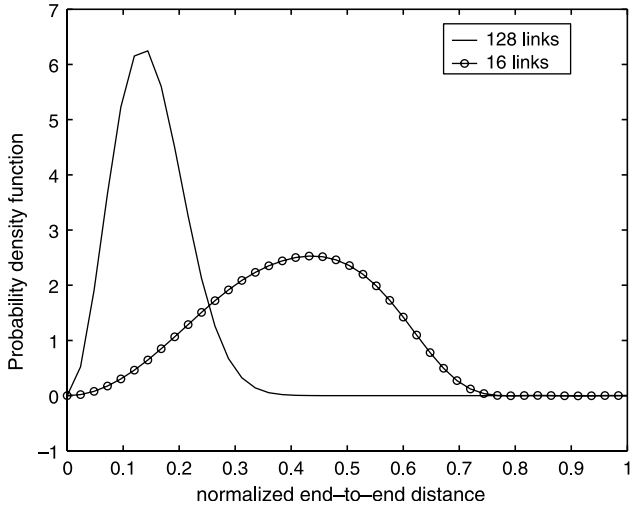

Fig. 11. Plot of the resulting end-to-end distance probability density functions for the polyalanine chain for different number of links. As one can see, the larger the number of links, the more concentrated to the left side the resulting plot becomes.

where **r** denotes the end-to-end vector and r denotes the end-to-end distance. $\langle r^2 \rangle$ corresponds to the mean square of the end-to-end distance. From these functions, one can obtain the PDFs of the end-to-end distance as

$$f(r) = 4\pi r^2 W(\mathbf{r}) \tag{70}$$

First we compare the freely-jointed chain model and the Gaussian chain model with the freely-rotating chain model from the proposed method. In this example, the number of links is 20 and the bond angle for the freely-rotating chain model is $\pi/4$. Referring to Table 1, we find that $\langle r^2 \rangle = 0.2502$. We substitute this value in Eq. (68) to obtain the PDF of the end-to-end distance for the Gaussian chain. When we use $\langle r^2 \rangle = NL^2$ in Eq. (68), the resulting PDF of the Gaussian chain model becomes in-distinguishable from that of freely-jointed chain model. In Fig. 12, we show the result of the comparison. As one can see, the freely-rotating chain model generates a different PDF than the freely-jointed model with the same number of links. Also, the PDF of the end-to-end distance of the Gaussian chain model with the same mean squared end-to-end distance as the freely-rotating chain model is different than that of the freely-rotating model.

Next, we compare the Gaussian chain model with our examples in previous sections. That is to say, we take the freely-rotating chain model, the independent energy chain model, and the interdependent pairwise energy chain model to compare with the Gaussian chain model when the length of a chain is large. Particularly, in this example we take the number of links is 128 and the bond angle is $\pi/4$ for the freely-rotating chain model. As for independent energy chain model, we utilize the same example in the first subsection. Also we take the polyalanine chain in the second subsection as an example of the interdependent pairwise energy chain model. As for the Gaussian chain model, we first compute $\langle r^2 \rangle$ from the results of the proposed method, and then obtain the Gaussian chain distribution from Eqs. (68) and (70). In Fig. 13 we show

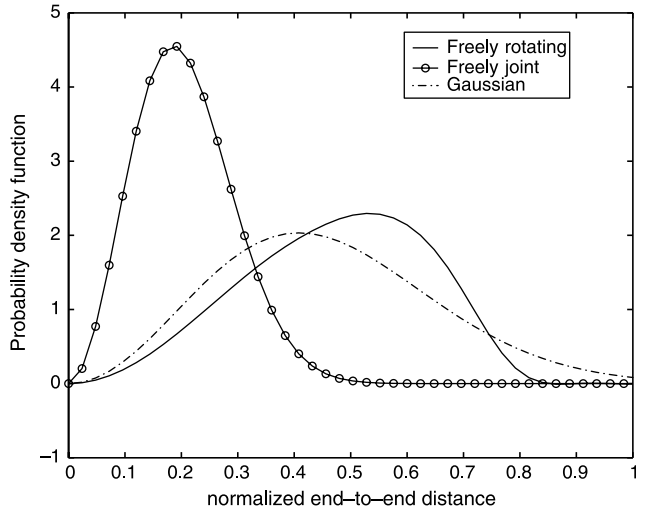

Fig. 12. Comparison among the Gaussian chain, the freely-jointed chain, and the freely-rotating chain model. The number of links is set to be 20 in the freely-jointed and the freely rotating chain models. As one can see, the resulting PDFs are different according to each model.

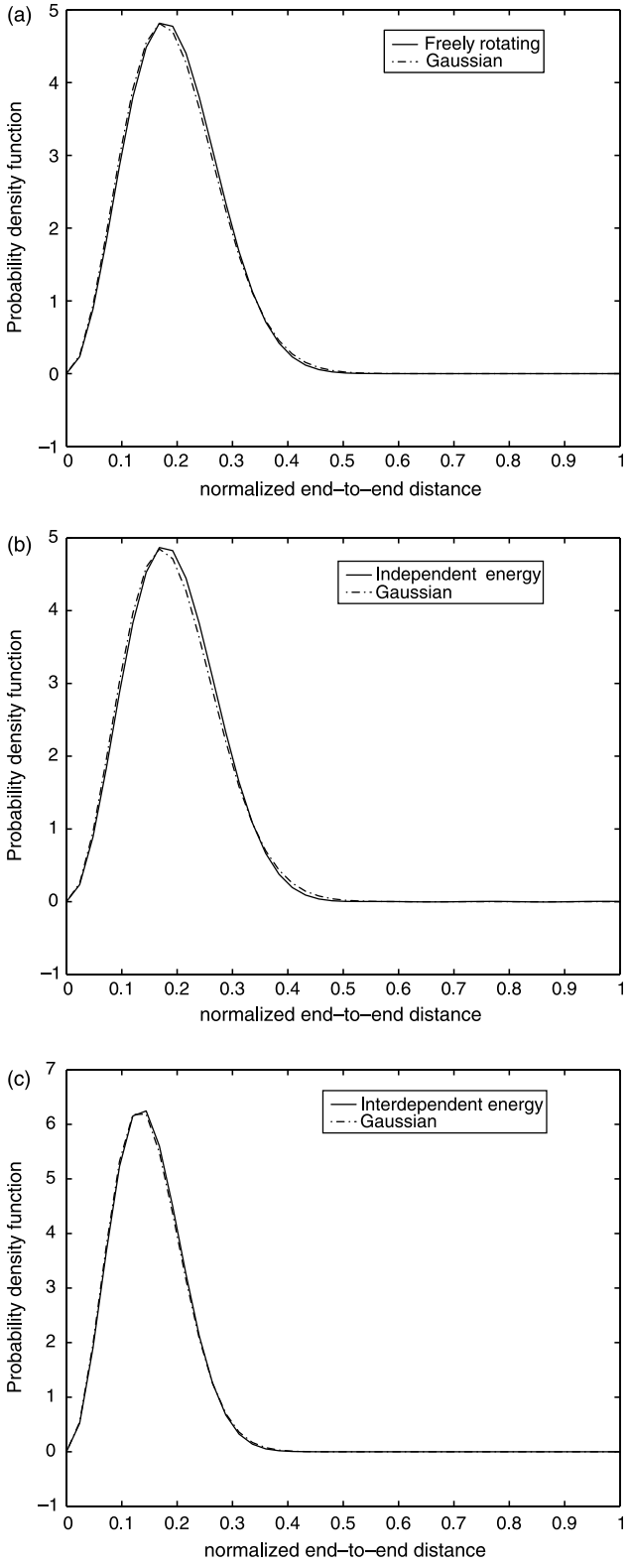

Fig. 13. (a) Comparison between the Gaussian chain and the freely-rotating chain model. (b) Comparison between the independent energy chain model and the Gaussian chain model. (c) Comparison between the interdependent pairwise energy chain model (polyalanine) and the Gaussian chain model. In all cases, the number of links is set to be 128. It shows that when the length of polymer chain is very long, then the freely-rotating chain model, the independent energy chain model, and the interdependent pairwise energy chain model all become the same as the Gaussian chain.

the results of comparison between each of the three different chain models and the Gaussian chain model. As one can see, these figures verify that all classical linear chain models behave as the Gaussian chain as the length of a chain becomes very large.

7. Conclusions

In this paper, we have presented a unified method to generate the probability density function of the end-to-end distance for discrete-chain phantom polymer models. Our method is based on previous work which utilizes the generalized convolution and extends it by employing the Fourier transform for $SE(3)$. The proposed method is general enough to be applied to any of the classical phantom polymer chain models. We have formulated the proposed method for three different discrete polymer chain models: the freely-rotating model, the independent torsional energy model, and interdependent pairwise energy model. We have also developed an efficient implementation method, particularly for the interdependent pairwise energy model, by approximating the exponential of the torsional potential function with the classical Fourier series. We have demonstrated the versatility of the proposed method by numerical examples. We expect that this method, which can generate both one-dimensional marginal PDFs and multi-dimensional PDFs (e.g. PDF for end-to-end distance and orientation of the distal end of a polymer chain), can be useful in a wider range of conformational studies on polymer chains, including artificial polymer chains and natural ones such as polypeptides and single-stranded RNA molecules.

Acknowledgements

This work was performed under NIH Grant R01GM075310.

References

[1] Flory PJ. Statistical mechanics of chain molecules. New York: Wiley-Interscience; 1969.
[2] Dill KA. Protein Sci 1999;8:1166.
[3] Mattice WL, Suter UW. Conformational theory of large molecules. New York: Wiley; 1994.
[4] Lapidus LJ, Steinbach PJ, Eaton WA, Szabo A, Hofrichter J. J Phys Chem B 2002;106:11628.
[5] Cannavacciuolo L, Pedersen JS. J Chem Phys 2004;120:8862.
[6] Winkler RG. J Chem Phys 2003;118:2919.
[7] Samuel J, Sinha S. Phys Rev E 66 050801 (R) (2002).
[8] Mark JE, Curro JG. J Chem Phys 1983;79:5705.
[9] Curro JG, Mark JE. J Chem Phys 1984;80:4521.
[10] Yuan W, Kloczkowski A, Mark JE, Sharaf MA. J Polym Sci, Part B: Polym Phys 1996;34:1647.
[11] Sharaf MA, Mark JE. Polymer 2002;43:643.
[12] Sharaf MA, Mark JE. Polymer 2004;45:3943.
[13] Chirikjian GS, Kyatkin AB. Engineering applications of noncommutative harmonic analysis. New York: CRC Press; 2001.
[14] Marko JF, Siggia ED. Macromolecules 1994;27:981.
[15] Haijin Z, Zhong-Can O. Phys Rev E 1998;58:4816.
[16] des Cloizeaux J, Jannink G. Polymers in solution: their modelling and structure. Oxford: Clarendon Press; 1990.
[17] Yamakawa H. Helical wormlike chains in polymer solutions. Berlin: Springer; 1997.
[18] Hamprecht B, Janke W, Kleinert H. Phys Lett A 2004;330:254.
[19] Bhattacharjee JK, Thirumalai D, Bryngelson JD. cond-mat/9705200.
[20] Stepanow S, Schütz M. Europhys Lett 2002;60:546.
[21] Zhou H. J Phys Chem B 2001;105:6763.
[22] Chirikjian GS, Wang Y. Phys Rev E 2000;62:880.
[23] Zhou Y, Chirikjian GS. J Chem Phys 2003;119:4962.
[24] Kostrowicki J, Scheraga HA. Comput Polym Sci 1995;5:47.
[25] Freire JJ, Rodrigo MM. J Chem Phys 1980;72:6376.
[26] Freire J, Fixman M. J Chem Phys 1978;69:634.
[27] Fixman M, Skolnick J. J Chem Phys 1976;65:1700.
[28] Fixman M. J Chem Phys 1973;58:1559.
[29] Rubio AM, Freire JJ. Macromolecules 1989;22:333.
[30] He S, Scheraga HA. J Chem Phys 1998;108:271.
[31] Plimpton S, Hendrickson B. J Comput Chem 1996;17:326.
[32] Lyulin AV, Dünweg B, Borisov OV, Darinskii AA. Macromolecules 1999;32:3264.
[33] Kloczkowski A, Sen TZ, Sharaf MA. Polymer 2005;46:4373.
[34] Marcelo G, Tarazona MP, Saiz E. Polymer 2004;45:1321.
[35] Chirikjian GS. Comput Theor Polym Sci 2001;11:143.
[36] Miller W. Commun Pure Appl Math 1964;17:527.
[37] Vilenkin NJ, Klimyk AU. Representation of Lie groups and special function. vol. 1. Dordrecht: Kluwer Academic; 1991.
[38] Gelfand IM, Minlos RA, Shapiro ZY. Representations of the rotation and Lorentz groups and their applications. New York: Macmillan; 1963.
[39] Varshalovich NJ, Moskalev AN, Khersonskii VK. Quantum theory of angular momentum. NJ: World Scientific Publishing; 1988.
[40] Branden C, Tooze J. Introduction to protein structure. 2nd ed. New York: Garland Publishing; 1999.
[41] Pappu RV, Srinivasan R, Rose GD. Proc Natl Acad Sci 2000;97:12565.
[42] Bahar I, Zuniga I, Dodge R, Mattice WL. Macromolecules 1991;24:2986.