JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 22, Number 9, 2015

Pp. 787–805
DOI: 10.1089/cmb.2015.0115

Research Articles

Computational Analysis of SAXS Data Acquisition

HUI DONG,[1] JIN SEOB KIM,[2] and GREGORY S. CHIRIKJIAN[2,*]

ABSTRACT

Small-angle x-ray scattering (SAXS) is an experimental biophysical method used for gaining insight into the structure of large biomolecular complexes. Under appropriate chemical conditions, the information obtained from a SAXS experiment can be equated to the pair distribution function, which is the distribution of distances between every pair of points in the complex. Here we develop a mathematical model to calculate the pair distribution function for a structure of known density, and analyze the computational complexity of these calculations. Efficient recursive computation of this forward model is an important step in solving the inverse problem of recovering the three-dimensional density of biomolecular structures from their pair distribution functions. In particular, we show that integrals of products of three spherical-Bessel functions arise naturally in this context. We then develop an algorithm for the efficient recursive computation of these integrals.

Key words: fast evaluation of integrals, pair distribution function, recurrence relations, SAXS, spherical-Bessel functions.

1. INTRODUCTION

STRUCTURE DETERMINATION OF biological macromolecules is important in biology because of the close relationship between shape and function. Many experimental methods for determining the structure of macromolecules exist in the field of biophysics. For single proteins and small complexes composed of one or more protein molecule and nucleic acids, macromolecular x-ray crystallography (MX) and nuclear magnetic resonance (NMR) have demonstrated great success in producing more than 100,000 structures reported in the Protein Data Bank (PDB) (Berman et al., 2000). Now, as the fields of biology and biophysics turn more toward the study of cellular phenomena, the role of understanding the structure and motion of large biomolecular complexes is becoming critical. These complexes can be thought of as the intellectual bridge between single-molecule studies and cellular phenomena.

The experimental methods used to study large complexes are very different than in the single-molecule case. Two of the main methods used in this context are small-angle x-ray scattering (SAXS) and cryo-electron microscopy (cryo EM). In this article we focus on the relationship between the shape of a biomolecular complex, and the data obtained from SAXS experiments. In the process, we uncover a computational problem and its solution: how to efficiently compute integrals of products of three spherical-Bessel functions.

[1]Fuzhou University, School of Mechanical Engineering and Automation, Fuzhou, China.
[2]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, Maryland.

2. BASIC FORMULATION IN SAXS

Let a biomolecular complex be viewed as a solid body, B. The key quantities that are obtained from SAXS experiments include the scattering density and the pair distance distribution function, respectively denoted as $I_B(p)$ and $p_B(r)$ throughout this article. Here p and r denote the radial coordinate in the reciprocal Fourier space and the radial coordinate in the real-space spherical coordinates system, respectively. The scattering density profile is usually obtained from SAXS experiments. Then one can compute the pair distance distribution (Svergun et al., 2013) as

$$p_B(r) = \frac{r^2}{2\pi^2}\int_0^\infty I_B(p)\frac{\sin(pr)}{pr}p^2 dp. \tag{1}$$

And the reverse is written as

$$I_B(p) = 4\pi\int_0^\infty p_B(r)\frac{\sin(pr)}{pr}dr. \tag{2}$$

The pair distance distribution function is the distribution of distances between every pair of points in the biomolecular complex, which contains important geometric information linked to the experimental data. In this article, we seek to develop an efficient method to compute Equation (1). The points in the complex body can be either discrete in the sense that they represent the coordinates of each atom, or continuous variables of space in the body when treated as a continuum, which is a very similar concept as used in describing the radial distribution function of a fluid. In this work, we employ the latter continuum description of body. To this end, we first define the characteristic function as

$$\chi_B(\mathbf{x}) \doteq \begin{cases} 1 & \text{if } \mathbf{x} \in B \\ 0 & \text{if } \mathbf{x} \notin B \end{cases} \tag{3}$$

which is interpreted as the unit body density. Here $\mathbf{x} \in \mathbb{R}^3$ denotes the position in 3D Euclidean space. A number of geometric quantities can be computed from it. For example, the volume of the body is computed as

$$V(B) = \int_{\mathbb{R}^3}\chi_B(\mathbf{x})\,d\mathbf{x} = \int_B 1 d\mathbf{x}$$

where $d\mathbf{x} = dxdydz$ is the usual integration measure for $\mathbb{R}^3$. When we use the spherical coordinates $\mathbf{x} = r\mathbf{u}$ where $\mathbf{u} \in \mathbb{S}^2$ is a unit vector of each point in the surface of the unit sphere, then this measure becomes $d\mathbf{x} = r^2 dr d\mathbf{u} = r^2 dr \sin\theta\, d\theta d\phi$ where θ and ϕ denote the polar and azimuthal angles.

There are several ways to compute the intensity and pair distance distribution functions. One way includes the so-called Debye approximation (Debye, 1915) where

$$I_B(p) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\chi_B(\mathbf{x}_1)\chi_B(\mathbf{x}_2)\frac{\sin(p\|\mathbf{x}_1-\mathbf{x}_2\|)}{p\|\mathbf{x}_1-\mathbf{x}_2\|}d\mathbf{x}_1 d\mathbf{x}_2 \tag{4}$$

and

$$p_B(r) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\chi_B(\mathbf{x}_1)\chi_B(\mathbf{x}_2)\delta(r-\|\mathbf{x}_1-\mathbf{x}_2\|)\,d\mathbf{x}_1 d\mathbf{x}_2. \tag{5}$$

Another way is to compute the pair distance distribution function using the so-called self-convolution as (Svergun et al., 2013; Kratky, 1963)

$$\begin{aligned} p_B(r) &= \frac{r^2}{4\pi}\int_{\mathbb{S}^2}\chi_B(\mathbf{x}) * \chi_B(-\mathbf{x})\,d\mathbf{u} \\ &= \frac{r^2}{4\pi}\int_{\mathbb{S}^2}d\mathbf{u}\int_{\mathbb{R}^3}\chi_B(\mathbf{y})\chi_B(\mathbf{y}+r\mathbf{u})\,d\mathbf{y} \end{aligned} \tag{6}$$

where * denotes the convolution in $\mathbb{R}^3$. If we define $\chi_{\overline{B}}(\mathbf{x}) = \chi_B(-\mathbf{x})$ (i.e., body B reflected through the origin as its mass center), then it follows that

$$p_B(r) = \frac{r^2}{4\pi}\int_{\mathbb{S}^2}\left(\chi_B * \chi_{\overline{B}}\right)(r\mathbf{u})\,d\mathbf{u} = \frac{r^2}{4\pi}\int_{\mathbb{S}^2}\int_{\mathbb{R}^3}\chi_B(\mathbf{y})\,\chi_{\overline{B}}(r\mathbf{u}-\mathbf{y})\,d\mathbf{y}\,d\mathbf{u}. \tag{7}$$

This alternative way does not involve any Dirac delta functions that are singular and has lower dimension of the integral than the expression (5).

We note that Equation (7) is invariant under translational shifts of the coordinate system in which the body is defined. Let $\chi_{B+\mathbf{t}}(\mathbf{x}) \doteq \chi_B(\mathbf{x}-\mathbf{t})$. Then

$$\begin{aligned}\left(\chi_{B+\mathbf{t}} * \chi_{\overline{B+\mathbf{t}}}\right)(\mathbf{x}) &= \int_{\mathbb{R}^3}\chi_{B+\mathbf{t}}(\mathbf{y})\,\chi_{\overline{B+\mathbf{t}}}(\mathbf{x}-\mathbf{y})\,d\mathbf{y}\\ &= \int_{\mathbb{R}^3}\chi_{B+\mathbf{t}}(\mathbf{y})\,\chi_{\overline{B}-\mathbf{t}}(\mathbf{x}-\mathbf{y})\,d\mathbf{y} = \int_{\mathbb{R}^3}\chi_B(\mathbf{y}-\mathbf{t})\,\chi_{\overline{B}}(\mathbf{x}-\mathbf{y}+\mathbf{t})\,d\mathbf{y}.\end{aligned}$$

Let $\mathbf{z} = \mathbf{y}-\mathbf{t}$, then

$$\begin{aligned}\left(\chi_{B+\mathbf{t}} * \chi_{\overline{B+\mathbf{t}}}\right)(\mathbf{x}) &= \int_{\mathbb{R}^3}\chi_B(\mathbf{y}-\mathbf{t})\,\chi_{\overline{B}}(\mathbf{x}-\mathbf{y}+\mathbf{t})\,d\mathbf{y}\\ &= \int_{\mathbb{R}^3}\chi_B(\mathbf{z})\chi_{\overline{B}}(\mathbf{x}-\mathbf{z})\,d\mathbf{z} = \left(\chi_B * \chi_{\overline{B}}\right)(\mathbf{x}).\end{aligned}$$

Also, since convolution on $\mathbb{R}^n$ is commutative, the roles of B and $\overline{B}$ can be reversed. Finally, integration over the sphere is invariant under rotation. As a result, $p_B(r)$ is invariant under the action of the full Euclidean group, $E(3)$, which includes translations, rotations, and reflections.

As we will show later, the spherical-Bessel functions are a natural basis to express $p_B(r)$. These can be viewed as being defined by a generating function (Abramowitz and Stegun, 1972)

$$\frac{\cos\sqrt{x^2-2xt}}{x} = \sum_{l=0}^{\infty}\frac{t^l}{l!}j_{l-1}(x),$$

and they have the Rodrigues formula (which in this case is called Rayleigh's formula) (Andrews et al., 1999; Arfken and Weber, 2010)

$$j_l(x) = (-1)^l x^l\left(\frac{1}{x}\frac{d}{dx}\right)^l\frac{\sin x}{x}.$$

3. CONVOLUTION WITH FOURIER TRANSFORM ON $\mathbb{R}^3$ AND PAIR DISTANCE DISTRIBUTION

In this section we discuss the spherical Fourier transform in 3D Euclidean space, which is a prerequisite for the recursive formula to computing the quantity of interest in SAXS. We then derive a new expression for the pair distribution function involving the integral of the product of three spherical-Bessel functions.

3.1. Spherical Fourier transform in $\mathbb{R}^3$

In spherical coordinates, it is known that

$$e^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\mathrm{i}^l j_l(pr)\,Y_l^m(\mathbf{u})\overline{Y_l^m(\mathbf{u}_p)} \tag{8}$$

and

$$e^{-\mathrm{i}\mathbf{p}\cdot\mathbf{x}} = \overline{e^{\mathrm{i}\mathbf{p}\cdot\mathbf{x}}} = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(-\mathrm{i})^l j_l(pr)\overline{Y_l^m(\mathbf{u})}Y_l^m(\mathbf{u}_p) \tag{9}$$

where $Y_l^m(\mathbf{u})$ denotes the spherical harmonics function of degree l and order m, which together with the spherical-Bessel function is important in describing shapes in spherical coordinates. However, as we

will see, this function does not contribute that much to computing $p_B(r)$ due to the spherical averaging effect as in Equation (6). We use i as the complex imaginary unit, $\sqrt{-1}$, to distinguish it from an index i. Here $\mathbf{x} = r\mathbf{u}$ is described by the spherical coordinates (r, θ, ϕ) (i.e., $\mathbf{u} \in \mathbb{S}^2$), and $\mathbf{p} = p\mathbf{u}_p$ is described by (p, θ_p, ϕ_p).

Then the Fourier transform in spherical coordinates is expressed as

$$\begin{aligned}\mathcal{F}_{\mathbb{R}^3}[f(r\mathbf{u})] \doteq \hat{f}(\mathbf{p}) &= \int_{\mathbb{R}^3} f(r\mathbf{u}) e^{-\mathrm{i}\mathbf{p}\cdot\mathbf{x}} d\mathbf{x} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} 4\pi(-\mathrm{i})^l \breve{f}_l^m(p)\, Y_l^m(\mathbf{u}_p).\end{aligned} \tag{10}$$

Here

$$\breve{f}_l^m(p) \doteq \int_{r=0}^{\infty} f_l^m(r)\, j_l(pr)\, r^2 dr \tag{11}$$

(which is the spherical Hankel transform) and

$$f_l^m(r) \doteq \int_{\mathbb{S}^2} f(r\mathbf{u})\, \overline{Y_l^m(\mathbf{u})}\, d\mathbf{u} = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} f(r, \theta, \phi)\, \overline{Y_l^m(\theta, \phi)}\, \sin\theta d\theta d\phi \tag{12}$$

(which is a spherical-harmonic transform).

The inverse formula is written as

$$f(r\mathbf{u}) = \frac{1}{2\pi^2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mathrm{i}^l \left(\int_{p=0}^{\infty} 4\pi(-\mathrm{i})^l \breve{f}_l^m(p)\, j_l(pr)\, p^2 dp \right) Y_l^m(\mathbf{u}) \tag{13}$$

where, after combining Equations (11) and (12),

$$\breve{f}_l^m(p) = \int_{r'} \int_{\mathbb{S}^2} f(r'\mathbf{u}') \overline{Y_l^m(\mathbf{u}')}\, j_l(pr') r'^2 dr' d\mathbf{u}'. \tag{14}$$

3.2. Convolution and pair distance distribution function

The spherical Fourier transform of the self-convolution is written as

$$\mathcal{F}_{\mathbb{R}^3}[\chi_B * \chi_{\overline{B}}] = \widehat{(\chi_B * \chi_{\overline{B}})}(p\mathbf{u}_p) = \hat{\chi}_B(p\mathbf{u}_p)\, \hat{\chi}_{\overline{B}}(p\mathbf{u}_p)$$

and the inverse formula is written as

$$\left(\chi_B * \chi_{\overline{B}}\right)(r\mathbf{u}) = \frac{1}{2\pi^2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \int_{p=0}^{\infty} \int_{\mathbf{u}_p \in \mathbb{S}^2} \mathrm{i}^l \widehat{\left(\chi_B * \chi_{\overline{B}}\right)}(p\mathbf{u}_p)\, j_l(pr) Y_l^m(\mathbf{u}) \overline{Y_l^m(\mathbf{u}_p)} p^2 dp d\mathbf{u}_p$$

where $\hat{\chi}_B(p\mathbf{u}_p)$ and $\hat{\chi}_{\overline{B}}(p\mathbf{u}_p)$ are obtained by Equation (10). Finally, the pair distance distribution function is calculated as

$$p_B(r) = \frac{r^2}{4\pi} \int_{\mathbb{S}^2} \left(\chi_B * \chi_{\overline{B}}\right)(r\mathbf{u})\, d\mathbf{u}. \tag{15}$$

What we encounter first is

$$\int_{\mathbf{u}_p} \overline{Y_l^m(\mathbf{u}_p)} Y_{l'}^{m'}(\mathbf{u}_p) Y_{l''}^{m''}(\mathbf{u}_p)\, d\mathbf{u}_p.$$

First, the product of the second and the third terms is expressed as (Varshalovich et al., 1988)

$$Y_{l'}^{m'}(\mathbf{u}_p) Y_{l''}^{m''}(\mathbf{u}_p) = \sum_{L,M} \sqrt{\frac{(2l'+1)(2l''+1)}{4\pi(2L+1)}} C_{l',0,l'',0}^{L,0} C_{l',m',l'',m''}^{L,M} Y_L^M(\mathbf{u}_p)$$

where $C^{l,m}_{l_1,m_1;l_2,m_2}$ denotes the Clebsch–Gordan coefficients (or CGCs) (Chirikjian, 2012; Varshalovich et al., 1988). Then use the fact that

$$\int_{\mathbf{u}_p} \overline{Y_l^m(\mathbf{u}_p)}\, Y_L^M(\mathbf{u}_p)\, d\mathbf{u}_p = \delta_{L,l}\,\delta_{M,m}.$$

Hence

$$\int_{\mathbf{u}_p} \overline{Y_l^m(\mathbf{u}_p)}\, Y_{l'}^{m'}(\mathbf{u}_p)\, Y_{l''}^{m''}(\mathbf{u}_p)\, d\mathbf{u}_p = \sqrt{\frac{(2l'+1)(2l''+1)}{4\pi(2l+1)}} C^{l,0}_{l',0,l'',0} C^{l,m}_{l',m',l'',m''}$$

Note that a property of CGCs is that they are only nonzero in the case when $m'' = m - m'$.

Also we will average over $\mathbf{u} \in \mathbb{S}^2$ to obtain the expression of $p_B(r)$, we use the following

$$\int_{\mathbf{u}\in\mathbb{S}^2} Y_l^m(\mathbf{u})\, d\mathbf{u} = \sqrt{4\pi}\delta_{l,0}\delta_{m,0}.$$

Then the associated Clebsch–Gordan coefficients become simplified as

$$C^{0,0}_{l',0,l'',0} = \delta_{l',l''}\,(-1)^{-l'}\sqrt{\frac{1}{2l'+1}}$$

and

$$C^{0,0}_{l',m',l',-m'} = \delta_{l',l''}(-1)^{l'-m'}\sqrt{\frac{1}{2l'+1}}.$$

Hence

$$\begin{aligned}\int_{\mathbf{u}} (\chi_B * \chi_{\overline{B}})(r\mathbf{u})\, d\mathbf{u} = \frac{1}{2\pi^2}\sum_{l',m'}(4\pi)^2(-1)^{l'} \int_p \int_{r'} \int_{\mathbf{u}'} \chi_B(r'\mathbf{u}')\overline{Y_{l'}^{m'}(\mathbf{u}')}\, d\mathbf{u}' j_{l'}(pr')\, r'^2 dr' \\ \times \int_{r''}\int_{\mathbf{u}''} \chi_{\overline{B}}(r''\mathbf{u}'')\overline{Y_{l'}^{-m'}(\mathbf{u}'')}\, d\mathbf{u}'' j_{l'}(pr'')\, r''^2 dr''(-1)^{-m'} j_0(pr)\, p^2 dp.\end{aligned}$$

Note that

$$\overline{Y_{l'}^{-m'}(\mathbf{u}')} = (-1)^{-m'} Y_{l'}^{m'}(\mathbf{u}')$$

Then

$$\begin{aligned}\int_{\mathbf{u}} (\chi_B * \chi_{\overline{B}})(r\mathbf{u})\, d\mathbf{u} = \frac{1}{2\pi^2}\sum_{l',m'}(4\pi)^2(-1)^{l'} \int_p \int_{r'} \int_{\mathbf{u}'} \chi_B(r'\mathbf{u}')\overline{Y_{l'}^{m'}(\mathbf{u}')} j_{l'}(pr')\, r'^2 dr' d\mathbf{u}' \\ \times \int_{r''}\int_{\mathbf{u}''} \chi_{\overline{B}}(r''\mathbf{u}'') Y_{l'}^{m'}(\mathbf{u}'') j_{l'}(pr'') r''^2 dr'' d\mathbf{u}''\; j_0(pr)\, p^2 dp.\end{aligned}$$

Moreover, one can find that

$$\begin{aligned}&\int_{\mathbb{R}^3} \chi_{\overline{B}}(r\mathbf{u}) j_{l'}(pr) Y_{l'}^{m'}(\mathbf{u}) r^2 dr d\mathbf{u} \\ &= \int_{\mathbb{R}^3} \chi_B(-r\mathbf{u}) j_{l'}(pr) Y_{l'}^{m'}(\mathbf{u}) r^2 dr d\mathbf{u} \\ &= \int_{\mathbb{R}^3} \chi_B(r\mathbf{u}') j_{l'}(pr) Y_{l'}^{m'}(-\mathbf{u}') r^2 dr d\mathbf{u}' \\ &= (-1)^{l'} \int_{\mathbb{R}^3} \chi_B(r\mathbf{u}') j_{l'}(pr) Y_{l'}^{m'}(\mathbf{u}')\, r^2 dr d\mathbf{u}' \qquad (16)\end{aligned}$$

by using the fact that $Y_l^m(-\mathbf{u}) = (-1)^l Y_l^m(\mathbf{u})$ (Varshalovich et al., 1988).

Then

$$
\begin{aligned}
p_B(r) &= \frac{r^2}{4\pi}\int_{\mathbb{S}^2}\left(\chi_B * \chi_{\overline{B}}\right)(r\mathbf{u})\,d\mathbf{u} \\
&= \frac{r^2}{2\pi^2}\int_{p=0}^{\infty}\left\{4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\left(\int_{r'}\int_{\mathbb{S}^2}\chi_B(r'\mathbf{u}')\,Y_{l'}^{m'}(\mathbf{u}')\,j_{l'}(pr')r'^2dr'd\mathbf{u}'\right)\right. \\
&\quad\left.\times\left(\int_{r'}\int_{\mathbb{S}^2}\chi_B(r'\mathbf{u}')\overline{Y_{l'}^{m'}(\mathbf{u}')}\,j_{l'}(pr')\,r'^2dr'd\mathbf{u}'\right)\right\}j_0(pr)\,p^2dp \\
&= \frac{r^2}{2\pi^2}\int_{p=0}^{\infty}\left\{4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\left(\int_{r'}\int_{\mathbb{S}^2}\chi_B(r'\mathbf{u}')\,Y_{l'}^{m'}(\mathbf{u}')\,j_{l'}(pr')\,r'^2dr'd\mathbf{u}'\right)\right. \\
&\quad\left.\times\left(\overline{\int_{r'}\int_{\mathbb{S}^2}\chi_B(r'\mathbf{u}')\,Y_{l'}^{m'}(\mathbf{u}')\,j_{l'}(pr')\,r'^2dr'd\mathbf{u}'}\right)\right\}j_0(pr)\,p^2dp.
\end{aligned}
\tag{17}
$$

Hence we define another key quantity in the formulation as

$$
(m_B)_{l,m}(p) \doteq \int_r (\chi_B^{\circ})_{l,m}(r)\,j_l(pr)\,r^2dr \tag{18}
$$

where

$$
(\chi_B^{\circ})_{l,m}(r) = \int_{\mathbb{S}^2}\chi_B(r\mathbf{u})Y_l^m(\mathbf{u})\,d\mathbf{u}. \tag{19}
$$

Usually in SAXS modeling, first Equation (19) is computed numerically for sampled values of r and this is substituted into Equation (18). Then the scattering density is computed numerically for sampled values of p as

$$
I_B(p) = 4\pi\sum_{l=0}^{\infty}\sum_{m=-l}^{l}(m_B)_{l,m}(p)\,\overline{(m_B)_{l,m}(p)}, \tag{20}
$$

and the pair distance distribution function is computed as

$$
p_B(r) = \frac{r^2}{2\pi^2}\int_0^{\infty} I_B(p)\,j_0(pr)\,p^2dp. \tag{21}
$$

which is the same as Equation (1), by noting that $\sin(pr)/pr = j_0(pr)$.

In the following section we introduce a new way of doing these calculations that involves fewer numerical integrals in Fourier space, thereby potentially reducing aliasing effects.

3.3. Explicit computation of pair distance distribution function

Given the molecular complex, B, we can directly apply Equations (18) and (20) for computing the pair distance distribution analytically. Specifically one can first analytically compute $(m_B)_{l,m}(p)$ using Equation (18), and square it to sum over l and m to compute $I_B(p)$ as in Equation (20). Then one can obtain pair distance distribution function $p_B(r)$ from Equation (21) directly.

This leads to

$$
\boxed{p_B(r) = \frac{2r^2}{\pi}\sum_{l=0}^{\infty}\int_0^{\infty}\int_0^{\infty}\left[\sum_{m=-l}^{l}(\chi_B^{\circ})_{l,m}(r'')\overline{(\chi_B^{\circ})_{l,m}(r')}\right]\begin{pmatrix}0 & l & l\\ r & r' & r''\end{pmatrix}_j r''^2r'^2dr''dr'} \tag{22}
$$

where

$$
\begin{pmatrix}0 & l & l\\ r & r' & r''\end{pmatrix}_j \doteq \int_0^{\infty} j_0(pr)\,j_l(pr'')\,j_l(pr')p^2dp
$$

which denotes ''3j integral'' (Dong and Chirikjian, 2014).*

*Here the subscript j denotes that integral resulted from spherical-Bessel functions, and distinguishes this notation from that used from the Wigner 3-J symbols (Chirikjian, 2012; Varshalovich et al., 1988), which are not related.

These equations are the central equations of this article. The following section explains how $3j$ integrals can be computed using efficient and stable recurrence formulas. Then the computation of $I_B(p)$ using Equation (2) follows.

4. INTEGRALS OF PRODUCTS OF THREE SPHERICAL-BESSEL FUNCTIONS

In this section, we present the motivation and analytical underpinnings for our computational contribution to matching candidate configurations of biomolecular complexes to the data collected by SAXS.

4.1. Motivation: Why do these arise in this problem?

Historically, integrals of product of three spherical-Bessel functions have been studied in the field of physics, such as nuclear physics, particle physics, and astrophysics, most of which are devoted to finding analytical solutions of these integrals as series solutions involving many nested summations (Gervois and Navelet, 1985a,b, 1989; Mehrem et al., 1991; Mehrem and Hohenegger, 2010; Mehrem, 2013). Unlike conventional studies, in this article, we are developing the recursive form of the integrals of the product of three spherical-Bessel functions that lend themselves to efficient computation. This current study is not only applicable to SAXS, but also can be applied to the fields mentioned above.

Focusing our main attention to SAXS, if a function of position is expanded in spherical coordinates, it is natural to use Bessel functions for the radial dependence and spherical harmonics for the angular dependence. When substituting into the formulas for the pair distribution function, integrals over the angular quantities can be computed in closed form and a series of products of Bessel functions results. But in SAXS experiments the distribution $p_B(r)$ is itself a function of radial coordinates, and a natural basis in which to express it is spherical-Bessel functions. Therefore, matching the model pair distribution function with the experimental data requires projecting the model into the same basis set as the data. Essentially, applying the inverse spherical-Hankel transform to the model introduces an integral of the product of three spherical-Bessel functions (the two originally in the model, and one that appears as part of the projection process). Efficient evaluation of these integrals therefore would facilitate the process of matching coefficients in models of SAXS densities and the experimental data.

4.2. Recursive computation

The spherical-Bessel functions are one of the classical sets of orthogonal functions used in mathematical physics. The first few of these can be computed in closed form in terms of elementary functions as $j_0(x) = \frac{\sin x}{x}$, $j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x}$, and $j_2(x) = (\frac{3}{x^2} - 1)\frac{\sin x}{x} - \frac{3\cos x}{x^2}$. Spherical-Bessel functions of higher index can be recursively computed from these initial values using the relationships

$$j_n(x) = \frac{x}{2n+1}[j_{n+1}(x) + j_{n-1}(x)] \tag{23}$$

and

$$j_n'(x) = \frac{1}{2n+1}[nj_{n-1}(x) - (n+1)j_{n+1}(x)] \tag{24}$$

where $'$ denotes the derivative with respect to x.

The three-term spherical-Bessel-function integrals that arise in the motivating applications in SAXS are of form

$$\begin{pmatrix} m & n & l \\ a & b & c \end{pmatrix}_j \doteq \int_0^\infty j_m(ar) j_n(br) j_l(cr) r^2 \, dr. \tag{25}$$

We use the subscript j here to distinguish this from other integrals that use similar notation. We believe that recursive algorithms for computing Equation (25) have not been attempted before, because only recently has the problem of integrals of three Bessel functions been articulated in Auluck (2012), and in that article

the Bessel function $J_m(x)$ [see, e.g., Watson (1995) for definition] were considered rather than the spherical-Bessel functions $j_l(x)$.

4.2.1. Useful properties. Obviously, since scalar multiplication is commutative, we always have that these integrals are invariant under permutations of columns:

$$\begin{pmatrix} m & n & l \\ a & b & c \end{pmatrix}_j = \begin{pmatrix} n & m & l \\ b & a & c \end{pmatrix}_j = \begin{pmatrix} l & n & m \\ c & b & a \end{pmatrix}_j = \begin{pmatrix} m & l & n \\ a & c & b \end{pmatrix}_j = \begin{pmatrix} l & m & n \\ c & a & b \end{pmatrix}_j = \begin{pmatrix} n & l & m \\ b & c & a \end{pmatrix}_j.$$

When $c>0$, if $r'=cr$, we get $dr'=cdr$, $r=c^{-1}r'$, and $dr=c^{-1}dr'$. Substituting into Equation (25),

$$\begin{aligned} &\begin{pmatrix} m & n & l \\ a & b & c \end{pmatrix}_j \\ &= \int_0^\infty j_m\left(\frac{a}{c}r'\right) j_n\left(\frac{b}{c}r'\right) j_l(r') c^{-2} r'^{-2} c^{-1} dr' \\ &= \frac{1}{c^3}\begin{pmatrix} m & n & l \\ \frac{a}{c} & \frac{b}{c} & 1 \end{pmatrix}_j. \end{aligned} \tag{26}$$

Another special case results, when $c=0, m \neq n$

$$\begin{aligned} \begin{pmatrix} m & n & l \\ a & b & c \end{pmatrix}_j &= \int_0^\infty j_m(ar) j_n(br) j_l(0) r^2 dr \\ &= \delta_{l,0} \int_0^\infty j_m(ar) j_n(br) r^2 dr, \end{aligned} \tag{27}$$

and when $c=0$, $m=n$

$$\begin{aligned} \begin{pmatrix} m & n & l \\ a & b & c \end{pmatrix}_j &= \int_0^\infty j_m(ar) j_n(br) j_l(0) r^2 dr \\ &= \delta_{l,0} \frac{\pi\delta(a-b)}{2a^2}. \end{aligned} \tag{28}$$

4.2.2. General recurrence relations. Based on the expression $j_n(x) = \frac{x}{2n+1}[j_{n+1}(x) + j_{n-1}(x)]$, we set

$$\begin{aligned} G(m,n,l,a,b,c) &\doteq \int_0^\infty j_m(ar) j_n(br) j_l(cr) r dr \\ &= \int_0^\infty \frac{ar}{2m+1}[j_{m+1}(ar) + j_{m-1}(ar)] j_n(br) j_l(cr) r dr \\ &= \frac{a}{2m+1}\left(\begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j + \begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j\right). \end{aligned} \tag{29}$$

By a similar computation, we can get

$$\begin{aligned} G(m,n,l,a,b,c) &= \int_0^\infty j_m(ar) j_n(br) j_l(cr) r dr \\ &= \int_0^\infty j_m(ar) \frac{br}{2n+1}[j_{n+1}(br) + j_{n-1}(br)] j_l(cr) r dr \\ &= \frac{b}{2n+1}\left(\begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j + \begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j\right) \end{aligned} \tag{30}$$

and

$$\begin{aligned} G(m,n,l,a,b,c) &= \int_0^\infty j_m(ar)\, j_n(br)\, j_l(cr) r dr \\ &= \int_0^\infty j_m(ar) j_n(br) \frac{cr}{2l+1} [j_{l+1}(cr) + j_{l-1}(cr)] r dr \\ &= \frac{c}{2l+1} \left(\begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j + \begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \right) \end{aligned} \tag{31}$$

These will be used in the next section to obtain the general recurrence relations, which is for computing Equation (25) rather than discretizing these integrals.

4.2.3. Generating recurrence relations using integration by parts. Using integration by parts to compute the function G defined in the previous section is another route to obtaining recurrence relations as follows.

$$\begin{aligned} G(m,n,l,a,b,c) &= \int_0^\infty j_m(ar) j_n(br) j_l(cr) r dr \\ &= \frac{1}{2} \int_0^\infty j_m(ar) j_n(br) j_l(cr) dr^2 \\ &= \frac{1}{2} \left[j_m(ar) j_n(br) j_l(cr) r^2 \Big|_0^\infty - \int_0^\infty \frac{\partial j_m(ar) j_n(br) j_l(cr)}{\partial r} r^2 dr \right] \\ &= \frac{1}{2} \left[0 - \int_0^\infty (a j_m'(ar) j_n(br) j_l(cr) + b j_m(ar) j_n'(br) j_l(cr) \right. \\ &\quad \left. + c j_m(ar) j_n(br) j_l'(cr)) r^2 dr \right] \\ &= \frac{1}{2} \int_0^\infty \left\{ \frac{a}{2m+1} ((m+1) j_{m+1}(ar) - m j_{m-1}(ar)) j_n(br) j_l(cr) \right. \\ &\quad + \frac{b}{2n+1} ((n+1) j_{n+1}(br) - n j_{n-1}(br)) j_m(ar) j_l(cr) \\ &\quad \left. + \frac{c}{2l+1} ((l+1) j_{l+1}(cr) - l j_{l-1}(cr)) j_m(ar) j_n(br) \right\} r^2 dr \\ &= \frac{a(m+1)}{2(2m+1)} \begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j - \frac{am}{2(2m+1)} \begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j \\ &\quad + \frac{b(n+1)}{2(2n+1)} \begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j - \frac{bn}{2(2n+1)} \begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j \\ &\quad + \frac{c(l+1)}{2(2l+1)} \begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j - \frac{cl}{2(2l+1)} \begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \end{aligned} \tag{32}$$

Then equating (29) = (32), (30) = (32), (31) = (32) gives

$$\begin{aligned} &\frac{a(m-1)}{2(2m+1)} \begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j + \frac{b(n+1)}{2(2n+1)} \begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j \\ &\quad + \frac{c(l+1)}{2(2l+1)} \begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j = \frac{a(2+m)}{2(2m+1)} \begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j \\ &\quad + \frac{bn}{2(2n+1)} \begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j + \frac{cl}{2(l+1)} \begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \end{aligned} \tag{33}$$

$$\frac{a(m+1)}{2(2m+1)}\begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j + \frac{b(n-1)}{2(2n+1)}\begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j + \frac{c(l+1)}{2(2l+1)}\begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j = \frac{am}{2(2m+1)}\begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j + \frac{b(2+n)}{2(2n+1)}\begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j + \frac{cl}{2(2l+1)}\begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \tag{34}$$

$$\frac{a(m+1)}{2(2m+1)}\begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j + \frac{b(n+1)}{2(2n+1)}\begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j + \frac{c(l-1)}{2(2l+1)}\begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j = \frac{am}{2(2m+1)}\begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j + \frac{bn}{2(2n+1)}\begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j + \frac{c(2+l)}{2(2l+1)}\begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \tag{35}$$

Rewriting (33), (34), and (35) in matrix form,

$$\begin{pmatrix} \frac{a(m-1)}{2(2m+1)} & \frac{b(n+1)}{2(2n+1)} & \frac{c(l+1)}{2(2l+1)} \\ \frac{a(m+1)}{2(2m+1)} & \frac{b(n-1)}{2(2n+1)} & \frac{c(l+1)}{2(2l+1)} \\ \frac{a(m+1)}{2(2m+1)} & \frac{b(n+1)}{2(2n+1)} & \frac{c(l-1)}{2(2l+1)} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} m+1 & n & l \\ a & b & c \end{pmatrix}_j \\ \begin{pmatrix} m & n+1 & l \\ a & b & c \end{pmatrix}_j \\ \begin{pmatrix} m & n & l+1 \\ a & b & c \end{pmatrix}_j \end{pmatrix} = \begin{pmatrix} \frac{a(2+m)}{2(2m+1)} & \frac{bn}{2(2n+1)} & \frac{cl}{2(2l+1)} \\ \frac{am}{2(2m+1)} & \frac{b(2+n)}{2(2n+1)} & \frac{cl}{2(2l+1)} \\ \frac{am}{2(2m+1)} & \frac{bn}{2(2n+1)} & \frac{c(2+l)}{2(2l+1)} \end{pmatrix} \begin{pmatrix} \begin{pmatrix} m-1 & n & l \\ a & b & c \end{pmatrix}_j \\ \begin{pmatrix} m & n-1 & l \\ a & b & c \end{pmatrix}_j \\ \begin{pmatrix} m & n & l-1 \\ a & b & c \end{pmatrix}_j \end{pmatrix}. \tag{36}$$

If we let $M_j = \begin{pmatrix} \frac{a(m-1)}{2(2m+1)} & \frac{b(n+1)}{2(2n+1)} & \frac{c(l+1)}{2(2l+1)} \\ \frac{a(m+1)}{2(2m+1)} & \frac{b(n-1)}{2(2n+1)} & \frac{c(l+1)}{2(2l+1)} \\ \frac{a(m+1)}{2(2m+1)} & \frac{b(n+1)}{2(2n+1)} & \frac{c(l-1)}{2(2l+1)} \end{pmatrix}$, then the determinant of this matrix is

$$\det(M_j) = \frac{abc(m+n+l+1)}{2(2l+1)(2m+1)(2n+1)},$$

which is never zero when $a,b,c>0$ and $m,n,l\geq 0$. Inverting this matrix gives four-term recurrence relations for the quantity of interest. This is very promising for computing the integrals that arise in SAXS, and represents an alternative to numerical integration, which has the obvious drawback of not being exact, requiring a finite truncation of the upper bound of integration, and being very time consuming to evaluate on each point in a fine discretized grid of values of a, b, c and many values of m, n, l.

4.2.4. Recurrence relations when $m=-1, 0$. We note that in Equation (22), we need a special form of $3j$ integral as $m=0$ and $n=l$. For that reason, in this section, we develop the recurrence relations for the special cases where n and l vary with m fixed as 0 or 1. The reason why we need to consider $m=-1$ as well as $m=0$ will be clear in the following derivation of the recurrence relations. First let us define

$$x_{n,l} \doteq \begin{pmatrix} -1 & n & l \\ a & b & c \end{pmatrix}_j$$

and

$$y_{n,l} \doteq \begin{pmatrix} 0 & n & l \\ a & b & c \end{pmatrix}_j.$$

We seek to find the vector recurrence relations with these two variables.

Let us consider the definition of $j_{-1}(x)$ as

$$j_{-1}(x) = \frac{\cos x}{x}. \tag{37}$$

This satisfies the recurrence formula Equation (23) and the definition of spherical-Bessel functions $j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x)$.

Let us consider the following term

$$\int_0^\infty \cos(ar)\, j_n(br)\, j_l(cr)\, dr. \tag{38}$$

This can be expressed into two different terms, by using Equation (23), as

$$\begin{aligned} &\int_0^\infty \cos(ar)\, j_n(br)\, j_l(cr)\, dr \\ &= \frac{ab}{2n+1}(x_{n+1,l} + x_{n-1,l}) \\ &= \frac{ac}{2l+1}(x_{n,l+1} + x_{n,l-1}). \end{aligned} \tag{39}$$

Then let us consider

$$\int_0^\infty \cos(ar) \frac{\partial j_n(br)}{\partial r}\, j_l(cr)\, dr. \tag{40}$$

This term can be expressed into two different forms. First,

$$\begin{aligned} &= \int_0^\infty \cos(ar) \frac{b}{2n+1} \{n j_{n-1}(br) - (n+1) j_{n+1}(br)\}\, j_l(cr)\, dr \\ &= \frac{b}{2n+1} \int_0^\infty \cos(ar) \{n j_{n-1}(br) - (n+1) j_{n+1}(br)\}\, j_l(cr)\, dr. \end{aligned} \tag{41}$$

The other expression can be obtained by using the integration by parts as

$$\begin{aligned} &= \cos(ar)\, j_n(br)\, j_l(br)\Big|_0^\infty - \int_0^\infty j_n(br) \frac{\partial}{\partial r} (\cos(ar)\, j_l(cr))\, dr \\ &= -\delta_{n,0}\delta_{l,0} + a \int_0^\infty \sin(ar)\, j_n(br)\, j_l(cr)\, dr - \int_0^\infty \cos(ar)\, j_n(br)\, c j_l'(cr)\, dr \\ &= -\delta_{n,0}\delta_{l,0} + a \int_0^\infty \sin(ar)\, j_n(br)\, j_l(cr)\, dr \\ &\quad - \frac{c}{2l+1} \int_0^\infty \cos(ar)\, j_n(br)\, (l j_{l-1}(cr) - (l+1) j_{l+1}(cr))\, dr. \end{aligned} \tag{42}$$

Now we take the partial derivative with respect to a to both the expressions. Then Equation (41) becomes

$$= -\frac{b}{2n+1} \int_0^\infty r \sin(ar) \{n j_{n-1}(br) - (n+1) j_{n+1}(br)\}\, j_l(cr)\, dr \tag{43}$$

and (42) becomes

$$\begin{aligned} &= \int_0^\infty \sin(ar)\, j_n(br)\, j_l(cr)\, dr + a \int_0^\infty r \cos(ar)\, j_n(br)\, j_l(cr)\, dr \\ &\quad - \frac{c}{2l+1} \int_0^\infty (-r) \sin(ar)\, j_n(br) (l j_{l-1}(cr) - (l+1) j_{l+1}(cr))\, dr. \end{aligned} \tag{44}$$

Similarly, we have

$$\begin{aligned}&\frac{\partial}{\partial a}\int_0^\infty \sin(ar)\frac{\partial j_n(br)}{\partial r} j_l(cr)\,dr\\ &\quad=\int_0^\infty \frac{\partial}{\partial a}\sin(ar)\frac{b}{2n+1}\{nj_{n-1}(br)-(n+1)j_{n+1}(br)\} j_l(cr)\,dr\\ &\quad=\frac{b}{2n+1}\int_0^\infty r\cos(ar)\{nj_{n-1}(br)-(n+1)j_{n+1}(br)\} j_l(cr)\,dr\end{aligned} \tag{45}$$

and

$$\begin{aligned}&\frac{\partial}{\partial a}\int_0^\infty \sin(ar)\frac{\partial j_n(br)}{\partial r} j_l(cr)\,dr\\ &\quad=-\int_0^\infty \cos(ar) j_n(br) j_l(cr)\,dr+a\int_0^\infty r\sin(ar) j_n(br) j_l(cr)\,dr\\ &\qquad-\frac{c}{2l+1}\int_0^\infty r\cos(ar) j_n(br)\,(lj_{l-1}(cr)-(l+1)j_{l+1}(cr))\,dr.\end{aligned} \tag{46}$$

Then from Equations (43) and (44), we have

$$\begin{aligned}-\frac{abn}{2n+1}y_{n-1,l}+\frac{ab(n+1)}{2n+1}y_{n+1,l}=&\int_0^\infty \sin(ar)\, j_n(br) j_l(cr)\,dr\\ &+a^2x_{n,l}+\frac{acl}{2l+1}y_{n,l-1}-\frac{ac(l+1)}{2l+1}y_{n,l+1}.\end{aligned} \tag{47}$$

Also from Equations (45) and (46), we have

$$\begin{aligned}\frac{abn}{2n+1}x_{n-1,l}-\frac{ab(n+1)}{2n+1}x_{n+1,l}=&-\int_0^\infty \cos(ar) j_n(br) j_l(cr)\,dr\\ &+a^2y_{n,l}-\frac{acl}{2l+1}x_{n,l-1}+\frac{ac(l+1)}{2l+1}x_{n,l+1}.\end{aligned} \tag{48}$$

Here $\int_0^\infty \cos(ar) j_n(br) j_l(cr)\,dr$ can be expressed with Equation (39). Combining them with Equation (48), we have

$$\begin{aligned}\frac{bn}{2n+1}x_{n+1,l}+\frac{c(l+1)}{2l+1}x_{n,l+1}&=\frac{b(n+1)}{2n+1}x_{n-1,l}+\frac{cl}{2l+1}x_{n,l-1}-ay_{n,l}\\ \frac{b(n+1)}{2n+1}x_{n+1,l}+\frac{cl}{2l+1}x_{n,l+1}&=\frac{bn}{2n+1}x_{n-1,l}+\frac{c(l+1)}{2l+1}x_{n,l-1}-ay_{n,l}.\end{aligned} \tag{49}$$

Then, let us consider

$$\int_0^\infty \sin(ar) j_n(br) j_l(cr)\,dr \tag{50}$$

similarly with Equation (39). Then we have two expressions for it as

$$\begin{aligned}&\int_0^\infty \sin(ar) j_n(br) j_l(cr)\,dr\\ &=\frac{ab}{2n+1}(y_{n-1,l}+y_{n+1,l})\\ &=\frac{ac}{2l+1}(y_{n,l-1}+y_{n,l+1}).\end{aligned} \tag{51}$$

We combine these expressions with Equation (47) to have

$$\begin{aligned}\frac{bn}{2n+1}y_{n+1,l}+\frac{c(l+1)}{2l+1}y_{n,l+1}&=\frac{b(n+1)}{2n+1}y_{n-1,l}+\frac{cl}{2l+1}y_{n,l-1}+ax_{n,l}\\ \frac{b(n+1)}{2n+1}y_{n+1,l}+\frac{cl}{2l+1}y_{n,l+1}&=\frac{bn}{2n+1}y_{n-1,l}+\frac{c(l+1)}{2l+1}y_{n,l-1}+ax_{n,l}.\end{aligned} \tag{52}$$

We arrange Equations (49) and (52) to have

$$\begin{pmatrix} \frac{bn}{2n+1} & \frac{c(l+1)}{2l+1} & 0 & 0 \\ \frac{b(n+1)}{2n+1} & \frac{cl}{2l+1} & 0 & 0 \\ 0 & 0 & \frac{bn}{2n+1} & \frac{c(l+1)}{2l+1} \\ 0 & 0 & \frac{b(n+1)}{2n+1} & \frac{cl}{2l+1} \end{pmatrix} \begin{pmatrix} x_{n+1,l} \\ x_{n,l+1} \\ y_{n+1,l} \\ y_{n,l+1} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \frac{b(n+1)}{2n+1} & \frac{cl}{2l+1} & -a & 0 & 0 \\ 0 & \frac{bn}{2n+1} & \frac{c(l+1)}{2l+1} & -a & 0 & 0 \\ a & 0 & 0 & 0 & \frac{b(n+1)}{2n+1} & \frac{cl}{2l+1} \\ a & 0 & 0 & 0 & \frac{bn}{2n+1} & \frac{c(l+1)}{2l+1} \end{pmatrix} \begin{pmatrix} x_{n,l} \\ x_{n-1,l} \\ x_{n,l-1} \\ y_{n,l} \\ y_{n-1,l} \\ y_{n,l-1} \end{pmatrix}. \tag{53}$$

The determinant of 2×2 matrix in the left-hand side of the equation is $\left(\frac{bc(n+l+1)}{(2n+1)(2l+1)}\right)^2$, which is not zero. Hence it is solvable and complete, once $m=-1,\ 0$ and $n=-1,\ 0$ are computed, which is the topic of the following section.

4.2.5. Recurrence relations when $m=-1,\ 0$ and $n=-1,\ 0$. We define

$$\begin{pmatrix} -1 & -1 & l \\ a & b & c \end{pmatrix}_j = \frac{1}{ab}\int_0^\infty \cos(ar)\cos(br)\, j_l(cr)\, dr \doteq z_l, \tag{54}$$

$$\begin{pmatrix} -1 & 0 & l \\ a & b & c \end{pmatrix}_j = \frac{1}{ab}\int_0^\infty \cos(ar)\sin(br)\, j_l(cr)\, dr \doteq v_l, \tag{55}$$

$$\begin{pmatrix} 0 & -1 & l \\ a & b & c \end{pmatrix}_j = \frac{1}{ab}\int_0^\infty \sin(ar)\cos(br)\, j_l(cr)\, dr \doteq u_l, \tag{56}$$

and

$$\begin{pmatrix} 0 & 0 & l \\ a & b & c \end{pmatrix}_j = \frac{1}{ab}\int_0^\infty \sin(ar)\sin(br)\, j_l(cr)\, dr \doteq w_l. \tag{57}$$

First, let us consider z_l; let us take $\frac{\partial}{\partial c}$ to

$$\int_0^\infty \cos(ar)\cos(br)\, j_l(cr)\, r^{-1} dr. \tag{58}$$

Then

$$\begin{aligned} &\frac{\partial}{\partial c}\int_0^\infty \cos(ar)\cos(br)\ j_l(cr)\ r^{-1}\ dr \\ &= \int_0^\infty \cos(ar)\ \cos(br)\ r j_l'(cr)\ r^{-1}\ dr \\ &= \frac{abl}{2l+1} z_{l-1} - \frac{ab(l+1)}{2l+1} z_{l+1}. \end{aligned} \tag{59}$$

Another expression is obtained by calculating the following

$$\begin{aligned} &\int_0^\infty \cos(ar)\ \cos(br)\ \frac{\partial j_l(cr)}{\partial c}\ r^{-1}\ dr \\ &= \int_0^\infty \cos(ar)\ \cos(br)\ \frac{r}{c}\frac{\partial j_l(cr)}{\partial r}\ r^{-1}\ dr = \frac{1}{c}\int_0^\infty \cos(ar)\ \cos(br)\ \frac{\partial j_l(cr)}{\partial r}\ dr \\ &= \frac{1}{c}\cos(ar)\ \cos(br)\, j_l(cr)\Big|_0^\infty - \frac{1}{c}\int_0^\infty j_l(cr)\ \frac{\partial}{\partial r}(\cos(ar)\ \cos(br))\ dr \\ &= -\frac{1}{c}\delta_{l,0} - \frac{1}{c}\int_0^\infty (-a)\sin(ar)\ \cos(br)\, j_l(cr)\ dr - \frac{1}{c}\int_0^\infty (-b)\cos(ar)\ \sin(br)\, j_l(cr)\ dr \end{aligned} \tag{60}$$

From Equations (59) and (60), we obtain

$$\frac{abl}{2l+1} z_{l-1} - \frac{ab(l+1)}{2l+1} z_{l+1} = -\frac{1}{c}\delta_{l,0} + \frac{a^2 b}{c} u_l + \frac{ab^2}{c} v_l. \tag{61}$$

Similar techniques can be applied to the cases for u_l, v_l, and w_l. When we consider

$$\frac{\partial}{\partial c} \int_0^\infty \cos(ar)\ \sin(br)\ j_l(cr)\ r^{-1}\ dr, \tag{62}$$

then we obtain

$$\frac{abl}{2l+1} v_{l-1} - \frac{ab(l+1)}{2l+1} v_{l+1} = \frac{a^2 b}{c} w_l - \frac{ab^2}{c} z_l. \tag{63}$$

When we consider

$$\frac{\partial}{\partial c} \int_0^\infty \sin(ar)\ \cos(br)\ j_l(cr)\ r^{-1}\ dr, \tag{64}$$

then we obtain

$$\frac{abl}{2l+1} u_{l-1} - \frac{ab(l+1)}{2l+1} u_{l+1} = -\frac{a^2 b}{c} z_l + \frac{ab^2}{c} w_l. \tag{65}$$

When we consider

$$\frac{\partial}{\partial c} \int_0^\infty \sin(ar)\ \sin(br)\ j_l(cr)\ r^{-1}\ dr, \tag{66}$$

then we obtain

$$\frac{abl}{2l+1} w_{l-1} - \frac{ab(l+1)}{2l+1} w_{l+1} = -\frac{a^2 b}{c} v_l - \frac{ab^2}{c} u_l. \tag{67}$$

In total, we have four equations as

$$\begin{aligned}
\frac{l+1}{2l+1} z_{l+1} &= \frac{l}{2l+1} z_{l-1} - \frac{a}{c} u_l - \frac{b}{c} v_l + \frac{1}{abc}\delta_{l,0} \\
\frac{l+1}{2l+1} v_{l+1} &= \frac{l}{2l+1} v_{l-1} - \frac{a}{c} w_l + \frac{b}{c} z_l \\
\frac{l+1}{2l+1} u_{l+1} &= \frac{l}{2l+1} u_{l-1} + \frac{a}{c} z_l - \frac{b}{c} w_l \\
\frac{l+1}{2l+1} w_{l+1} &= \frac{l}{2l+1} w_{l-1} + \frac{a}{c} v_l + \frac{b}{c} u_l
\end{aligned} \tag{68}$$

Let $\boldsymbol{\xi}_l = [z_l\ v_l\ u_l\ w_l]^T$. Then four equations can be arranged as

$$\frac{l+1}{2l+1} \boldsymbol{\xi}_{l+1} = \frac{l}{2l+1} \boldsymbol{\xi}_{l-1} + L\boldsymbol{\xi}_l + \frac{1}{abc}\delta_{l,0}\mathbf{e}_1 \tag{69}$$

where $\mathbf{e}_1 = [1\ 0\ 0\ 0]^T$ and

$$L = \begin{pmatrix} 0 & -\frac{b}{c} & -\frac{a}{c} & 0 \\ \frac{b}{c} & 0 & 0 & -\frac{a}{c} \\ \frac{a}{c} & 0 & 0 & -\frac{b}{c} \\ 0 & \frac{a}{c} & \frac{b}{c} & 0 \end{pmatrix}.$$

This completes the recurrence relations for $m = -1,\ 0$ and $n = -1,\ 0$.

Regarding the initial conditions, we need to compute the cases when $l = 0$. Then from Equation (69), we can obtain other terms sequentially. The closed forms of the initial conditions can be found as

$$\begin{pmatrix} -1 & -1 & 0 \\ a & b & c \end{pmatrix}_j = \begin{cases} 0 & c < |a-b| \\ \frac{\pi}{8abc} & c = |a-b| \\ \frac{\pi}{4abc} & |a-b| < c < a+b, \\ \frac{3\pi}{8abc} & c = a+b \\ \frac{\pi}{2abc} & c > a+b \end{cases} \tag{70}$$

$$\begin{pmatrix} -1 & 0 & 0 \\ a & b & c \end{pmatrix}_j = \begin{cases} \frac{1}{8abc}\log\left(\frac{(-a+b+c)^2(a+b+c)^2}{(a+b-c)^2(a-b+c)^2}\right) & c < |a-b| \text{ or } c \geq a+b \\ -\frac{1}{4abc}\log\left(\frac{a^2-(b-c)^2}{-a^2+(b+c)^2}\right) & |a-b| \leq c < a+b \end{cases}, \tag{71}$$

and

$$\begin{pmatrix} 0 & 0 & 0 \\ a & b & c \end{pmatrix}_j = \begin{cases} 0 & c < |a-b| \text{ or } c > a+b \\ \frac{\pi}{4abc} & |a-b| < c < a+b \\ \frac{\pi}{8abc} & c = |a-b| \text{ or } c = a+b \end{cases}. \tag{72}$$

5. COMPUTATIONAL COMPLEXITY

In numerical computations, the bandlimit for summations over l and m must be finite, and integrals over r and p must be sampled at discrete values up to a finite limit. Suppose that we have $O(N)$ discrete values for each of these variables. In the subsections that follow we analyze the computational cost and issues affecting numerical accuracy when computing the pair distribution function using traditional approaches and when using our 3j-integral approach.

5.1. Cost by direct evaluation

Since Equation (19) is used in both approaches, let us consider it first. Given l, m, and each point on r, there are two integrations (over the unit sphere $\mathbb{S}^2$), which makes the computational cost as $O(N^2)$.

Now the computational cost for the approach to use Equations (18), (20), and (21) is estimated as follows. Regarding the calculation of (18), one has to consider each pair of l and m, and each point of p. Also there is one integration on r. Hence the total cost for calculating (18) becomes $O(N^4)$. Furthermore, (20) involves two summations and storage to an array of p, which leads to the cost of $O(N^3)$. Then by the same reasoning, the computational cost for (21) becomes $O(N^2)$. If we are particularly concerned with (20) and (21), the total computational cost is estimated as $O(N^4)$.

A similar cost estimate for computing Equation (22) can be performed. Suppose we have precomputed and store all of 3j integral values. First, the computational cost of brute-force precomputation for 3j integral is estimated as $O(N^5)$ with $O(N)$ values for each variable, and the space cost to store all of these is $O(N^4)$. But there are one-time costs. In computing Equation (22), we begin with the summation over m of $(\chi_B^\circ)_{l,m}(r'')\overline{(\chi_B^\circ)_{l,m}(r')}$. This has a cost of $O(N^4)$, and afterward m is removed from further nested computations. Since what remains in Equation (22) involves one summation and two integrations for each fixed value of r, the cost of directly computing Equation (22) is $O(N^4)$ when done for all values of r. This is a similar computational cost with the previous approach.

It should be noted that all of the computational costs described here do not assume any properties of recursive computations. The potential to gain efficiencies using recursions is described in the next section.

5.2. Cost by fast Hankel and fast spherical-harmonic transforms

Instead of directly computing the terms, one can use the fast transform approaches. General three-term linear recurrence relations have been studied extensively (Amos and Burgmeier, 1973; Gautschi, 1967). In addition, algorithms for the fast Bessel/spherical-Bessel transforms (Candel, 1981; Goldstein and Thaler, 1959; Johansen and Sorensen, 1979) and the fast spherical harmonics (Driscoll and Healy, 1994; Healy et al., 2003; Gorski et al., 2005; Gumerov and Duraiswami, 2005) have been investigated.

Though ''fast'' algorithms for spherical-Bessel transforms have reported complexities of as low as $O(N \log N)$ and $O(N^2 \log^2 N)$, respectively, they can suffer from numerical stability problems. Nevertheless, we revisit the issue of computational cost if one hypothetically uses the fastest of these algorithms.

For the approach in Equations (18), (20), and (21), the $O(N^4)$ bottleneck reduces to $O(N^3 \log {}^2N)$ since the spherical-harmonic transform is computed for each of $O(N)$ values of r.

Recursions help in computing Equation (22) in several ways. First, the same computational bottleneck mentioned above appears in this method as well. Second, rather than precomputing $3j$ integrals, in principle they can be computed on the fly. That is, because the $3j$ integrals satisfy recurrences, the potential to develop a ''fast $3j$ transform'' exists, which would make this competitive in terms of speed with the more traditional approach in Equations (18), (20), and (21). In addition, there are numerical stability and accuracy advantages to our new approach, as discussed below.

5.3. Numerical accuracy

In addition to computational cost, there are other advantages of recursively computing the $3j$ integrals and evaluating the pair distribution function using Equation (22). First, unlike Equation (20), which requires converting the problem to Fourier space and back, the approach of using $3j$ integrals keeps everything in real space. And since integrals over p are not sampled nor are they computed with finite bandlimits, this can avoid any possible aliasing, which is common in sampling in the Fourier approach. That is, recurrence relations that we have derived compute the exact values of the $3j$ integrals as if they had been computed analytically over the continuous half-infinite integral. Secondly, recurrence relations for $3j$ integrals are more stable than numerical integration and fast spherical-Bessel transforms. The reason for this is because the recurrence relations that result from inverting the matrix on the left side of Equation (36) all involve coefficients that depend only on m, n, l, with a, b, c factoring out. Therefore, the sorts of instabilities that arise when computing fast associated Legendre and fast spherical-Bessel transforms are avoided. It is counter-intuitive that the integral of products of functions has nicer recurrence relations than the original functions themselves. Both of aforementioned factors emphasize the importance of using the recurrence relations.

6. NUMERICAL EXAMPLES

We consider lactoferrin molecules as a numerical example. Lactoferrin molecules are know to have the hinge-bending motion [described by the motion among three rigid-cluster parts (Kim et al., 2005)] due to their iron-ion binding functions, which results in two different conformations (open and closed conformation, respectively represented as 1LFH and 1LFG), as shown in Figure 1. We apply the method for a single body of each conformation for the purpose of verifying the proposed model. Another purpose is to investigate whether the obtained $p_B(r)$ can distinguish different conformations of a single molecule. Note that the method proposed in this article can be applied to the detection of the structural differences as shown through the current example (open and closed in particular). Once we detect the structural differences, the proposed method can provide the important foundation to the investigation of the rigid body motions or flexibility between substructures of the complex, which forms the topic of future study.

The characteristic function is reconstructed as follows. First, given $\mathbf{x}_i$ denoted as the coordinates of C_α in PDB of the protein, $\varphi_B(\mathbf{x}) = \sum_{i=1}^{n} \delta(\mathbf{x} - \mathbf{x}_i)$, where n denotes the number of C_α. Then the reconstruction formula gives

$$\chi_B(r\mathbf{u}) = \frac{2}{\pi} \sum_{l=0}^{L} \sum_{m=-l}^{l} \int_0^\infty p^2 dp\, j_l(pr)\, Y_l^m(\mathbf{u}) \left(\int_{r'} \int_{\mathbb{S}^2} \varphi_B(r'\mathbf{u}')\, j_l(pr')\, \overline{Y_l^m(\mathbf{u}')}\, r'^2 dr' d\mathbf{u}' \right)$$

Note that $\int_r \int_{\mathbb{S}^2} (\cdot) r^2 dr d\mathbf{u}$ should be performed around the center of mass of the molecule, which is essential scattering center of the molecule. The term with spatial integration can be replaced by the summation as

$$\int_{r'} \int_{\mathrm{S}^2} \varphi_B(r'\mathbf{u}')\, j_l(pr')\, \overline{Y_l^m(\mathbf{u}')}\, r'^2 dr' d\mathbf{u}' = \sum_{i=1}^{n} j_l(pr_i)\, \overline{Y_l^m(\mathbf{u}_i)} \tag{73}$$

due to Dirac delta functions.

Figure 2 shows the results obtained by using Equation (22). We verified that numerical computations by using both Equations (22) and (21) produce the same results. Note that in Equation (71), there is a special

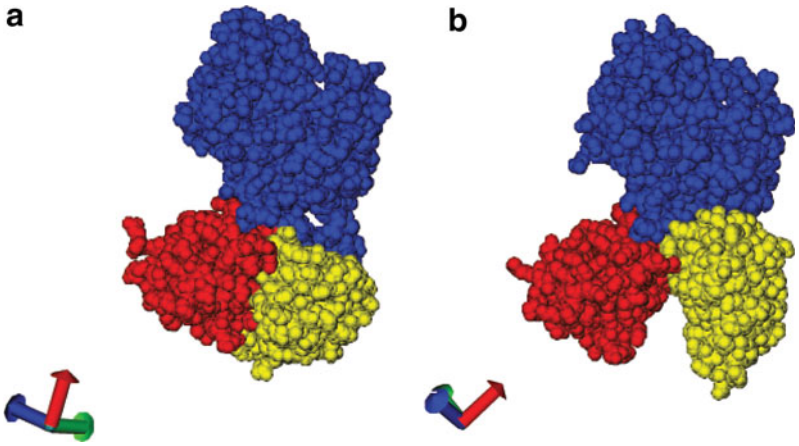

FIG. 1. Conformations of lactoferrin: **(a)** closed conformation (1LFG); **(b)** open conformation (1LFH). There are three rigid clusters: head in blue, and left and right clusters are denoted in red and yellow, respectively (Kim et al., 2005).

set of conditions that makes the integral divergent. In that case, we put an arbitrarily large number to apply the recurrence relations, which still works for Equation (22). Figure 2a and b shows that indeed SAXS can distinguish conformational information. And Figure 2c and d verifies our numerical computations of the pair distribution function with direct evaluation by binning all inter-residue distances and creating a histogram. Considering that $p_B(r)$'s in Figure 2a are probability density functions (i.e., normalized so as to its area is 1), the difference of $p_B(r)$'s between 1LFG and 1LFH is significant enough to reflect the conformational difference, which is also confirmed in Figure 2c and d with histograms of all inter-residue distances.

7. CONCLUSIONS AND FUTURE WORK

In this article we model the data acquisition process in SAXS, which is a major method for gaining insight into the structure of large biomolecular complexes. We established that integrals of products of three spherical-Bessel functions arise when a body is described as a unit density (or characteristic) function that is expanded in terms of spherical-Bessel functions in the radial direction. Then the the self-convolution of the density produces expressions containing products of spherical-Bessel functions, which when expressed in the original basis requires computing the product of three such functions.

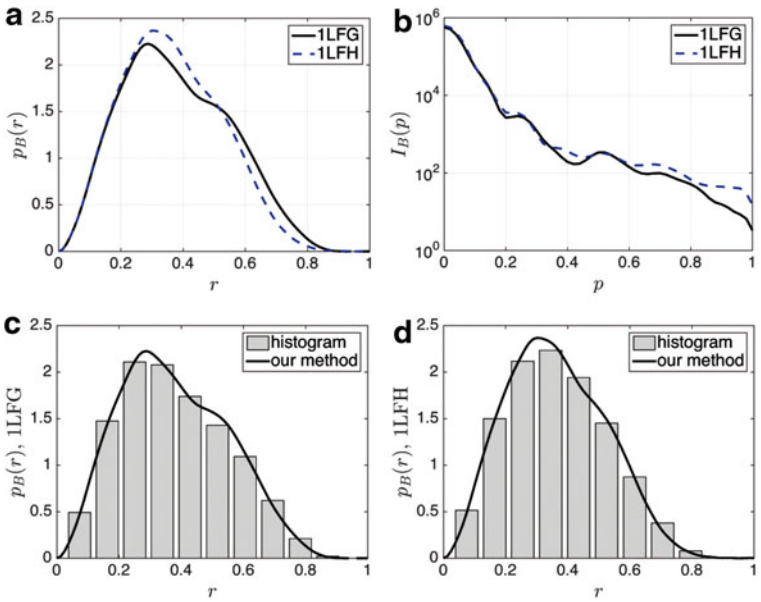

FIG. 2. **(a)** Plots of $p_B(r)$ for 1LFG and 1LFH computed by the proposed approach. **(b)** Plots of $I_B(p)$ by the approach. **(c, d)** Comparison of $p_B(r)$ with brute-force calculation by calculating the histograms of inter-distance of the molecules. **(c)** and **(d)**, respectively, correspond to 1LFG and 1LFH. $p_B(r)$ plots are normalized by the area so that they become probability densities.

Here we provided the motivation for this problem and showed how the integrals of products of three spherical-Bessel functions can be computed recursively. This consists of two parts. First, we derived the general recurrence relations, and second we derived the recurrence relations that are particularly suitable in the computation of $p_B(r)$. In doing so we made use of the recurrence relations for individual spherical-Bessel functions and used integration tricks such as changes of variables and integration by parts. We observed the counter-intuitive phenomenon that the integral of products of three spherical-Bessel functions has recurrence relations that are more well-behaved than the original functions themselves.

Finally, we demonstrated our method on a biologically meaningful example, the protein lactoferrin, and showed that SAXS can distinguish conformational changes. And we verified our methodology against brute-force computation of the pair distribution function by comparing with the histogram of all inter-residue distances.

Equipped with this motivation and the recurrence relations, the next step will be to apply these to develop numerical codes to solve problems of interest in structural biology. Future study includes the investigation of the detection of rigid body motions between sub structures of the given complex, and the treatment of flexibility of subparts during the motions.

ACKNOWLEDGMENT

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM113240.

AUTHOR DISCLOSURE STATEMENT

The authors declare that no competing financial interests exist.

REFERENCES

Abramowitz, M., and Stegun, I., eds. 1972. *Handbook of Mathematical Functions*. Dover Edition, New York.

Amos, D., and Burgmeier, J. 1973. Computation with three-term linear nonhomogeneous recurrence relations. *SIAM Rev.* 15, 335–351.

Andrews, G., Askey, R., and Roy, R. 1999. *Special Functions, Encyclopedia of Mathematics and Its Applications*, volume 71. Cambridge University Press, Cambridge, UK.

Arfken, G., and Weber, H. 2010. *Mathematical Methods for Physicists*, 7th edition. Academic Press, San Diego, CA.

Auluck, S. 2012. On the integral of the product of three Bessel functions over an infinite domain Fourier-space representation of nonlinear dynamics of continuous media in cylindrical geometry. *Math. J.* 14, 1–39.

Berman, H., Westbrook, J., Feng, Z., et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.

Candel, S. 1981. An algorithm for the Fourier-Bessel transform. *Comput. Phys. Commun.* 23, 343–353.

Chirikjian, G.S. 2012. *Stochastic Models, Information Theory and Lie Groups*, volume 2. Birkhäuser, Geneva.

Debye, P. 1915. Zerstreuung von Röntgenstrahlen. *Ann. Physik*. 28, 809–823.

Dong, H., and Chirikjian, G.S. 2014. A computational model for data acquisition in SAXS, 695–702. BDB '14 Proceeding of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics.

Driscoll, J., and Healy Jr., D. 1994. Computing Fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.* 15, 202–250.

Gautschi, G. 1967. Computational aspects of three-term recurrence relations. *SIAM Rev.* 9, 24–82.

Gervois, A., and Navelet, H. 1985a. Integrals of three Bessel functions and Legendre functions. I. *J. Math. Phys.* 26, 633–644.

Gervois, A., and Navelet, H. 1985b. Integrals of three Bessel functions and Legendre functions. II. *J. Math. Phys.* 26, 645–655.

Gervois, A., and Navelet, H. 1989. Infinite integrals involving three spherical Bessel functions. *SIAM J. Math. Anal.* 20 1006–1018.

Goldstein, M., and Thaler, R. 1959. Recurrence techniques for the calculation of Bessel functions. *Math. Tables Other Aids Comput.* 13, 102–108.

Gorski, K., Hivon, E., Banday, A., et al. 2005. HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys. J.* 622, 759.

Gumerov, N., and Duraiswami, R. 2005. *Fast Multiple Methods for the Helmholtz Equation in Three Dimensions*. Elsevier, New York.

Healy Jr., D., Rockmore, D., Kostelec, P., and Moore, S. 2003. FFTs for the 2-sphere-improvements and variations. *J. Fourier Anal. Appl.* 9, 341–385.

Johansen, H., and Sorensen, K. 1979. Fast Hankel-transforms. *Geophys. Prospect.* 27, 876–901.

Kim, M., Jernigan, R., and Chirikjian, G. 2005. Rigid-cluster models of conformational transitions in macromolecular machines and assemblies. *Biophys. J.* 89, 43–55.

Kratky, O. 1963. X-ray small angle scattering with substances of biological interest in diluted solutions. *Prog. Biophys. Mol. Biol.* 13, 105–173.

Mehrem, R. 2013. Analytical evaluation of an infinite integral over four 45 spherical Bessel functions. *Appl. Math. Comput.* 219, 10506–10509.

Mehrem, R., and Hohenegger, A. 2010. A generalization for the infinite integral over three spherical Bessel functions. *J. Phys. A Math. Theor.* 43, 455204.

Mehrem, R., Longdergan, J., and Macfarlane, M. 1991. Analytic expressions for integrals of products of spherical Bessel functions. *J. Phys. A Math. Gen.* 24, 1435–1453.

Svergun, D.I., Koch, M.H., Timmins, P.A., and May, R.P. 2013. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. Oxford University Press, Oxford, UK.

Varshalovich, D.A., Moskalev, A.N., and Khersonskii, V.K. 1988. *Quantum Theory of Angular Momentum*. World Science Publishing, New York.

Watson, G. 1995. *A Treatise on the Theory of Bessel Functions*. Cambridge University Press, Cambridge, UK.

Address correspondence to:
Prof. Gregory S. Chirikjian
Department of Mechanical Engineering
Johns Hopkins University
223 Latrobe Hall, 3400 N. Charles Street
Baltimore, MD 21218

E-mail: gregc@jhu.edu