Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 9 May 2012
Accepted 28 June 2012

Computation of fluctuation scattering profiles *via* three-dimensional Zernike polynomials

Haiguang Liu,[a]‡ Billy K. Poon,[a] Augustus J. E. M. Janssen[b] and Peter H. Zwart[a]*

[a]Physical Biosciences Division, Lawrence Berkeley National Laboratories, One Cyclotron Road, Berkeley, CA 94720 USA, and [b]Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, 5600MB Eindhoven, The Netherlands. Correspondence e-mail: PHZwart@lbl.gov

Ultrashort X-ray pulses from free-electron laser X-ray sources make it feasible to conduct small- and wide-angle scattering experiments on biomolecular samples in solution at sub-picosecond timescales. During these so-called fluctuation scattering experiments, the absence of rotational averaging, typically induced by Brownian motion in classic solution-scattering experiments, increases the information content of the data. In order to perform shape reconstruction or structure refinement from such data, it is essential to compute the theoretical profiles from three-dimensional models. Based on the three-dimensional Zernike polynomial expansion models, a fast method to compute the theoretical fluctuation scattering profiles has been derived. The theoretical profiles have been validated against simulated results obtained from 300 000 scattering patterns for several representative biomolecular species.

1. Introduction

The overwhelming majority of known protein structures at the atomic scale have been determined by X-ray crystallography. The availability of these high-resolution structural models is the foundation of a deeper understanding of fundamental processes in biology (Orengo *et al.*, 1999), and the development of new therapeutic drugs (Schneider & Fechner, 2005) and novel classes of nano-materials (Barth *et al.*, 2005). Owing to difficulties in obtaining crystals of large macromolecular complexes, like membrane proteins or large molecular machines, X-ray crystallography is rarely the technique of choice for deriving structure information of such biomolecules or complexes (Miao *et al.*, 2008). Current techniques used to study large macromolecular complexes, like electron microscopy (Medalia *et al.*, 2002; Frank, 2002), derive structural information of complexes in non-native environments and do not easily allow for the investigation of time-dependent large-scale structural changes. Techniques like small- and wide-angle X-ray scattering (SAXS/WAXS) are suitable and allow for time-resolved studies, but have the drawback that the data have relatively low information content (Volkov & Svergun, 2003). A possible route for increasing the information content in solution scattering experiments while monitoring the large-scale structural changes of macromolecular complexes in aqueous environments is fluctuation X-ray scattering (fXS), a method proposed by Kam in the late 1970s (Kam, 1977; Kam *et al.*, 1981). A fluctuation X-ray scattering experiment is performed by collecting scattering patterns of a dilute sample of scatterers at exposure times shorter than the time required for particles to reorient themselves *via* rotational diffusion. These experiments are ideally performed on free-electron lasers such as the Linac Coherent Light Source (LCLS; Emma *et al.*, 2010), the European X-ray Free Electron Laser (XFEL; Vartanyants *et al.*, 2007) or the (to be constructed) Next Generation Light Source (NGLS; Schoenlein, 2011). Under the aforementioned conditions, the measured scattering patterns will no longer be angularly isotropic, but will contain speckles. It can be shown that a large number of these scattering patterns can be used to estimate the average angular autocorrelation of the scattering pattern of a single particle (Kam *et al.*, 1981; Saldin *et al.*, 2010, 2011).

Earlier work by Saldin *et al.* (2009) has shown that the angular autocorrelations can be used in the determination of low-resolution structural envelopes in a manner similar to which SAXS data are used to determine structural envelopes (Saldin *et al.*, 2009). By parameterizing a macromolecular envelope with spherical harmonics, a straightforward procedure is obtained that allows for the optimization of real-space expansion coefficients given observed angular correlations. Although the spherical-harmonics expansion used to parameterize the macromolecular structure has been widely used, its main drawback is that it cannot properly describe shapes containing cavities or pores. A more appropriate polynomial expansion, capable of describing complex shapes not limited to star-shaped objects, can be obtained by using three-dimensional Zernike polynomials (Novotni & Klein, 2003; Mak *et al.*, 2008; Liu, Morris *et al.*, 2012). Three-dimensional Zernike polynomials are extensions of the well known two-dimensional Zernike polynomials frequently used in

‡ Current address: Physics Department, Arizona State University, Tempe, AZ 85287, USA.

describing optics (Wang & Silva, 1980) and have similar properties. A basic introduction to three-dimensional Zernike polynomials can be found by Canterakis (1999).

Three-dimensional Zernike polynomials are useful tools in solution scattering because with a relatively low expansion order (20, for instance) one is able to faithfully reproduce a macromolecular shape and its associated scattering function. This fact has lead to novel ways to compute SAXS profiles (Liu, Morris *et al.*, 2012) and accelerated shape recovery from SAXS data (unpublished). In this article we report an economic route of computing the two-point angular autocorrelations, $C_2(q, \Delta\varphi)$, from the Zernike expansion coefficients. This work is a first step towards *ab initio* structure solution from fluctuation scattering data at low to medium resolution ranges. In a forthcoming paper, details on how to utilize this method in three-dimensional model reconstruction will be reported.

2. Methods

The average angular autocorrelation can be extracted from N experimental patterns, acquired in femtosecond X-ray scattering experiments, *via*

$$C_2(q, \Delta\varphi) = \frac{1}{N}\sum_i \sum_\varphi I_i(q, \varphi) I_i(q, \varphi + \Delta\varphi), \qquad (1)$$

where $I_i(q, \varphi)$ is the scattering intensity of pattern i at a pixel position corresponding to reciprocal space point (q, φ). Kam *et al.* (1981) have shown that, given a sufficient number of scattering patterns, this autocorrelation will converge to a fixed value. The expected value of $C_2(q, \Delta\varphi)$ can be computed from the associated molecular structure as outlined below.

2.1. Three-dimensional Zernike polynomial models

The three-dimensional Zernike model is a compact description of three-dimensional models using convenient orthogonal polynomials. After scaling down to fit in a unit sphere, the electron density of a three-dimensional model can be represented by

$$\rho(\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{l=0}^{n}\sum_{m=-l}^{l} c_{nlm} R_{nl}(r) Y_{lm}(\omega_r), \qquad (2)$$

in which $\mathbf{r}$ is a three-dimensional vector (r, ω_r). The original three-dimensional model can be obtained by scaling up the unit sphere by $r_{\max}$, the radius of a sphere circumscribing the model. $R_{nl}(r)$ is the three-dimensional Zernike polynomial radial function, and the orders l and n have the same parity, *i.e.* $(n - l)$ is an even number.

The expansion coefficients, c_{nlm}, can be computed *via* the following procedure: the three-dimensional model is firstly scaled down to fit into a unit sphere by dividing $r_{\max}$ to obtain the scaled density distribution function, $\rho(\mathbf{r})$; and then the c_{nlm} coefficients can be computed utilizing the orthogonality,

$$c_{nlm} = \frac{3}{4\pi}\int_{|\mathbf{r}|<1} \rho(r) R_{nl}(r) Y^*_{lm}(r)\, \mathrm{d}r. \qquad (3)$$

Novotni & Klein (2003) have derived a more efficient way of computing the Zernike moments using the geometry moments, instead of direct integration.

It is worthwhile pointing out that the computation of the Zernike moments, c_{nlm}, is the time-limiting step. The details of the implementation can be found in the original papers (Novotni & Klein, 2003; Mak *et al.*, 2008). In short, the Protein Data Bank (PDB) models are mapped to three-dimensional grids to generate a voxelized object, and from which the c_{nlm} are subsequently calculated. The grid spacing is set to 0.7 Å, so the number of non-zero voxels is proportional to the number of atoms. For very large molecular systems, the grid size can be increased to speed up the computation. This makes the Zernike approximation advantageous over other methods, especially when high-resolution data are not required or available.

Following Liu, Morris *et al.* (2012) and Appendix *A*, the Fourier transform of this model is equal to

$$A(\mathbf{q}) = 4\pi \sum_{n}^{\infty}\sum_{l}^{n}\sum_{m=-l}^{+l} i^l (-1)^{(n-l)/2} c_{nlm} Y^*_{lm}(\omega_q) b_n(qr_{\max}) \qquad (4)$$

with

$$b_n(qr_{\max}) = \frac{j_n(qr_{\max}) + j_{n+2}(qr_{\max})}{2n+3}. \qquad (5)$$

We now aim to expand $A(\mathbf{q})$ in shells of fixed q, and only vary the angles,

$$\begin{aligned} A_q(\omega_q) &= 4\pi \sum_{n}^{\infty}\sum_{l}^{n}\sum_{m=-l}^{+l} i^l (-1)^{(n-l)/2} b_n(qr_{\max}) c_{nlm} Y^*_{lm}(\omega_q) \\ &= 4\pi \sum_{n}^{\infty}\sum_{l}^{n}\sum_{m=-l}^{+l} w_{nl}(qr_{\max}) c_{nlm} Y^*_{lm}(\omega_q), \end{aligned} \qquad (6)$$

$$w_{nl}(qr_{\max}) = i^l (-1)^{(n-l)/2} b_n(qr_{\max}). \qquad (7)$$

If we now regroup constants and set the maximum expansion order to $n_{\max}$, we obtain

$$A_q(\omega_q) = 4\pi \sum_{l}^{n_{\max}}\sum_{m=-l}^{+l} Y^*_{lm}(\omega_q) \sum_{n}^{n_{\max}} w_{nl}(qr_{\max}) c_{nlm} \qquad (8)$$

or

$$\begin{aligned} A_q(\omega_q) &= 4\pi \sum_{l}^{n_{\max}}\sum_{m=-l}^{+l} a_{lm}(qr_{\max}) Y^*_{lm}(\omega_q) \\ a_{lm}(qr_{\max}) &= \sum_{n}^{n_{\max}} w_{nl}(qr_{\max}) c_{nlm}. \end{aligned} \qquad (9)$$

The outlined route to compute coefficients a_{lm} using available Zernike moments can be seen to depend linearly on expansion order $n_{\max}$. This is in contrast to the traditional method (Kam, 1977) when a_{lm} is computed directly from the atomic coordinates (Liu, Morris *et al.*, 2012; expression 6) and depends linearly on the number of atoms N_{atoms},

$$a_{lm}(q) = \sum_{j=1}^{N_{\rm atoms}} f_j(q) j_l(qr_j) Y^*_{lm}(\omega_j). \quad (10)$$

In the following subsections, the expressions of Kam (1977) and Saldin *et al.* (2009) are closely followed in order to provide the reader with a complete overview of subsequent steps required to compute fluctuation scattering curves.

2.2. Intensities

The relation between the Zernike coefficients and the complex expansion coefficients for the structure factor at any fixed resolution, q, can be utilized to obtain spherical-harmonics-based expansion coefficients for the intensity at this fixed resolution,

$$\begin{aligned} I_q(\omega_q) &= A^*_q(\omega_q) A_q(\omega_q) \\ &= 16\pi^2 \sum_{l}^{n_{\rm max}} \sum_{m=-l}^{+l} \sum_{l'}^{n_{\rm max}} \sum_{m'=-l'}^{+l'} a^*_{l'm'}(qr_{\rm max}) a_{lm}(qr_{\rm max}) \\ &\quad \times Y^*_{lm}(\omega_q) Y_{l'm'}(\omega_q). \end{aligned} \quad (11)$$

Although the above expression is not overtly complicated, we need to re-expand it in a spherical-harmonics series,

$$I_q(\omega_q) = 16\pi^2 \sum_{l}^{n_{\rm max}} \sum_{m=-l}^{+l} I_{lm}(q) Y_{lm}(\omega_q). \quad (12)$$

The expansion coefficients, $I_{lm}(q)$, can be expressed as a function of $a_{lm}(qr_{\rm max})$ *via* a Gaunt series (Gaunt, 1928), using

$$\int Y_{lm} Y_{l'm'} Y_{l''m''}\, {\rm d}\Omega = G^{mm'm''}_{ll'l''} \quad (13)$$

with

$$\begin{aligned} G^{mm'm''}_{ll'l''} &= (-1)^m [(2l+1)(2l'+1)(2l''+1)/4\pi]^{1/2} \\ &\quad \times \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (14)$$

One thus obtains

$$I_{lm}(q) = \sum_{l'} \sum_{l''} \sum_{m'} \sum_{m''} a_{l'm'}(qr_{\rm max}) a_{l''m''}(qr_{\rm max}) G^{mm'm''}_{ll'l''}, \quad (15)$$

where

$$\begin{pmatrix} l & l' & l'' \\ m & m' & m'' \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}$$

are Wigner 3*j*-symbols.

The latter expansion facilitates easy computation of the autocorrelation (Saldin *et al.*, 2009) as shown in the following section.

2.3. The expansion of autocorrelations

According to Saldin *et al.* (2009) and following their notation, the correlation function, C_2, at a given q value can be expressed as a weighted sum of Legendre polynomials,

$$C_{2,q}(\Delta\varphi) = \sum_l F_l(\Delta\varphi) B_l(q), \quad (16)$$

with

$$\begin{aligned} F_l(\Delta\varphi) &= \frac{1}{4\pi} P_l\big[\cos^2\theta(q) + \sin^2\theta(q)\cos(\Delta\varphi)\big], \\ B_l(q) &= \sum_m |I_{lm}(q)|^2, \end{aligned} \quad (17)$$

where $P_l(\ldots)$ denotes a Legendre polynomial and

$$\theta(q) = \pi/2 - \sin^{-1}(q/2\kappa). \quad (18)$$

κ is equal to the wavenumber $2\pi/\lambda$ with λ the wavelength of the incident radiation. The expansion coefficients $B_l(q)$ can thus be computed directly from the Zernike moments by considering relations (9), (15) and (17).

3. Results

In this section the theoretical fXS profiles calculated using the three-dimensional Zernike polynomial method are compared with the results calculated using the spherical-harmonics approach, as well as the fXS profiles extracted from simulated scattering patterns of single molecules. After the validation of the method, we will show the relation between the resolution and the Zernike expansion order. The computing complexity and speed will also be discussed.

3.1. Validation

In order to verify the above expressions, scattering patterns of proteins in random orientations were simulated up to q values of 0.5 Å^{-1} as described in Appendix *B*. Besides model data generated *via* expressions (9), (15) and (17), $B_l(q)$ were computed *via* a spherical-harmonics expansion of the intensities (Saldin *et al.*, 2009). Using three different example proteins, the fXS profiles obtained from three distinct approaches are compared in Fig. 1. The $B_l(q)$ curves obtained with the Zernike expansion are in excellent agreement with the curves calculated using the spherical-harmonics approach, revealing the theoretical fXS profiles under ideal conditions (noise-free with infinitely small pixel size). The agreement between the Zernike-based method and the data obtained *via* numerical simulation is satisfactory at low resolution but quickly becomes worse as the resolution increases. The discrepancy between the theoretical fXS profiles and the data extracted from the simulated diffraction patterns might be due to the finite resolution of the detectors (see §4).

3.2. Expansion order

The effect of the three-dimensional Zernike expansion order on the calculated data is shown in Fig. 2. The truncation effects depicted in Fig. 2 are comparable with those seen when computing SAXS profiles as described by Liu, Morris *et al.* (2012). As described in the previous study, the resolution at which truncation ripples appear is largely a function of the expansion order, $n_{\rm max}$, and size, $R_{\rm max}$, of the particle owing to the scale-free nature of the Zernike polynomials (Novotni & Klein, 2003; Mak *et al.*, 2008; Liu, Morris *et al.*, 2012).

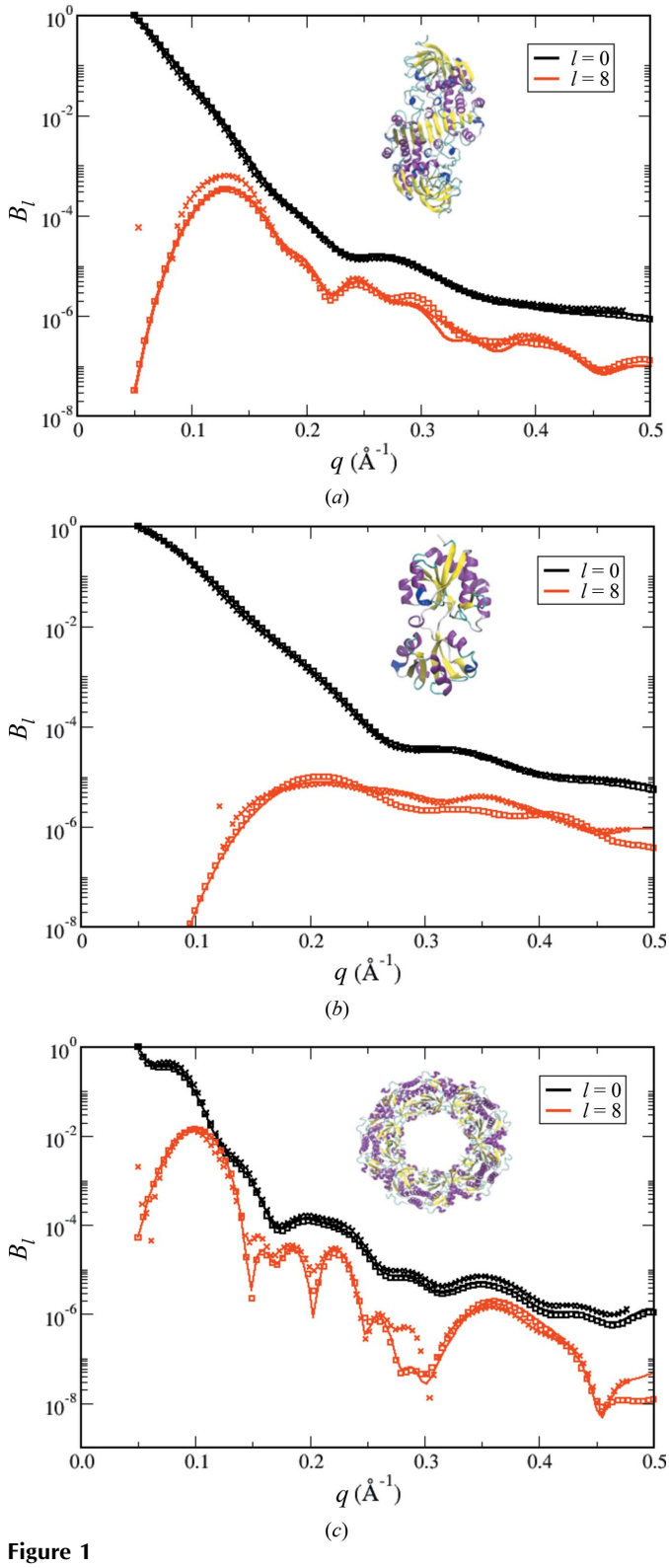

Figure 1
Comparisons of fluctuation scattering profiles. The fXS profiles are calculated using three methods: the Zernike-based method (solid curve), spherical-harmonics approach (open squares) and extracted from simulations (crosses). For clarity, only the l = 0 and l = 8 curves are shown as representatives here. Curves depicted are for models (*a*) 1ee2, (*b*) 2lao and (*c*) 2e2g. The theoretical curves from the Zernike-based method agree with the spherical-harmonics-based method. In spite of discrepancies between the simulations results and the theoretical curves, the main features are preserved. The expansion order $n_{\max}$ was set to 40.

3.3. Complexity and timing

As discussed elsewhere (Liu, Morris *et al.*, 2012), the complexity of the Zernike method for computing SAXS data scales linearly with the number of atoms (N) and is separately dependent on the number of data points computed (M); therefore, the computing complexity is $O(M + N)$. As mentioned, the time-limiting step is the computation of the Zernike moments, and the dependence on M is due to the precomputation of spherical Bessel functions. The resulting computing complexity can be approximated by $O(N)$, because the computation of Bessel functions only compromises a small overhead on the total run time. On the other hand, the spherical-harmonics-based method depends linearly on the number of atoms and the number of data points computed, resulting in a complexity of $O(NM)$. For computing fluctuation scattering data, both the spherical-harmonics-based method and the Zernike-based method show a linear dependence on the number of data points and number of atoms (Fig. 3). The reason for the linear dependence when computing $B_l(q)$ curves using the Zernike method can be seen from (9), in which a q-dependent weight is computed. As can be seen from (9), the three-dimensional Zernike moments c_{nlm} need only be computed once for a given model. In contrast, the q-dependent expansion coefficients, $a_{lm}(q)$, as obtained from a spherical-harmonics expansion [see equation (10)], depends on the number of atoms. This reduction in numerical complexity provides a significant time advantage in the calculation of $B_l(q)$ coefficients *via* the Zernike method. These effects are illustrated in Fig. 3.

4. Discussions and conclusions

The results from Fig. 1 indicate that the proposed route for computing model fXS data is effective as the curves obtained *via* simulation, spherical harmonics and the Zernike-based method are all in agreement. The main benefit of the proposed method is the gain in computational efficiency, a prerequisite for real-time structure refinement or iterative real–reciprocal-space *ab initio* structure determination methods.

Although a notable difference between simulated and calculated data is found, especially at higher q values, the discrepancy can be attributed to smearing of data owing to finite pixel size as well as sampling errors and other experimental phenomena.

The timing results indicate that the Zernike-based method is more efficient than the spherical-harmonics-based method because a time-consuming step that depends on the number of atoms (calculating three-dimensional Zernike moments) is performed only once. For the spherical-harmonics-based method, a loop over all atoms has to be performed for each data point, making this method more computationally costly. This is especially critical when applying a local perturbation approach for structural reconstruction or refinement, where only part of the system will be changed at a time. In this case, since the Zernike moments are a summation of contributions from all parts, only the perturbed part of the system needs to

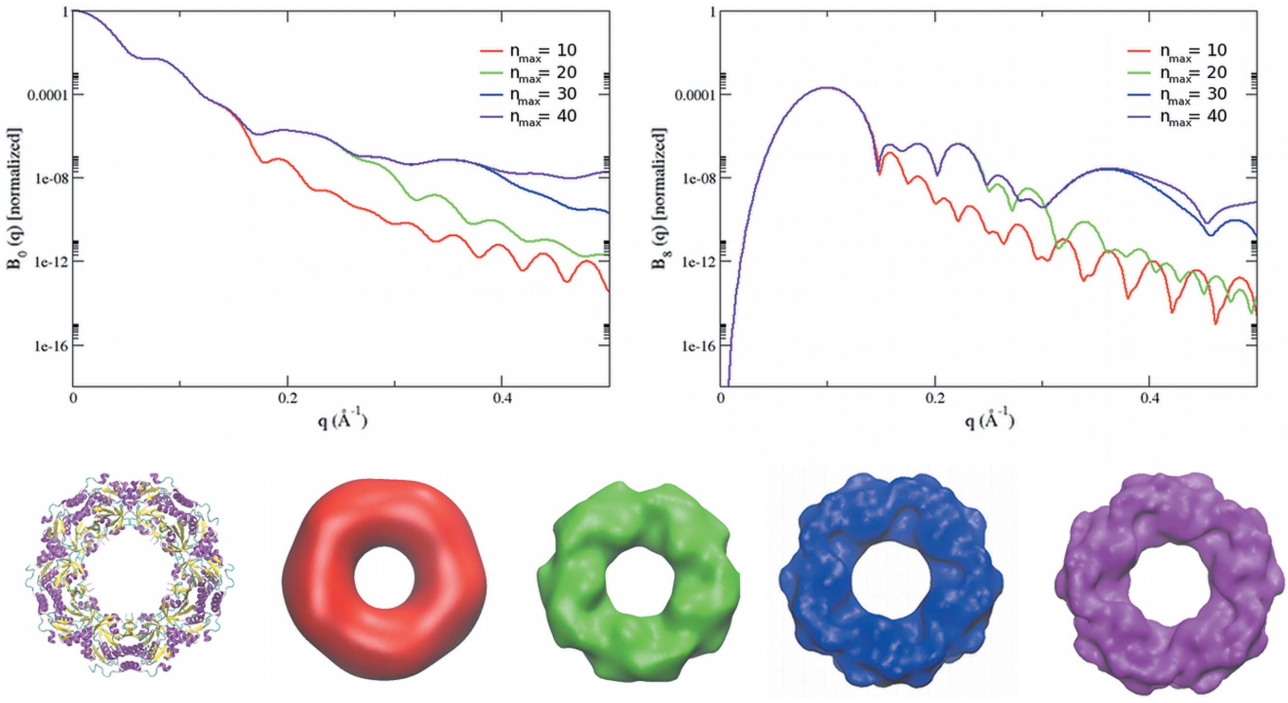

Figure 2
Model resolution dependence on expansion order. Accurate modeling of a higher-resolution fXS profile requires more polynomials, *i.e.* a higher Zernike expansion order. The donut-shaped protein (PDB ID: 2e2g) is used as an example to demonstrate the relationships between real-space resolution, reciprocal-space scattering vector and the maximum expansion order. A higher expansion order needs to be used when truncation errors are observed.

be updated to obtain the fXS profiles for the newly generated model. Although this principle applies to the spherical-harmonics-based method as well [see equation (10)], the memory requirements and computational efficiency for the Zernike method are more favourable.

Furthermore, the proposed Zernike method can, in principle, be improved by removing the q dependence. This can be accomplished by obtaining a series expansion form of $I_{lm}(q)$ that depends only on the three-dimensional Zernike moments and r_{max}, likely along the lines of work outlined by Pavelcik *et al.* (2002). For the SAXS intensity, which is closely related to $B_0(q)$, this expression is quite straightforward (Liu, Morris *et al.*, 2012), indicating that, for higher orders of l, similar simple expressions can be found. This is beyond the scope of this work.

To conclude, an efficient method based on three-dimensional Zernike polynomials for fluctuation X-ray scattering profile calculations is presented. The performance and accuracy are validated against the data obtained from the spherical-harmonics expansion method and simulation results. This method provides an opportunity in *ab initio* model reconstruction based on the fXS profiles, which are embedded in the fluctuation X-ray scattering experimental data.

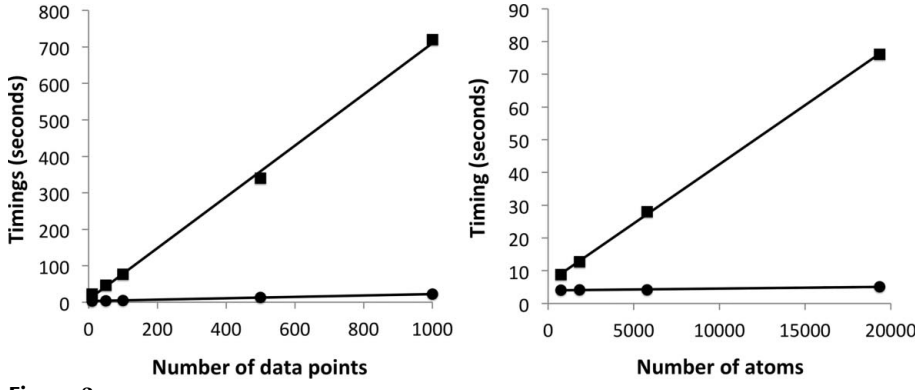

Figure 3
Computing complexity and speed. A computing time comparison for the spherical harmonic and Zernike-based approaches. Both methods depict a linear computational complexity with respect to the number of q values in the data set (left) as well as the number of atoms in the model (right). The speed increase for the Zernike-based method (circles) is approximately a factor of 70 as compared with the spherical-harmonics-based method (squares).

APPENDIX *A*
Fourier transform of three-dimensional Zernike radial functions

Expression (17) of Liu, Morris *et al.* (2012) provides details of the Fourier transform of a Zernike polynomial. A formal proof of the expression (40) as stated by Mathar (2008) could not be found and is clarified below. It is required to show that

$$I_{nl}(q) = \int_0^1 j_l(qr) R_{nl}(r) r^2 \, \mathrm{d}r = (-1)^{[(n-l)/2]} (2n+3)^{1/2} j_{n+1}(q)/q. \quad (19)$$

A concise argument for the validity of (19) follows. We have from Abramowitz & Stegun (1972), (10.1.1) and (9.1.10),

$$j_l(z) = \Gamma\left(\frac{3}{2}\right)\left(\frac{1}{2}z\right)^l \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}z^2)^k}{k!\Gamma(k+l+\frac{3}{2})}. \tag{20}$$

Using this in (19) yields

$$I_{nl}(q) = \Gamma\left(\frac{3}{2}\right)\left(\frac{1}{2}q\right)^l \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}z^2)^k}{k!\Gamma(k+l+\frac{3}{2})} \times \int_0^1 r^{l+2k+2} R_{nl}(r)\,\mathrm{d}r. \tag{21}$$

To evaluate the integrals on the right-hand side of (21), we use the definition of $R_{nl}(r)$, see equation (43) of Mathar (2009),

$$R_{nl}(r) = (2n+3)^{1/2}\, r^l P_p^{(0,l+\frac{1}{2})}(2r^2-1), \qquad p = \frac{1}{2}(n-l), \tag{22}$$

where $P_j^{(\alpha,\beta)}(x)$ is the general Jacobi polynomial. With the substitution $x = 2r^2 - 1$, we then obtain

$$\int_0^1 r^{l+2k+2} R_{nl}(r)\,\mathrm{d}r = (2n+3)^{1/2}\, 2^{-l-k-\frac{5}{2}} \times \int_{-1}^1 (1+x)^{k+l+\frac{1}{2}} P_p^{(0,l+\frac{1}{2})}(x)\,\mathrm{d}x. \tag{23}$$

Next, we use Rodrigues' formula in Abramowitz & Stegun (1972) [equation (21.11.1)], so that

$$(1+x)^{l+\frac{1}{2}} P_p^{(0,l+\frac{1}{2})}(x) = \frac{(-1)^p}{2^p p!}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^p \left[(1-x)^p (1+x)^{p+l+\frac{1}{2}}\right]. \tag{24}$$

Upon p partial integrations in (23), one obtains

$$\int_0^1 r^{l+2k+2} R_{nl}(r)\,\mathrm{d}r = (2n+3)^{1/2}\, 2^{-p-l-k-\frac{5}{2}} \frac{k!}{p!(k-p)!} \times \int_{-1}^1 (1-x)^p (1+x)^{k+l+\frac{1}{2}}\,\mathrm{d}x, \tag{25}$$

when $k \geq p$, and (25) vanishes when $0 \leq k \leq p$. The remaining integral can be expressed in terms of the Beta integral as in Abramowitz & Stegun (1972) [equation (6.2)]. Using the substitution $t = \frac{1}{2}(1+x)$, the integral becomes

$$\int_0^1 r^{l+2k+2} R_{nl}(r)\,\mathrm{d}r = \frac{1}{2}(2n+3)^{1/2} \frac{k!\Gamma(k+l+\frac{3}{2})}{(k-p)!\Gamma(p+k+l+\frac{5}{2})} \tag{26}$$

for $k \geq p$, while (26) vanishes for $0 \leq k \leq p$. Using this in (21) gives

$$I_{nl}(q) = \frac{1}{2}\Gamma\left(\frac{3}{2}\right)\left(\frac{1}{2}q\right)^l \sum_{k=p}^{\infty} \frac{(-\frac{1}{4}q^2)^k}{(k-p)!\Gamma(p+k+l+\frac{5}{2})}. \tag{27}$$

Replacing k by $(j+p)$, $j = 0, 1, \ldots,$ one obtains

$$I_{nl}(q) = (-1)^p\, \frac{1}{2}\Gamma\left(\frac{3}{2}\right)\left(\frac{1}{2}q\right)^{l+2p} \times (2n+3)^{1/2} \sum_{j=0}^{\infty} \frac{(-\frac{1}{4}q^2)^j}{j!\Gamma(j+l+2p+\frac{5}{2})}. \tag{28}$$

Then (19) follows from (20) and the fact that $n = l + 2p$.

APPENDIX *B*
Simulation of scattering data

Scattering data were simulated for molecules in a vacuum, without any sources of noise. Using atomistic models from the PDB (http://www.pdb.org/), a single molecule was randomly rotated and placed in a 1 μm^3 box. A full scattering image, using physically relevant detector parameters that captured q values up to 0.5 Å^{-1}, was then calculated *via* direct summation,

$$A(\mathbf{q}) = \sum_i^{n_{\mathrm{atoms}}} f_i(\mathbf{q}) \exp(i\mathbf{q}\cdot\mathbf{r}_i),$$
$$I(\mathbf{q}) = |A(\mathbf{q})|^2, \tag{29}$$

where f_i and $\mathbf{r}_i$ are the scattering form factor and coordinate of atom i, respectively. Pixels in the image were then binned for specific q and φ values, and the autocorrelations calculated using equation (1). Simulations were repeated for 300 000 images to ensure convergence of the autocorrelations.

After the autocorrelations were calculated from the simulated images, a series of Legendre polynomials were used to fit the autocorrelations according to equation (16). The coefficients were calculated with the general expression for expanding any function with a Fourier–Legendre series,

$$B_l(q) = \frac{2l+1}{2} \int_{-1}^1 P_l(x) f(x)\,\mathrm{d}x,$$
$$x = \cos^2\theta(q) + \sin^2\theta(q)\cos(\Delta\varphi), \tag{30}$$

where l is the order of the Legendre polynomial, $f(x)$ is the autocorrelation function, $C_2(q, \Delta\varphi)$, and x is used to map $\Delta\varphi$ onto $[-1, 1]$ for the Legendre polynomial, P_l. Since the integral in (30) was evaluated numerically, the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm was used to minimize the least-squares difference between the simulated autocorrelation and its series expansion form.

All simulations and calculations were implemented in Python and C/C++ using the *Small Angle Scattering Toolbox* (Liu, Hexemer & Zwart, 2012) as part of the *Computational Crystallography Toolbox* (http://cctbx.sourceforge.net/) (Grosse-Kunstleve *et al.*, 2002). Source code is available upon request and from http://sastbx.als.lbl.gov/.

PHZ would like to thank Professor Dr A. P. Zwart and Professor Dr M. Vlasiou for stimulating discussions. This work was supported by Laboratory Directed Research and Development (LDRD) funding from Berkeley Laboratory, provided by the Director, Office of Science, of the US Department of

Energy under Contract No. DE-AC02-05CH11231. HGL is grateful for the support from the Human Frontier Science Program (HFSP) award 024940, and the discussions with Professor Dr J. C. H. Spence.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions: with Formulas, Graphs and Mathematical Tables*. New York: Dover.

Barth, J. V., Costantini, G. & Kern, K. (2005). *Nature (London)*, **437**, 671–679.

Canterakis, N. (1999). *Proceedings of the 11th Scandinavian Conference on Image Analysis*, Kangerlussuaq, Greenland, pp. 85–93.

Emma, P. *et al.* (2010). *Nat. Photon.* **4**, 641–647.

Frank, J. (2002). *Annu. Rev. Biophys. Biomol. Struct.* **31**, 303–319.

Gaunt, J. (1928). *Proc. R. Soc. London Ser. A*, **122**, 513–532.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. (2002). *J. Appl. Cryst.* **35**, 126–136.

Kam, Z. (1977). *Macromolecules*, **10**, 927–934.

Kam, Z., Koch, M. & Bordas, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 3559–3562.

Liu, H., Hexemer, A. & Zwart, P. H. (2012). *J. Appl. Cryst.* **45**, 587–593.

Liu, H., Morris, R. J., Hexemer, A., Grandison, S. & Zwart, P. H. (2012). *Acta Cryst.* A**68**, 278–285.

Mak, L., Grandison, S. & Morris, R. J. (2008). *J. Mol. Graph. Model.* **26**, 1035–1045.

Mathar, R. J. (2008). *Balt. Astron.* **17**, 383–398.

Mathar, R. J. (2009). *Serb. Astron. J.* **179**, 107–120.

Medalia, O., Weber, I., Frangakis, A., Nicastro, D., Gerisch, G. & Baumeister, W. (2002). *Science*, **298**, 1209–1213.

Miao, J., Ishikawa, T., Shen, Q. & Earnest, T. (2008). *Annu. Rev. Phys. Chem.* **59**, 387–410.

Novotni, M. & Klein, R. (2003). *Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications*, pp. 216–225. New York: ACM.

Orengo, C. A., Todd, A. E. & Thornton, J. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 374–382.

Pavelcik, F., Zelinka, J. & Otwinowski, Z. (2002). *Acta Cryst.* D**58**, 275–283.

Saldin, D. K., Poon, H. C., Bogan, M. J., Marchesini, S., Shapiro, D. A., Kirian, R. A., Weierstall, U. & Spence, J. C. H. (2011). *Phys. Rev. Lett.* **106**, 115501.

Saldin, D. K., Shneerson, V. L., Fung, R. & Ourmazd, A. (2009). *J. Phys. Condens. Matter*, **21**, 134014.

Saldin, D. K., Shneerson, V. L., Howells, M. R., Marchesini, S., Chapman, H. N., Bogan, M., Shapiro, D., Kirian, R. A., Weierstall, U., Schmidt, K. E. & Spence, J. C. H. (2010). *New J. Phys.* **12**, 035014.

Schneider, G. & Fechner, U. (2005). *Nat. Rev. Drug Discov.* **4**, 649–663.

Schoenlein, R. (2011). *Frontiers in Optics*, p. FMC2. Washington: Optical Society of America.

Vartanyants, I. A., Robinson, I. K., McNulty, I., David, C., Wochner, P. & Tschentscher, Th. (2007). *J. Synchrotron Rad.* **14**, 453–470.

Volkov, V. V. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 860–864.

Wang, J. Y. & Silva, D. E. (1980). *Appl. Opt.* **19**, 1510–1518.