Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 5 August 2008
Accepted 26 January 2009

© 2009 International Union of Crystallography
Printed in Singapore – all rights reserved

Molecular shapes from small-angle X-ray scattering: extension of the theory to higher scattering angles

V. L. Shneerson and D. K. Saldin*

Department of Physics, University of Wisconsin – Milwaukee, PO Box 413, Milwaukee, WI 53201, USA. Correspondence e-mail: dksaldin@uwm.edu

A low-resolution shape of a molecule in solution may be deduced from measured small-angle X-ray scattering $I(q)$ data by exploiting a Hankel transform relation between the coefficients of a multipole expansion of the scattered amplitude and corresponding coefficients of the electron density. In the past, the radial part of the Hankel transform has been evaluated with the aid of a truncated series expansion of a spherical Bessel function. It is shown that series truncation may be avoided by analytically performing the radial integral over an entire Bessel function. The angular part of the integral involving a spherical harmonic kernel is performed by quadrature. Such a calculation also allows a convenient incorporation of a molecular hydration shell of constant density intermediate between that of the protein and the solvent. Within this framework, we determine the multipole coefficients of the shape function by optimization of the agreement with experimental data by simulated annealing.

1. Introduction

Two of the techniques that have given us the greatest knowledge of the high-resolution three-dimensional structure of proteins are X-ray crystallography (Drenth, 1994) and NMR (Wüthrich, 1986). However, these methods have their limitations: it is often difficult to grow crystals of high-molecular-weight assemblies suitable for crystallography and the application of NMR is generally limited to small proteins of molecular weight less than about 30 kDa.

Most cellular functions are performed by macromolecular complexes, whose structures are determined by their largely aqueous environments. Techniques capable of determining such structures in solution are therefore very much in demand. Small-angle X-ray and neutron scattering (SAXS and SANS) are two techniques [for a review, see Petoukhov & Svergun (2007)] that provide low-resolution information about the structures of molecules and molecular complexes over a broad range of conditions and particle sizes.

These techniques measure angularly averaged scattered intensities (of X-rays and neutrons) as a function of the magnitude of the momentum transfer, q. Such $I(q)$ curves are strongly peaked in the forward-scattering direction ($q = 0$). However, important information about the scattering particles, such as their molecular mass, radius of gyration, hydration volume and maximum diameter, may be derived directly from the shape of the $I(q)$ distribution (Guinier & Fournet, 1955).

More recently, largely due to the work of Stuhrmann (1970*a*,*b*) and Svergun & Stuhrmann (1991), it has been realized that, remarkably, it may be possible to determine not only directionally independent quantities, such as the above, but even the three-dimensional shape of a dissolved particle *directly* from measured SAXS data. It was shown that the decomposition of the experimental $I(q)$ curve into an expansion of the scattered amplitudes in an angular momentum basis enables the determination of an angular momentum representation of a molecular shape function $F(\omega)$, where ω represents, say, a set of polar (θ) and azimuthal (φ) angles. The limitations of such a single-valued function in representing molecular shapes which may contain complicated features such as multiple subparticles, internal voids *etc.* has led to the development of techniques that represent a complicated particle as a set of dummy atoms or residues on a three-dimensional grid, whose configuration is determined from the experimental data by global optimization techniques, such as genetic algorithms (Chacón *et al.*, 1998) and simulated annealing (Svergun, 1999).

Nevertheless, methods that determine the multipole coefficients of a molecular shape function are mathematically elegant, and remain useful for compact particles of simpler shape. The development of spherical harmonic representations of more complex shapes (Morris *et al.*, 2005) also suggests the possibility of the extension of such methods to such shapes, which may include protein binding pockets. Stuhrmann (1970*a*) showed that the coefficients $A_{lm}(q)$ of a spherical harmonic expansion of the scattered amplitudes may be related to those, $f_{lm}^{(u)}$, of the uth power of $F(\omega)$ *via* a Hankel transform relating the electron density of the molecule and the scattered amplitudes. The radial integrals involved in evaluating the Hankel functions are made tractable by assuming a series representation of the spherical Bessel functions in the Hankel transform. The truncation of this series after a finite number of terms necessarily limits the applicability of the

theory to relatively low values of the momentum transfer q. With the availability of experimental SAXS measurements at higher values of q, it may be desirable to overcome this limitation. We demonstrate in this paper how this may be done in theory by analytic integrations over the entire spherical Bessel functions. We also examine the limits of q beyond which the SAXS intensities calculated by the two approaches deviate. We show that the analytical structure of the radial integrals allows the convenient inclusion in the theory of a hydration shell with a uniform density around a protein molecule of unknown shape in solution. We also demonstrate the use of such an algorithm for the determination of a molecular shape function from both simulated and measured SAXS data.

2. Determination of a molecular shape function directly from SAXS data

The starting point of the existing theory of the shape function (see *e.g.* Svergun & Stuhrmann, 1991) is a representation of both the molecular electron density $\rho^{(\mathrm{p})}(\mathbf{r})$ and scattered X-ray amplitudes $A^{(\mathrm{p})}(\mathbf{q})$ as the multipole expansions

$$\rho^{(\mathrm{p})}(\mathbf{r}) = \sum_L \rho_L^{(\mathrm{p})}(r) Y_L(\omega), \tag{1}$$

where the position vector $\mathbf{r}$ may be specified in terms of polar coordinates (r, ω), and

$$A^{(\mathrm{p})}(\mathbf{q}) = \sum_L A_L^{(\mathrm{p})}(q) Y_L(\Omega), \tag{2}$$

where the photon's momentum transfer vector $\mathbf{q}$ is assumed to have a polar coordinate representation (q, Ω), L (= $\{lm\}$) is a combined index that specifies both azimuthal and magnetic quantum numbers, and Y_L is a spherical harmonic.

Scattering theory suggests that these two sets of expansion coefficients are related by the Hankel transform,

$$A_L^{(\mathrm{p})}(q) = i^l (2/\pi)^{1/2} \int_0^\infty \rho_L^{(\mathrm{p})}(r) j_l(qr) r^2 \, \mathrm{d}r. \tag{3}$$

A shape function $F(\omega)$ may be defined by approximating the particle electron density by

$$\rho^{(\mathrm{p})}(\mathbf{r}) = \begin{cases} 1 & \text{if } 0 \le r \le F(\omega) \\ 0 & \text{if } r \ge F(\omega) \end{cases}, \tag{4}$$

where

$$F(\omega) = \sum_L f_L Y_L(\omega) \tag{5}$$

may be expanded in terms of (complex) multipole coefficients f_L. Then, writing

$$\rho_L^{(\mathrm{p})}(r) = \int \rho^{(\mathrm{p})}(\mathbf{r}) Y_L^*(\omega) \, \mathrm{d}\omega \tag{6}$$

with the shape-function representation [equation (4)] of the electron density, and the representation of the spherical Bessel functions in equation (3) by the power series

$$j_l(qr) = \sum_{s=0}^\infty d_{ls} (qr)^{l+2s} \tag{7}$$

with coefficients

$$d_{ls} = (-1)^s / \{2^s s! [2(l+s)+1]!!\}, \tag{8}$$

the integral in equation (3) may be performed analytically to yield

$$A_L^{(\mathrm{p})}(q) = i^l (2/\pi)^{1/2} \sum_{s=0}^\infty \left[d_{ls} f_L^{(l+2s+3)} / (l+2s+3) \right] q^{l+2s}, \tag{9}$$

where

$$f_L^{(u)} = \int [F(\hat{\mathbf{r}})]^u Y_L^*(\hat{\mathbf{r}}) \, \mathrm{d}\hat{\mathbf{r}} \tag{10}$$

are the multipole coefficients of the uth power of the shape function.

Svergun & Stuhrmann (1991) evaluated only the coefficients with the lowest power, $u = 1$, by numerical integration. Those of higher powers u were computed from these by a recursion relation. The *SASHA* program of Svergun and co-workers (available from http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html) calculates the coefficients f_L of the molecular shape function $F(\omega)$ from an experimental $I(q)$ SAXS curve by optimizing its agreement with a simulated SAXS spectrum,

$$I(q) = [1/(4\pi)^{1/2}] \sum_L |A_L(q)|^2, \tag{11}$$

treating the f_L coefficients as variable parameters.[1]

3. Exact analytic evaluation of the multipole coefficients of the scattered amplitudes

In the following, we show that any approximation of the radial integrals due to a truncation of the Bessel-function power-series expansion above may be circumvented by performing analytical integrals over entire spherical Bessel functions. This is made possible by using the representation of these functions in terms of trigonometric functions. First, note that using equations (4) to (6), equation (3) may be written

$$A_L^{(\mathrm{p})}(q) = i^l (2/\pi)^{1/2} \int R_l[F(\omega)] Y_L^*(\omega) \, \mathrm{d}\omega, \tag{12}$$

where

$$R_l[F(\omega)] = \int_0^{F(\omega)} j_l(qr) r^2 \, \mathrm{d}r. \tag{13}$$

Using the representations (Abramovitz & Stegun, 1974) of spherical Bessel functions

$$j_n(z) = f_n(z) \sin z + (-1)^{n+1} f_{-n-1}(z) \cos z, \tag{14}$$

where the coefficients $f_n(z)$ may be found from the recursion relations

$$f_{n-1}(z) + f_{n+1}(z) = (2n+1) z^{-1} f_n(z), \quad (n = 0, \pm 1, \pm 2, \ldots) \tag{15}$$

[1] The factor of $1/(4\pi)^{1/2}$ in equation (11) is unimportant in practical applications due to a need to scale experiment to theory.

with initial values

$$f_0(z) = z^{-1}, \quad f_1(z) = z^{-2}, \tag{16}$$

the Bessel functions may be written in terms of trigonometric functions. The lowest-order spherical Bessel functions take the form

$$\begin{aligned} j_0(z) &= (\sin z)/z, \\ j_1(z) &= (\sin z)/z^3 - (\cos z)/z, \\ j_2(z) &= \left(3/z^3 - 1/z\right)\sin z - (3/z^3)\cos z, \\ &\ldots, \end{aligned} \tag{17}$$

which allow a convenient analytic evaluation of the radial integrals [equation (13)] to infinite order in the power-series expansion [equation (7)]. Up to angular momentum quantum number $l = 8$, these are

$$R_0[F(\omega)] = (1/q^3)[\sin(qF) - qF\cos(qF)], \tag{18}$$

$$R_1[F(\omega)] = -(i/q^3)[-2 + 2\cos(qF) + qF\sin(qF)], \tag{19}$$

$$R_2[F(\omega)] = -(1/q^3)[qF\cos(qF) - 4\sin(qF) + 3\mathrm{Si}(qF)], \tag{20}$$

$$\begin{aligned} R_3[F(\omega)] = \{i/[(qF)q^3]\}[&8qF - 15\sin(qF) + 7qF\cos(qF) \\ &+ (qF)^2\sin(qF)], \end{aligned} \tag{21}$$

$$\begin{aligned} R_4[F(\omega)] = -\{1/[2(qF)^2q^3]\}[&-105qF\cos(qF) \\ &- 15(qF)^2\mathrm{Si}(qF) + 105\sin(qF) \\ &- 22(qF)^2\sin(qF) + 2(qF)^3\cos(qF)], \end{aligned} \tag{22}$$

$$\begin{aligned} R_5[F(\omega)] = -\{i/[(qF)^3q^3]\}[&-16(qF)^3 - 315qF\cos(qF) \\ &- 105(qF)^2\sin(qF) + 315\sin(qF) + 16(qF) \\ &+ 16(qF)^3\cos(qF) + (qF)^4\sin(qF)], \end{aligned} \tag{23}$$

$$\begin{aligned} R_6[F(\omega)] = \{1/[12(qF)^4q^3]\}[&-20790qF\cos(qF) \\ &+ 1575(qF)^3\cos(qF) - 105(qF)^4\mathrm{Si}(qF) \\ &+ 20790\sin(qF) + 176(qF)^4\sin(qF) \\ &- 8(qF)^5\cos(qF)], \end{aligned} \tag{24}$$

$$\begin{aligned} R_7[F(\omega)] = -\{i/[(qF)^5q^3]\}[&128(qF)^5 \\ &+ 135135(qF)\cos(qF) + 58905(qF)^2\sin(qF) \\ &- 13860(qF)^3\cos(qF) - 1890(qF)^4\sin(qF) \\ &- 135135\sin(qF) + 145(qF)^5\cos(qF) \\ &+ 5(qF)^6\sin(qF)] \end{aligned} \tag{25}$$

and

$$\begin{aligned} R_8[F(\omega)] = -\{1/[16(qF)^6q^3]\}[&-5405400(qF)\cos(qF) \\ &- 2432430(qF)^2\sin(qF) + 630630(qF)^3\cos(qF) \\ &+ 100485(qF)^4\sin(qF) - 10395(qF)^5\cos(qF) \\ &- 315(qF)^6\mathrm{Si}(qF) + 5405400\sin(qF) \\ &- 592(qF)^6 + 16(qF)^7\cos(qF)], \end{aligned} \tag{26}$$

where $\mathrm{Si}(z)$ stands for the sine integral

$$\mathrm{Si}(z) = \int_0^z (\sin t)/t\,\mathrm{d}t, \tag{27}$$

which may be conveniently evaluated by the algorithm developed by MacLeod (1996). Using these closed-form expressions for the radial integrals $R_l[F(\omega)]$ allows the evaluation of the multipole coefficients of the scattered amplitudes by a single angular integral of the form of equation (12), a task which may be performed rapidly and accurately using Gaussian quadrature.

4. Effects of the solvent and hydration shell

It is relatively straightforward within this formalism to model also the effect of a solvent containing the protein molecules, as well as a thin hydration shell around each molecule. We assume a model in which the protein electron density is assumed to be a constant, ρ_p, inside the molecular envelope specified by $F(\omega)$. If ρ_s represents the mean solvent density, the calculated SAXS signal $I(q)$ from protein molecules in the solution is the angular average

$$\begin{aligned} I_\mathrm{calc}(q) &= \langle|(\rho_\mathrm{p} - \rho_\mathrm{s})A^{(\mathrm{p})}(\mathbf{q})|^2\rangle_\Omega \\ &= [1/(4\pi)^{1/2}]\sum_L |(\rho_\mathrm{p} - \rho_\mathrm{s})A_L^{(\mathrm{p})}|^2, \end{aligned} \tag{28}$$

where the second equality follows from the orthonormality of spherical harmonics with respect to integration over the solid angle Ω (Svergun *et al.*, 1995).

If the protein molecules are surrounded by a hydration shell of density ρ_h, intermediate between that of the protein and solvent, this expression is modified to

$$\begin{aligned} I_\mathrm{calc}(q) &= \langle|(\rho_\mathrm{p} - \rho_\mathrm{s})A^{(\mathrm{p})}(\mathbf{q}) + (\rho_\mathrm{h} - \rho_\mathrm{s})A^{(\mathrm{h})}(\mathbf{q})|^2\rangle_\Omega \\ &= [1/(4\pi)^{1/2}]\sum_L |(\rho_\mathrm{p} - \rho_\mathrm{s})A_L^{(\mathrm{p})} + (\rho_\mathrm{h} - \rho_\mathrm{s})A_L^{(\mathrm{h})}|^2, \end{aligned} \tag{29}$$

where $A^{(\mathrm{h})}(\mathbf{q})$ is the amplitude of the scattering from the hydration shell and $A_L^{(\mathrm{h})}$ is its multipole coefficient.

In order to include the protein and its hydration shell within the envelope defined by the molecular shape function $F(\omega)$, the scattered amplitude from the protein may be evaluated from

$$A_L^{(\mathrm{p})}(q) = i^l(2/\pi)^{1/2}\int R_l[F(\omega) - \Delta/\gamma]Y_L^*(\omega)\,\mathrm{d}\omega, \tag{30}$$

with the upper limit on the radial integral [equation (13)] now $F(\omega) - \Delta/\gamma$ rather than $F(\omega)$, where Δ is the diameter of the hydration shell (assumed to be 3 Å) and γ is the cosine of the angle between the radius vector and the normal to the outer

Table 1
Values of the complex coefficients f_{lm} of the multipole expansion of the shape function $F(\omega)$.

The complex coefficients are listed for different values of l and non-negative m up to $l = 4$ and were used for the calculations of the SAXS spectra of Figs. 1–4.

l	m	f_{lm} (real)	f_{lm} (imaginary)
0	0	2.529218	0.000000
1	0	−0.376664	0.000000
1	1	0.159247	−0.129320
2	0	0.294870	0.000000
2	1	0.976603	−0.016234
2	2	0.204422	0.017612
3	0	0.063160	0.000000
3	1	0.092240	0.054767
3	2	0.121282	0.194941
3	3	−0.050339	0.250137
4	0	0.030176	0.000000
4	1	0.054744	−0.012466
4	2	0.167895	−0.081864
4	3	0.119843	−0.003168
4	4	0.135816	−0.110773

surface of the hydration shell. The scattered amplitude from the hydration shell is defined by

$$A_L^{(\mathrm{h})}(q) = i^l (2/\pi)^{1/2} \int \{R_l[F(\omega)] - R_l[F(\omega) - \Delta/\gamma]\} Y_L^*(\omega)\, \mathrm{d}\omega. \tag{31}$$

Together with the analytical radial integrals [equations (18) to (27)], the last three equations enable the calculation of the SAXS intensity from a model of the protein of electron density ρ_p, surrounded by a hydration shell of electron density ρ_h of thickness 3 Å and immersed in a solvent of electron density ρ_s, as implemented in the *CRYSOL* program of Svergun *et al.* (1995). This differs from the model assumed by the *SASHA* program, in which the density of the hydration shell reduces linearly from its inner surface to its outer surface (Svergun, 1997).

We next proceed to demonstrate the use of our algorithm for two tasks. In §5 we calculate the SAXS spectrum expected for a given molecular shape function, as characterized by a given set of multiple coefficients, by our method and by the program of Svergun and co-workers. We demonstrate that the two methods of calculation agree well in this case up to a value of q of about 0.2 Å^{-1}, beyond which the results of the two programs differ somewhat. Finally, a little beyond $q = 0.6$ Å^{-1}, the term-by-term integration of a truncated spherical Bessel function power series yields a divergent SAXS signal, whereas our analytical radial integration over entire Bessel functions appears to produce convergent values (as judged by the smoothness of the simulated curve). In §6, we show how our approach allows a recovery of a molecular shape function from a SAXS spectrum.

5. Simulation of a SAXS spectrum for a given molecular shape function

As a first test of our theory, we compared SAXS spectra calculated by our method and by the power-series expansion method for a molecular shape function specified by different sets of multipole coefficients f_{lm} from a file sent to us by Dr Dmitri Svergun. Since the quantity $F(\omega)$ involves the sum of terms of the form $f_{lm}Y_{lm}(\omega) + f_{l-m}Y_{l-m}(\omega)$, and since $Y_{l-m}(\omega) = (-1)^m Y_{lm}^*(\omega)$, the reality of $F(\omega)$ is ensured by the choice of $(-1)^m f_{l-m} = f_{lm}^*$, so only the coefficients f_{lm} with positive m need be specified. Table 1 lists the first 15 of these independent (complex) coefficients for values of l up to 4.

The calculations were performed both by a routine sent to us by Dr Svergun (part of the *SASHA* program) in which the radial integral [equation (3)] is performed by the power-series expansion method of §2, and by the analytical radial integration method described in §3. The spectra calculated by the two methods were scaled to coincide at $q = 0$. Both calculations assumed a 3 Å-thick hydration shell (although calculated on slightly different models, as described in the previous section).

Fig. 1 compares the SAXS spectra calculated from the f_{lm} coefficients in Table 1 for l values up to a maximum $l_\mathrm{max} = 2$, and for a range of q from 0 to 0.2 Å^{-1}. The spectra calculated by the two methods are essentially indistinguishable in this case.

The same holds when $l_\mathrm{max} = 4$ for the same range of q, as illustrated in Fig. 2.

The situation is different when the results of the two methods of calculation are compared in the range of q from 0.2 to 0.8 Å^{-1}. For $l_\mathrm{max} = 2$ (Fig. 3), small deviations in the results of the two methods are noticeable in the range from $q =$ 0.3 Å^{-1} onwards, while for $l_\mathrm{max} = 4$ (Fig. 4), deviations are noticeable closer to $q = 0.2$ Å^{-1}. For $l_\mathrm{max} = 2$, the power-series expansion method is completely divergent by about $q =$ 0.8 Å^{-1}, while for $l_\mathrm{max} = 4$ the divergence is apparent by about $q = 0.7$ Å^{-1}; the results of our method do not show such a divergence in either case.

6. Calculation of the molecular shape function from experimental SAXS data

The use of our analytical expressions for the radial integrals involved in the evaluation of the scattered amplitudes from the molecules and their hydration shells allows us to find information about both the shape of a dissolved molecule as well as its hydration shell by optimizing the agreement of the calculated SAXS spectrum with measured SAXS data $I_\mathrm{exp}(q)$ (say) by minimizing a function of the form

$$\chi^2 = \sum_q |I_\mathrm{calc}(q) - I_\mathrm{exp}(q)|^2/\sigma_q^2 \tag{32}$$

[where σ_q is the estimated error in the measurement of $I_\mathrm{exp}(q)$] with respect to the unknown parameters f_L specifying the molecular shape, and to the parameter Δ, specifying the thickness of the hydration shell. For the mean densities of the protein molecule, hydration shell and solvent density we assumed the accepted approximate values of 0.40, 0.36 and 0.33 e Å^{-3}, respectively, and we assumed the thickness of the hydration shell, Δ, to be 3 Å. The optimization was performed by means of a simulated annealing algorithm (Kirkpatrick *et al.*, 1983), starting with random values of the real and

imaginary components of the complex quantities f_L within reasonable physical ranges.

7. Comparison of calculated molecular shapes with those from *SASHA*

Figs. 3 and 4 suggest that the SAXS intensities calculated by the *SASHA* program and ours may differ significantly, especially for higher values of q. *SASHA* is widely used for the calculation of molecular shapes from SAXS spectra. Therefore it is of interest to compare the molecular shapes calculated from the same SAXS spectra by *SASHA* and by our method. For our test data, we took the experimental SAXS data (in the file 'lyzexp.dat') for lysozyme distributed as a test data set for *SASHA* from Dr Svergun's web site at http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html. This data set contained data from very close to $q = 0$ up to about $q = 0.5$ Å^{-1}.

The molecular shape functions calculated by *SASHA* from these SAXS data for the ranges (i) $q = 0$ to 0.25 Å^{-1} and (ii) $q = 0$ to 0.50 Å^{-1}, and displayed using the graphics program *MASSHA* (also distributed from the same web site), are shown in Figs. 5 and 6, respectively.

In both cases, the molecular model was displayed with its longest axis lengthwise across the page, and the molecular shape was rotated and translated in order to best fit the model. The molecular envelope calculated from the smaller data range was found to fit easily around the backbone trace of the

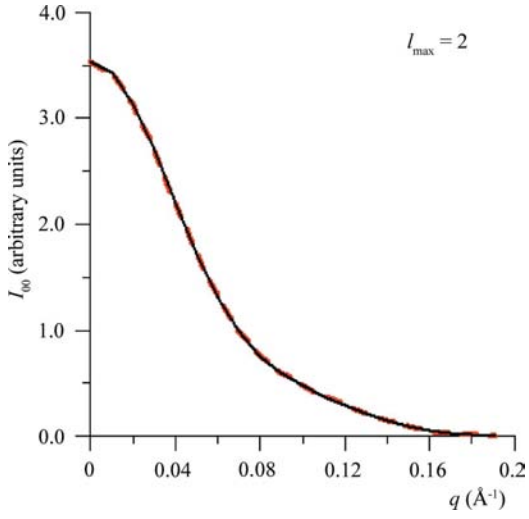

Figure 1
Comparison of SAXS spectra for q in the range 0 to 0.2 Å^{-1}, calculated from the f_{lm} coefficients of a molecular shape function up to $l_{max} = 2$ by radial integrals involving (*a*) term-by-term integration of a truncated power-series expansion of spherical Bessel function (dashed red line) and (*b*) analytical integration of the entire Bessel functions (solid black line).

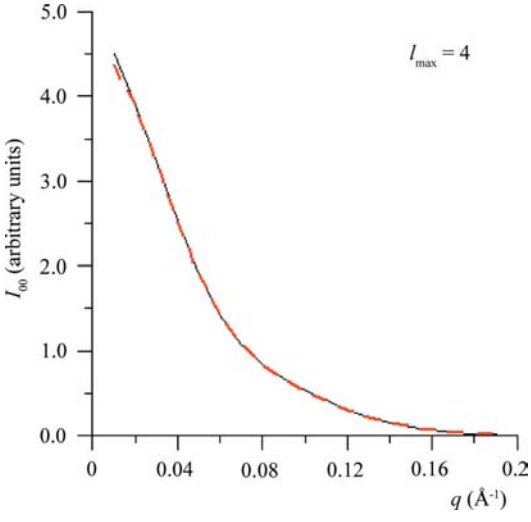

Figure 2
As for Fig. 1, except that $l_{max} = 4$.

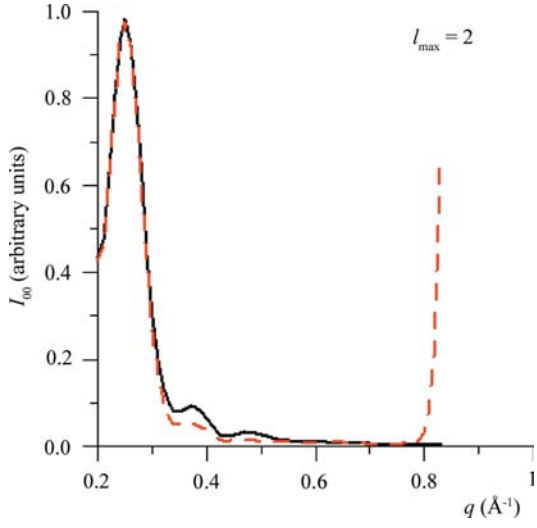

Figure 3
As for Fig. 1, except q is in the range 0.2 to 0.8 Å^{-1}. The intensity units should be multiplied by a factor of 10^{-2} to make them consistent with the scale of Fig. 1.

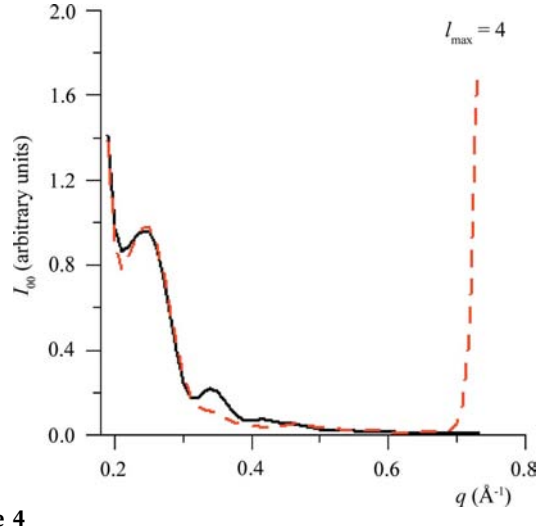

Figure 4
As for Fig. 2, except q is in the range 0.2 to ~0.8 Å^{-1}. The intensity units should be multiplied by a factor of 10^{-2} to make them consistent with the scale of Fig. 2.

molecule from the 2BPU entry in the Protein Data Bank (Fig. 5). However, when the molecular envelope was calculated from the larger q range, it was impossible to find an orientation of the envelope to completely cover the molecule (Fig. 6).

We then used instead the theory based on the analytical radial integrals, as described in §3, to find the optimum values of the the molecular envelope coefficients f_{lm} from the same experimental data. In this case, the resulting envelopes were found to completely cover the molecule for both ranges of q, as may be seen in Figs. 7 and 8.

8. Discussion and conclusions

A major advance in the analysis of SAXS spectra was the realization that not only traditional spherically averaged properties of a dissolved molecule, such as its radius of gyration, may be deduced from a SAXS spectrum, but even the shape of the molecule (Svergun & Stuhrmann, 1991). This method of analysis depends on the representation of the molecule by a shape function characterized by a set of spherical harmonic expansion coefficients. By comparing SAXS curves calculated from 25 different proteins with molecular masses between 10 and 300 kDa, Svergun & Koch (2002) concluded that data up to a momentum transfer q of 0.7 Å^{-1} (corresponding to a spatial resolution larger than $2\pi/q \simeq 9$ Å), where SAXS curves manifest their greatest variability, are most sensitive to the molecular shape. Although SAXS curves show less variability amongst different proteins in such a range, some information about the overall molecular fold may be gleaned from data in the range of q between about 0.6 and 1.2 Å^{-1} (resolution between ~10 and 5 Å). Information about the secondary structure of proteins may be obtained only from ranges of q greater than about 1.2 Å^{-1} (resolutions smaller than about 5 Å), where SAXS spectra are similar for a wide range of proteins. [A discussion of the information that can be obtained from different ranges of q in protein crystallography has been given by Morris *et al.* (2004).]

In this paper, we are concerned mainly about the portion of a SAXS spectrum with the greatest variability amongst different proteins, namely that with q less than about 0.7 Å^{-1}, from which information may be deduced about the molecular shape. Deduction of spherical harmonic expansion coefficients of the molecular shape function requires the integration of a radial function with a spherical Bessel function kernel. The standard method of calculating this integral involves a term-by-term integration of terms derived from a power-series expansion of the Bessel function (*e.g.* Svergun & Stuhrmann, 1991). The necessary truncation of such a series is a potential source of error, which would be expected to increase with an increase of the momentum transfer q.

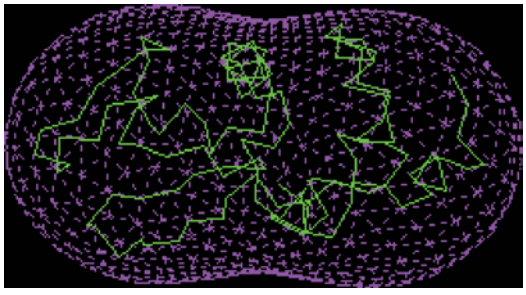

Figure 5
Molecular shape function of lysozyme calculated *via* the expansion coefficients f_{lm} of its spherical harmonic expansion by the executable code *SASHA* distributed by Svergun and co-workers from the experimental data file 'lyzexp.dat' distributed with the code for the range $q = 0$ to 0.25 Å^{-1}. The maximum value of l was taken to be $l_{max} = 2$. The molecular envelope is displayed by the graphics program *MASSHA*, also from the same web site. The molecular envelope is compared with the backbone trace of lysozyme from the protein data bank (PDB) entry 2BPU.

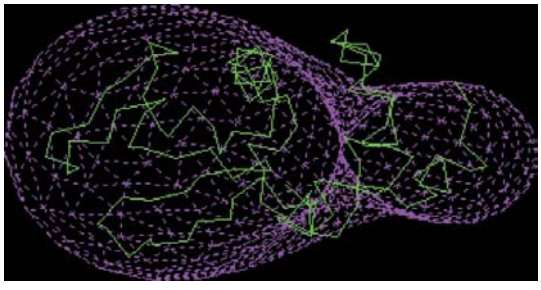

Figure 6
As for Fig. 5 except that the molecular shape calculation was from the full data range $q = 0$ to 0.5 Å^{-1}.

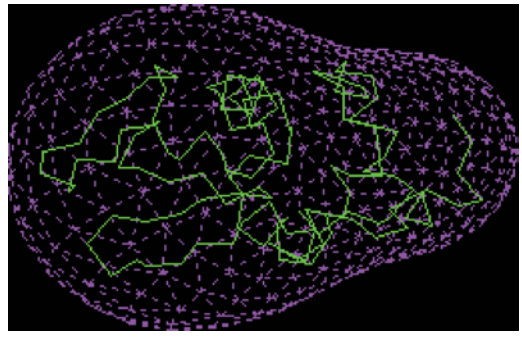

Figure 7
As for Fig. 5 except that the calculation was performed with the SAXS intensity simulated with the use of the analytic radial integrals described in §3.

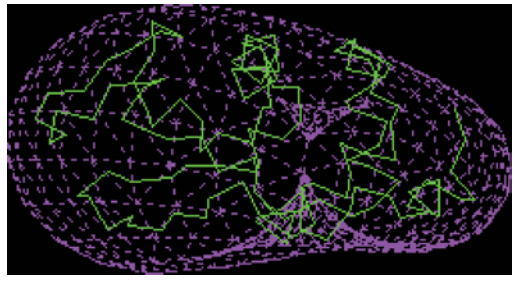

Figure 8
As for Fig. 6 except that the calculation was performed with the SAXS intensity simulated with the use of the analytic radial integrals described in §3.

We show in this paper that truncation of such a series expansion may be avoided by performing the radial integrals analytically using the trigonometric representation of the spherical Bessel function, and then by performing the angular integrals involving spherical harmonics by quadrature, a procedure which requires relatively few evaluations of the angular integrands. Incorporation of different electron densities of the molecular solvent and hydration shell is also straightforward with this approach. We suggest that the proposed method may be capable of recovering low-resolution molecular shapes from SAXS data of a wider range of q. Of course, the larger the value of q, the greater is the contribution from the internal structure of a molecule (Svergun, 1994), an effect not modelled in the present treatment. Nevertheless, as we point out, there are relatively small, but noticeable, deviations in calculations of SAXS spectra by our method and by the prior method even for q values as small as about 0.2 or 0.3 Å^{-1}. Molecular shapes calculated even from SAXS data in the range 0 to 0.5 Å^{-1} by the two methods appear to be noticeably different.

We thank Dr Dmitri Svergun for providing us with his computer routine for calculating a SAXS spectrum from a set of multipole coefficients f_{lm} specifying a molecular shape function (part of his *SASHA* program). We are also grateful to Dr Svergun for helpful comments on an earlier draft of this paper. We thank Professor Michael Weinert for providing his program for performing angular integrals over spherical harmonics by quadrature. We acknowledge support from the US Department of Energy (grant Nos. DE-FG02-84ER45076 and DE-FG02-06ER46277).

References

Abramovitz, M. & Stegun, I. A. (1974). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, p. 438. New York: Dover.
Chacón, P., Morán, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). *Biophys. J.* **74**, 2760–2775.
Drenth, J. (1994). *Principles of Protein X-Ray Crystallography*. New York: Springer-Verlag.
Guinier, A. & Fournet, G. (1955). *Small Angle Scattering of X-Rays*. New York: Wiley and Sons.
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
MacLeod, A. J. (1996). *Numer. Algorithms*, **12**, 259–272.
Morris, R. J., Blanc, E. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 227–240.
Morris, R. J., Najmanovich, R. J., Kahraman, A. & Thornton, J. M. (2005). *Bioinformatics*, **21**, 2347–2355.
Petoukhov, M. V. & Svergun, D. I. (2007). *Curr. Opin. Struct. Biol.* **17**, 562–571.
Stuhrmann, H. B. (1970*a*). *Z. Phys.* **72**, 177–184.
Stuhrmann, H. B. (1970*b*). *Acta Cryst.* A**26**, 297–306.
Svergun, D. I. (1994). *Acta Cryst.* A**50**, 391–402.
Svergun, D. I. (1997). *J. Appl. Cryst.* **30**, 792–797.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D., Barberato, C. & Koch, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
Svergun, D. I. & Koch, M. H. (2002). *Curr. Opin. Struct. Biol.* **12**, 654–660.
Svergun, D. I. & Stuhrmann, H. B. (1991). *Acta Cryst.* A**47**, 736–744.
Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. New York: Wiley.