Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 13 November 2012
Accepted 2 March 2013

Three-dimensional single-particle imaging using angular correlations from X-ray laser data

Haiguang Liu,[a] Billy K. Poon,[b] Dilano K. Saldin,[c] John C. H. Spence[a] and Peter H. Zwart[b]*

[a]Department of Physics, Arizona State University, Tempe, AZ 85287, USA, [b]Physical Biosciences Division, Lawrence Berkeley National Laboratories, One Cyclotron Road, Berkeley, CA 94720, USA, and [c]Department of Physics, University of Wisconsin–Milwaukee, 1900 E. Kenwood Boulevard, Milwaukee, WI 53211, USA. Correspondence e-mail: phzwart@lbl.gov

Femtosecond X-ray pulses from X-ray free-electron laser sources make it feasible to conduct room-temperature solution scattering experiments far below molecular rotational diffusion timescales. Owing to the ultra-short duration of each snapshot in these *fluctuation scattering experiments*, the particles are effectively frozen in space during the X-ray exposure. In contrast to standard small-angle scattering experiments, the resulting scattering patterns are anisotropic. The intensity fluctuations observed in the diffraction images can be used to obtain structural information embedded in the average angular correlation of the Fourier transform of the scattering species, of which standard small-angle scattering data are a subset. The additional information contained in the data of these fluctuation scattering experiments can be used to determine the structure of macromolecules in solution without imposing symmetry or spatial restraints during model reconstruction, reducing ambiguities normally observed in solution scattering studies. In this communication, a method that utilizes fluctuation X-ray scattering data to determine low-resolution solution structures is presented. The method is validated with theoretical data calculated from several representative molecules and applied to the reconstruction of nanoparticles from experimental data collected at the Linac Coherent Light Source.

1. Introduction

The structure of biological macromolecules is the key to understanding their function and behavior in living cells. The number of structures deposited in the Protein Data Bank (PDB) (Bernstein *et al.*, 1977) currently exceeds 82 000, yet many important structures of biological molecules and their complexes have not been determined. Currently, more than 80% of structures are solved using X-ray crystallography, which relies on the growth of well diffracting crystals that can survive a high dose of X-rays during data collection. Other methods, like nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM), have been successful tools for structure determination but have historically lacked high-throughput capabilities and pose size restrictions on systems studied. Structures of molecular complexes and their behavior in solution are often studied by combining the information from high-resolution crystal structures of domains together with experimental small- or wide-angle X-ray scattering (SAXS/WAXS) data in solution (Putnam *et al.*, 2007). This combined multi-resolution approach allows one to understand the structure and dynamics of macromolecules in solution and often plays an important role in revealing a holistic viewpoint of the system under study (Krukenberg *et al.*, 2011).

A recent development is the use of X-ray free-electron lasers (XFELs), such as the Linac Coherent Light Source (LCLS) (Emma *et al.*, 2010), in structural biology. For molecules or complexes where growing sufficiently large crystals for synchrotron X-ray crystallography is difficult, the method of serial femtosecond macromolecular nanocrystallography has been developed (Chapman *et al.*, 2011). The basic idea behind this new technique is to record the instantaneous elastic scattering using a pulse so brief that it terminates before radiation damage has time to develop ('diffract before destroy') (see Spence *et al.*, 2012, for a review). Although great strides have been made in serial nanocrystallography (Boutet *et al.*, 2012; Kern *et al.*, 2012; Aquila *et al.*, 2012), the structures of macromolecules determined in this fashion are still in the solid state, limiting the capability to deduce the structure and dynamics of macromolecules in a variety of dissimilar conformations without altering crystallization conditions. The diffract-before-destroy nature of serial nanocrystallography applies to single-particle and solution scattering as well (Neutze *et al.*, 2000; Barty *et al.*, 2012; Seibert *et al.*, 2011) and offers the theoretical opportunity to image single molecules at

room temperature (rather than in the frozen state needed to reduce radiation damage in protein crystallography) in a variety of conformations (Giannakis *et al.*, 2012; Schwander *et al.*, 2012) or, in the case of femtosecond solution scattering, to provide additional experimental information as compared to standard synchrotron SAXS/WAXS, as will be discussed below.

For the case of XFEL microdiffraction scattering with one particle per shot, four major challenges exist. First of all, the signal-to-noise (S/N) ratio is small despite the large number of incident photons (*e.g.* 10^{12} in a single pulse). A typical, very large virus scatters about 10^6 photons per pulse, with scattering falling off as the inverse fourth power of scattering angle, resulting in few photons per shot per Shannon pixel at 1 nm resolution, so that background counts will quickly dominate. Secondly, as the orientation of the sample for each shot is unknown, this must be determined before data can be merged (to improve the S/N problem in the first challenge) and the phase problem then solved for the resulting three-dimensional data set (Shneerson *et al.*, 2008). Thirdly, diffraction images should only originate from single particles if they are to be useful for the currently available reconstruction methods. While the hit rate for single-particle patterns is typically a few percent (most X-ray pulses miss particles altogether), the hit rate for solution scattering from a continuously flowing liquid stream is 100%. Finally, biological macromolecules exhibit conformational heterogeneity that typically increases with size. This latter problem exacerbates the former three, making the three-dimensional molecular reconstruction of single-particle scattering data a difficult computational challenge (Fung *et al.*, 2009; Loh *et al.*, 2010; Yoon *et al.*, 2011).

An alternative approach to the single-particle strategies described above is the method proposed by Kam (1977) and Kam *et al.* (1981), illustrated here in Fig. 1. In the latter papers a technique called *fluctuation X-ray scattering* (fXS) is introduced that does not require a single particle per shot. In this technique, it is shown that the average of angular correlation functions of diffraction patterns, each taken from an ensemble of randomly oriented identical particles, will converge to that for one particle. XFEL diffraction data are perfect for the application of Kam's theory, since it can provide solution scattering at room temperature, where, using the 'diffract-before-destroy' mode, resolution is not limited by radiation damage. Furthermore, while the low-signal diffraction patterns and unknown orientation create challenges for other methodologies, it is simple to compute the angular autocorrelations and extract the associated fluctuation X-ray scattering profiles, since these do not require a knowledge of particle orientation in order to merge data. The fXS method when carried out on particles in solution is furthermore immune to the hit-rate problems observed in single-particle scattering experiments. In the latter case, hit rates decline as the beam diameter approaches the submicron size of single particles owing to experimental alignment difficulties and instabilities (Weierstall *et al.*, 2012). In contrast to the single-particle scattering, in which hit rates are typically below 5%, the hit rate for fXS will be 100% when using an X-ray beam much wider than one particle. The Kam method assumes that coherent interparticle interference effects are suppressed, so that the sum of many single-particle-per-shot patterns is the same (apart from S/N) as a single shot containing many particles.

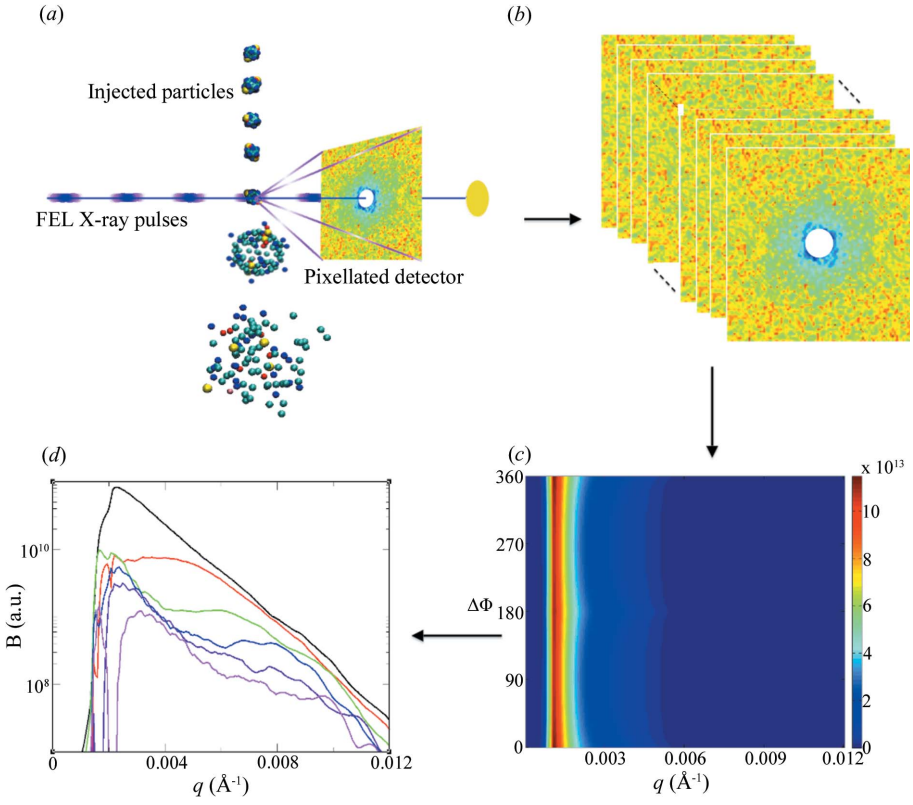

Figure 1
The fluctuation X-ray scattering profile computation process from many femtosecond diffraction patterns. Many diffraction patterns are collected from the femtosecond X-ray scattering experiments at an XFEL light source (*a*) → (*b*). The autocorrelation $C_2(q)$ is computed for each diffraction pattern (DP) and then averaged over all collected DPs (*b*) → (*c*). The fXS profiles $B_l(q)$ are then calculated from the converged $C_2(q)$ (*c*) → (*d*). The focus of this work is to reconstruct the three-dimensional model from the fXS profiles shown as in (*d*).

Angular autocorrelations, as obtained from experimental data, can be reduced to a series of resolution-dependent expansion coefficients, akin to SAXS data (Saldin *et al.*, 2009). These angular autocorrelation functions are the self-convolution of the diffraction pattern intensity taken around concentric rings in the pattern. These expansion coefficients are related to the

amplitudes of a spherical harmonic expansion of the scattering pattern of a single molecule. The individual fXS curves carry essential structural information embedded in the original scattering patterns. The zeroth-order curve, for instance, is equivalent to the square of the standard SAXS data. The original proposal by Kam was to determine the three-dimensional scattering volume from these fXS curves and subsequently apply standard phase-retrieval methods to yield a real-space model. The computational challenges to tackle this hyper-phase problem are, however, non-trivial and it has been suggested that they prevent a complete structure recovery (Elser, 2011) from fXS. In two dimensions (particles differing only by rotation about a single axis), however, the problem does become tractable and both theoretical studies (Saldin, Shneerson *et al.*, 2010) and experimental proof-of-principle experiments have been carried out with great success (Saldin, Poon *et al.*, 2010; Chen *et al.*, 2012).

In three dimensions, particle symmetry constraints have been used in the structure solution process to allow for an efficient solution of the hyper-phase problem (Saldin, Poon, Schwander *et al.*, 2011), but no general solution has yet been demonstrated for non-symmetric particles. Following the success with real-space procedures in the structure solution of SAXS data (Svergun *et al.*, 2001) and two-dimensional fXS data (Chen *et al.*, 2012), a novel method is proposed here in which a reverse Monte Carlo (McGreevy & Pusztai, 1988) procedure is used to determine the structure from fXS data. Unlike the Kam approach, in which the angular correlation function is directly converted to a real-space density by a double phasing approach (Saldin, Poon, Bogan *et al.*, 2011), our approach here is based on optimization of the angular correlation function of a trial model for best fit to the experimental angular correlation function, summed from many diffraction patterns, each from multiple particles. The structure solution procedure proposed uses three-dimensional Zernike polynomials to represent molecular models (Novotni & Klein, 2003; Mak *et al.*, 2008; Liu, Morris *et al.*, 2012) and utilizes the relation between the three-dimensional Zernike moments to the fXS profiles (Liu, Poon *et al.*, 2012). In this paper, the reverse Monte Carlo method has been tested using simulated data from various molecular structures. The method is then applied to the reconstruction of images of ellipsoidal iron oxide nanoparticles (nanorice) using experimental LCLS femtosecond diffraction data (Loh *et al.*, 2010) available from the CXIDB (Kassemeyer *et al.*, 2012).

2. Methods

2.1. Fourier transform of a three-dimensional Zernike polynomial

The three-dimensional Zernike model is a compact description of three-dimensional shapes using convenient orthogonal polynomials. The definition of a three-dimensional Zernike polynomial of order (n, l, m) is

$$Z_{nlm}(\mathbf{r}) = R_{nl}(r)Y_{lm}(\Omega), \tag{1}$$

where $R_{nl}(r)$ is the three-dimensional Zernike radial function and $Y_{lm}(\Omega)$ is the spherical harmonic of order (l, m). Because three-dimensional Zernike polynomials are orthogonal in the unit sphere (*i.e.* $r \leq 1$), any twice differentiable function enclosed in the sphere, after suitable rescaling, can be approximated with a weighted polynomial series:

$$\rho(\mathbf{r}) \simeq \sum_{n=0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} c_{nlm} Z_{nlm}(\mathbf{r}), \tag{2}$$

where the complex coefficients c_{nlm} are the three-dimensional Zernike moments:

$$c_{nlm} = \frac{3}{4\pi} \int_{|\mathbf{r}|\leq 1} \rho(\mathbf{r}) Z^*_{nlm}(\mathbf{r})\, \mathrm{d}\mathbf{r}. \tag{3}$$

The three-dimensional Zernike moments can be efficiently computed *via* the Novotni & Klein algorithm (Novotni & Klein, 2003).

The Fourier transform of the three-dimensional Zernike polynomial approximation can be used to compute the complex scattering function of $\rho(\mathbf{r})$. It can be shown that, for a model with radius $r_{\max}$, the complex scattering factor at $\mathbf{q}$ is given by

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

with

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

Details of the derivation are provided in Liu, Morris *et al.* (2012) and Liu, Poon *et al.* (2012).

2.2. The relation between Zernike moments and the average angular autocorrelation

Experimentally, the average angular correlation function can be extracted from data acquired in femtosecond X-ray scattering experiments using

$$C_2(q, q'\Delta\varphi) = \frac{1}{N} \sum_i \sum_\varphi I_i(q, \varphi) I_i(q', \varphi + \Delta\varphi), \tag{6}$$

where $I_i(q, \varphi)$ is the scattered intensity of pattern i at a pixel position corresponding to reciprocal-space point (q, φ). Given a sufficient number of scattering patterns (each from one or multiple identical particles), this correlation will converge to a fixed value (Kam, 1977; Kam *et al.*, 1981; Kirian *et al.*, 2011; Saldin *et al.*, 2009). The procedure of extracting the embedded fXS profile from femtosecond X-ray diffraction patterns is summarized in Fig. 1. As indicated in the latter expression, the correlation function can be computed either within a fixed resolution shell, $q = q'$, or involve two resolution rings $q \neq q'$. In the remainder of this paper, we will focus on the use of autocorrelations for which $q = q'$ and denote this correlation as $C_{2,q}(\Delta\varphi)$. The use and utility of correlations between resolution rings, *i.e.* $q \neq q'$, will be highlighted in the discussion of this paper (§3.5).

The angular autocorrelation $C_2(q)$ can be decomposed into a weighted series of Legendre polynomials (Saldin *et al.*, 2009):

$$C_{2,q}(\Delta\varphi) = \sum_l F_l(\Delta\varphi) B_l(q), \tag{7}$$

where $B_l(q)$ are the weights and

$$F_l(\Delta\varphi) = \frac{1}{4\pi} P_l\left[\cos^2\theta(q) + \sin^2\theta(q)\cos(\Delta\varphi)\right]. \tag{8}$$

$P_l(\cdot)$ is a Legendre polynomial and

$$\theta(q) = \pi/2 - \sin^{-1}(q/2\kappa), \tag{9}$$

where κ is equal to the wavenumber $2\pi/\lambda$ with λ the wavelength of the incident radiation. The weights $B_l(q)$ and the estimated uncertainties $\sigma_l(q)$ can be obtained using standard numerical techniques from the experimentally obtained $C_2(q)$ patterns (Saldin *et al.*, 2009). The full collection of $B_l(q)$ curves compose an fXS data set.

The three-dimensional Zernike moments themselves are related to the $B_l(q)$ curves as described in detail in Liu, Poon *et al.* (2012) and summarized below. The coefficients c_{nlm} are linked to a spherical harmonic expansion of the intensities as a function of resolution *via* a Gaunt series:

$$I_{lm}(q) = \sum_{l'} \sum_{l''} \sum_{m'} \sum_{m''} a_{l'm'}(qr_{\max}) a_{l''m''}(qr_{\max}) G_{ll'l''}^{mm'm''}, \tag{10}$$

where $G_{ll'l''}^{mm'm''}$ are Gaunt coefficients and coefficients $a_{lm}(qr_{\max})$ are a function of c_{nlm} *via*

$$a_{lm}(qr_{\max}) = \sum_{n}^{n_{\max}} w_{nl}(qr_{\max}) c_{nlm} \tag{11}$$

with

$$w_{nl}(qr_{\max}) = i^l(-1)^{(n-l)/2} b_n(qr_{\max}). \tag{12}$$

Finally, one can show that (Kam, 1977; Saldin *et al.*, 2009)

$$B_l(q) = \sum_m |I_{lm}(q)|^2. \tag{13}$$

2.3. Structure solution procedure

The real-space model reconstruction method utilizes two basic real-space operations to modify an initial trial solution. Modification to the proposed trial solutions are random *dilations* or *erosions* of the existing body (Fig. 2). In a dilation, the existing body is extended at a certain position, whereas in an erosion, a piece of the model is carved out and discarded. The main reasoning behind these operations is that they are local, affecting only the surface of the current model and, in principle, do not tend to generate trial solutions that consist of many disconnected parts. It must be stressed, however, that no topology-conserving constraints or compactness-enforcing restraints are used in the reconstruction, in contrast to what is used during similar SAXS shape-reconstruction methods (Svergun *et al.*, 2001; Svergun, 1999; Franke & Svergun, 2009). Proposed random modifications to the model are accepted on the basis of the well known Metropolis–Hastings criteria (Metropolis *et al.*, 1953).

2.3.1. Model perturbations. Because the resolution range of the experimental data lies well within the range in which one can model macromolecules relatively well using a uniform density approximation, the optimization problem is limited to placing beads of density onto a grid, similar to the task performed in SAXS (Svergun, 1999). Owing to the Monte Carlo approach used in the optimization, small changes to a trial model are favored over large modifications. The three-dimensional model is built on a predefined grid with a size based on the analysis of the SAXS data. By default, the starting model is a hollow sphere with an outer radius equal to 80% of the grid radius and an inner radius equal to 30% of the grid radius. The model reconstruction is processed on a cubic grid with ($40 \times 40 \times 40$) voxels unless specifically mentioned. Although the final results do not depend on the starting model, convergence times can be significantly reduced by a judicious choice of the starting model. Changes to the trial model are done using the following two-step procedure:

(i) Pick a voxel at random from a predetermined list of voxels that are allowed to be modified. The predetermined list of voxels only contains voxels that lie within a preset radius of other border voxels.

(ii) If the density of this voxel is equal to 1, set the density of all neighbors to 1 (dilation). If the density of this voxel is 0, set all neighbors to this voxel to 0 (erosion).

Only model perturbations in which the full model actually changes are carried out; 'null modifications' are not allowed and, when detected, the procedure is attempted again. Because changes to the trial model are biased to the surface only, the modification procedure tends to produce densities that are connected and does not produce scattered, discon-

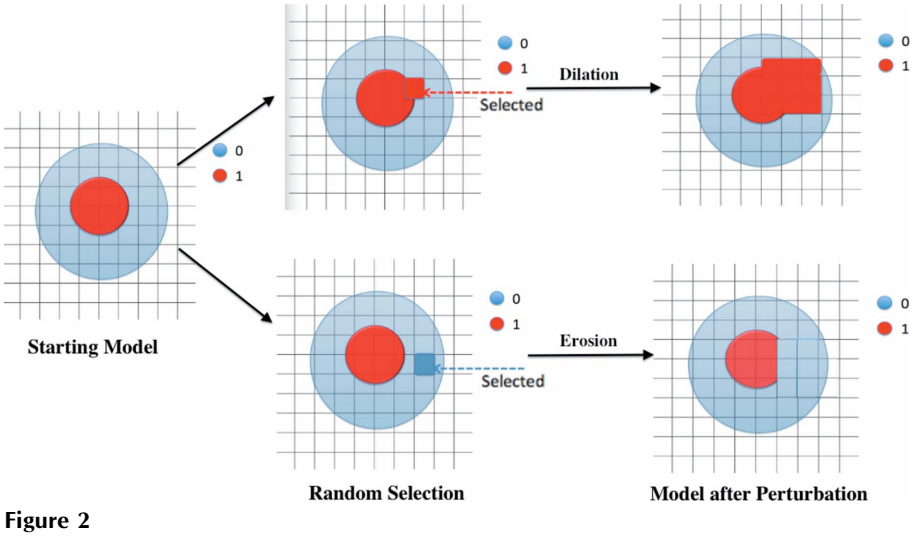

Figure 2
Model perturbation method. Two perturbation modes are used: dilation and erosion.

nected clumps of density. After each perturbation, the Zernike moments, c_{nlm}, of the modification to the model are computed and are either added (in the case of model dilation) or subtracted (in the case of model erosion) from the three-dimensional Zernike moments of the trial model. The new set of c_{nlm} are used to compute model fXS curves as described above. The rate-limiting step in the latter procedure is the computation of the modified three-dimensional Zernike moments, c_{nlm}. The timings in this step depend on the number of voxels involved in the model perturbation. Changes to the trial model are accepted on the basis of the Metropolis–Hastings criteria as explained below.

2.3.2. Simulated annealing optimization. The target of the optimization is the goodness-of-fit between the model and experimental fXS profiles, measured by χ^2 defined as

$$\chi^2 = \frac{1}{N_q L}\sum_{l=0}^{L}\sum_{j=1}^{N_q}\left[\frac{B_l^{\rm expt}(q_j) - kB_l^{\rm model}(q_j)}{\sigma_l^{\rm expt}(q_j)}\right]^2. \qquad (14)$$

The perturbations described in the previous section are either accepted or rejected to improve the model–data agreement. The acceptance probability is controlled by a Boltzmann distribution of χ^2, following the Metropolis–Hastings criteria. The simulated annealing (SA) algorithm is employed to avoid becoming trapped in local minima (Kirkpatrick *et al.*, 1983). One of the critical parameters for this SA algorithm is the starting temperature. The starting temperature is chosen on the basis of the standard deviation of χ^2, calculated from a number of random perturbations of the starting model (100 perturbations by default). The cooling scheme used is exponential decay, where the temperature at cooling step t is $T(t) = T_0\alpha^t$, with $\alpha = 0.9$ as the default value. At each temperature, 500 perturbation steps will be carried out. The optimization is terminated if either the temperature is lower than 0.01% of the starting temperature or the acceptance ratio is below 10.0%.

2.3.3. Model superpositioning. Multiple trial solutions from independent runs can be compared by aligning the three-dimensional models. Given the fact that independent trial solutions have the same $r_{\rm max}$, the center of mass for the final models is mostly aligned to the center of the sphere; therefore only a rotational alignment needs to be performed. Because the trial models available have been modeled with three-dimensional Zernike polynomials, a correlation-based alignment using fast Fourier methods can be employed. The spatial correlation coefficient is defined as

$$\mathrm{CC} = \frac{\langle\rho_{\rm fixed}(\mathbf{r})\rho_{\rm moving}(\mathbf{r}|\alpha,\beta,\gamma)\rangle - \langle\rho_{\rm fixed}(\mathbf{r})\rangle\langle\rho_{\rm moving}(\mathbf{r}|\alpha,\beta,\gamma)\rangle}{\sigma(\rho_{\rm fixed})\sigma(\rho_{\rm moving})}, \qquad (15)$$

where $\rho_{\rm fixed}(\mathbf{r})$ refers to the reference model and $\rho_{\rm moving}(\mathbf{r}|\alpha,\beta,\gamma)$ refers to the to-be-aligned object after a rotation with Euler angles (α,β,γ). It can be shown that angles (α,β,γ) that maximize CC can be found using a series of two-dimensional Fourier transforms (Appendix *A*).

Table 1
Summary of the model reconstruction from the simulated data.

The density correlation coefficient (cc) was calculated using the method described in Appendix *A* with $n_{\rm max} = 20$. The cc's are averaged from ten runs and the number in parentheses is the standard deviation of the cc's.

Name	cc (start)	cc (fXS)	cc (SAXS)	PDB ID	$r_{\rm max}$ (Å)
LAO	0.46	0.77 (0.03)	0.68 (0.07)	2lao	35.0
Lysozyme	0.50	0.76 (0.02)	0.69 (0.03)	6lyz	25.0
Peroxiredoxin	0.45	0.87 (0.04)	0.39 (0.06)	2e2g	75.0
STMV	0.57	0.83 (0.06)	0.50 (0.08)	1a34	100.0

Table 2
Key parameters for model reconstruction.

Name	$r_{\rm max}$	$q_{\rm max}$ (Å^{-1})	Voxel size (Å)
LAO	35.0	0.25	1.75
Lysozyme	25.0	0.25	1.25
Peroxiredoxin	75.0	0.25	3.75
STMV	100.0	0.10	5.00

3. Results and discussion

3.1. Model reconstruction from simulated data

The performance of the algorithm has been tested using theoretical fXS profiles of four model systems, namely (*a*) lysine-, arginine-, ornithine-binding protein (LAO); (*b*) lysozyme; (*c*) peroxiredoxin protein; and (*d*) satellite tobacco mosaic virus (STMV). Each example represents particular molecular groups: lysozyme and LAO are typical globular-like proteins, the peroxiredoxin protein complex has a donut architecture, whereas the STMV virus has icosahedral symmetry with a central cavity. fXS profiles were computed using the methods described previously (Liu, Poon *et al.*, 2012), up to a q value of 0.25 Å^{-1}. To gauge the reconstruction accuracy, we computed the spatial correlation coefficient between the reconstructed model and the original PDB model, aligned to the proposed solution, rendered onto three-dimensional grids as a body with uniform density (Appendix *A*). As can be seen from Table 1, the average correlation coefficients of the reconstruction *versus* the known model are above 75%, indicating the high quality of the fit (Fig. 3). Key parameters used in model constuction are found in Table 2. In

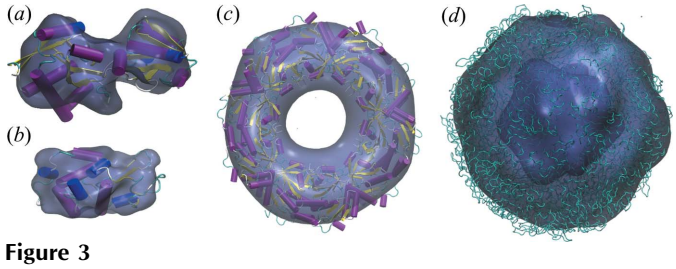

Figure 3
The four molecular structures used in the testing cases. The reconstructed models are superposed to the original PDB models, where the PDB models are represented using cartoons, and the reconstructed models are shown in the form of density isosurfaces. (*a*) LAO binding protein, (*b*) lysozyme, (*c*) peroxiredoxin and (*d*) STMV. They are not shown to scale; see Table 1 for the actual sizes.

order to assess the increase in information contained in an fXS data set as compared to a standard SAXS data set, the reconstruction algorithm was applied on $B_0(q)$, the square of a SAXS curve. The SAXS shape-reconstruction attempts are only marginally successful with average correlation coefficients below 70%. Furthermore, the shape reconstructions from SAXS data sometimes result in multiple disconnected bodies, highlighting a loss of information when using only SAXS data as compared to the fXS data set.

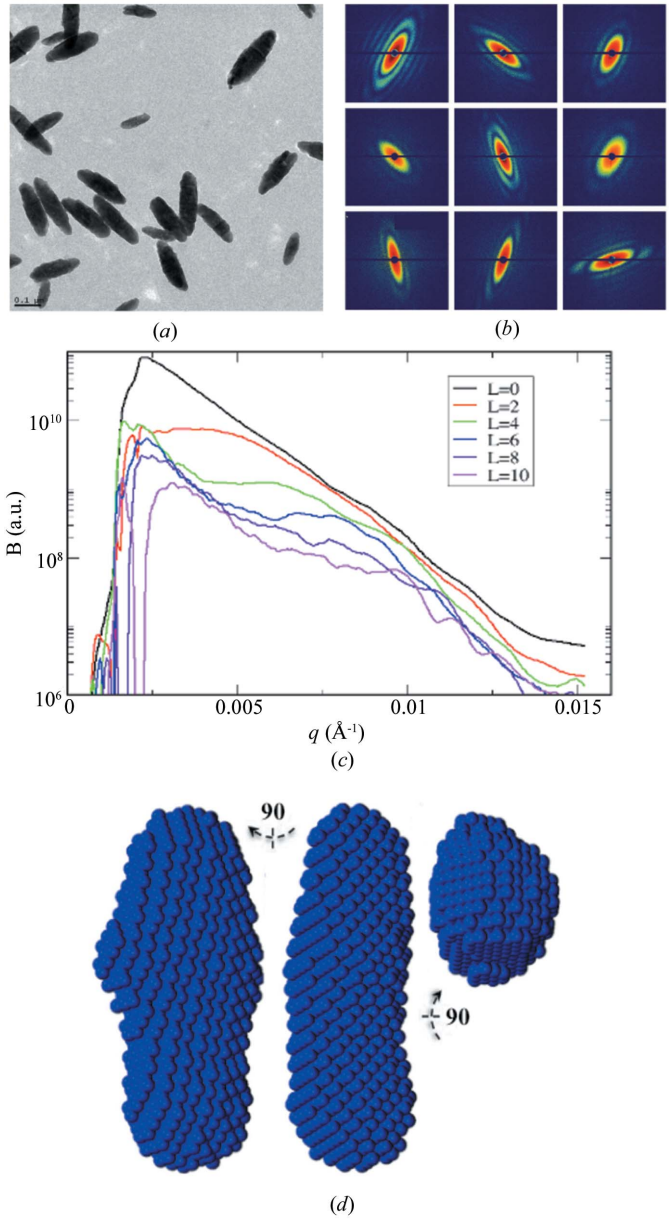

Figure 4
The nanorice data and reconstructed model. (*a*) TEM image of the nanorice sample. (*b*) Representative diffraction patterns collected at the LCLS. (*c*) fXS profiles extracted from a subset of about 800 diffraction patterns. (*d*) The reconstructed three-dimensional models in bead representations in three view angles. (*a*) and (*b*) are reproduced from Kassemeyer *et al.* (2012) with permission from the Optical Society of America.

3.2. Model reconstruction from experimental data

The proposed reconstruction method was also tested using experimental data collected at the LCLS. Here, we present the results for ellipsoidal iron oxide nanoparticles, also known as nanorice. As observed in TEM (transmission electron microscopy), the nanorice has a size variation from 150 to 250 nm in the longest axis (Fig. 4*a*). Publicly available X-ray scattering data of nanorice (Bogan *et al.*, 2008) were downloaded from the CXIDB (Kassemeyer *et al.*, 2012). Figs. 4(*a*) and 4(*b*) show TEM images and representative single-particle diffraction patterns of the nanorice sample, respectively. The available scattering patterns were filtered to exclude multiple-hit patterns. The reason for this filtering step was to ensure a high-quality data set given the limited size of the data set. Although multiple-particle shots are tolerated in an fXS experiment, experimental autocorrelation profiles converge faster when only single particles are used (Kirian *et al.*, 2011). After image selection, about 800 scattering patterns remained from which an fXS data set was obtained (Fig. 4*c*). Because of the limited data-set size and associated convergence issues, $B_l(q)$ curves were limited up to $l = 4$ (Fig. 4*c*). Experimental data up to about $q_{\max} = 0.0156$ Å^{-1} (~413 Å) were used during model reconstruction. The model reconstruction is carried out on a three-dimensional grid of 50 × 50 × 50 and the starting hollow-sphere radius is 1200 Å. As a result, the voxel size is about 48^3 Å^3. The reconstructed model captures the structural features of nanorice particles, as shown in Fig. 4(*d*), providing a proof of principle of our method on experimental data. It is worth pointing out that the obtained model is an averaged model from all nanorice particles that contribute to the scattering data. To achieve a high-resolution model, the sample particles need to be in agreement up to the desired resolution.

3.3. Performance and speed

The underlying procedure for computing fXS profiles and the need to only recompute three-dimensional Zernike moments of the small trial perturbations allow for fast recomputation of the target function, a prerequisite for efficient Monte Carlo-based methods. Average timings for all examples are shown in Fig. 5 for various expansion orders. The expansion order, $n_{\max}$, determines the resolution of the final model and the maximum momentum transfer limit that can be used during fitting (Liu, Poon *et al.*, 2012). Reconstruction times lie well under 2 h for all models, but can likely be speeded up by further algorithmic improvements.

3.4. Uniqueness

Given the analyses by Elser (2011) showing that structure determination from fXS data is an under-constrained problem, the demonstrated reconstructions seem to contradict these findings. The arguments in the latter study indicate that the data-to-parameter ratio is about 70% under the assumption of independent density values in real space. This lack of information should result in degenerate solutions that do not need to resemble the target structure. This would be the case for either solving the hyper-phase problem or when using a real-

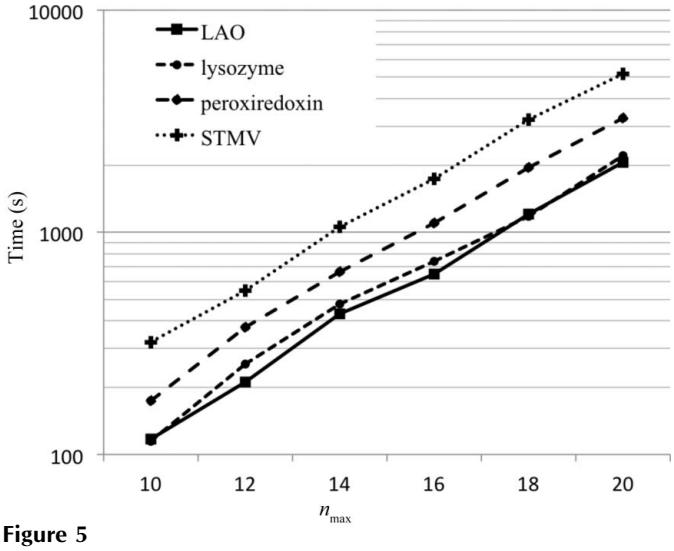

Figure 5
The performance of the reconstruction algorithm. The computing time increases exponentially with respect to the maximum expansion order. With $n_{max} = 20$, the computing time is still within the capacity of a single processor, since it takes about 1.5 h to obtain the converged model for the most difficult case, the STMV. For higher orders, parallelization with computer clusters or GPUs should be used to speed up the reconstruction.

space approach. In practice, however, solutions are not degenerate and match the target shape. The reason for this seemingly contradictory result (Elser, 2011) is most likely because the number of independent variables in real space is significantly reduced by the fact that, at low resolution, regions of molecular density are connected and smooth. This additional information is used successfully in restraining the structure determination from SAXS data to guide the solution process to produce physically meaningful structures (Svergun, 1999; Svergun *et al.*, 2001). Although connectivity restraints are not used directly in the proposed algorithm, the update scheme tends to favor the exploration of the part of solution space that produces connected structures, thus naturally limiting the size of the feasible set of solutions, resulting in meaningful reconstructions.

3.5. The use of angular correlations across resolution shells

As shown in equation (6), correlations between resolution shells (q') can be computed in a straightforward manner. Given that these correlations provide additional experimental information, it is not unlikely that the use of this information will strengthen the real-space structure solution procedure outlined in this manuscript. In fact, the use of these correlations are essential for structure solution procedures that attempt to solve the 'double phase problem' (Saldin, Poon *et al.*, 2010; Saldin, Shneerson *et al.*, 2010; Saldin *et al.*, 2009). Another exciting possible application of the cross correlation data is to use this information as an unbiased quality measure of the proposed model. The use of this data in cross validation can reduce over-fitting (Brünger, 1992) and assist in model identification (Schneider & Sheldrick, 2002).

4. Conclusions

We have demonstrated the feasibility of *ab initio* model reconstruction based on fXS profiles that can be extracted from ultra-fast X-ray scattering experiments. The new X-ray free-electron laser light sources, together with the development of fast detectors and computational models, now allow Kam's (1977) theory to be applied in practical applications for the first time. As we have shown from the simulation and experimental data, three-dimensional low-resolution structures can be obtained from fXS data. The additional information content in the data reduces or even eliminates the need for external spatial constraints, thus significantly reducing ambiguities in the determination of low-resolution solution models.

The algorithm and procedures presented here demonstrate a diminished need for explicit spatial restraints when solving structures from fXS data as compared to traditional SAXS data. This is especially the case in the early stages of the structure solution method. Nevertheless, the application of prior knowledge, either based on symmetry or basic topological features, can speed up the structure solution process. It must be noted that an initial guess of the size of the particle is required to start the reconstruction. This information can be easily obtained from an analysis of the SAXS data, $[B_0(q)]^{1/2}$ (Liu, Hexemer *et al.*, 2012). As scattering data extend to higher resolution, higher-order Zernike polynomials are required to model the fXS profiles and more spatial details can be obtained. A basic rule of thumb states that $q_{max}R = l_{max}$, where R is equal to the radius of the particle and q_{max} is the maximum available momentum transfer value of the data (Pendry, 1974). This rule also guides the choice for the expansion order, n_{max}, in the reconstruction, requiring $n_{max} \geq q_{max}R$.

Although experimental challenges for the fXS technique still need to be addressed, the presented results indicate that a successful fXS experiment produces significantly more information compared to a standard synchrotron SAXS/WAXS experiment. This is a result of the two-dimensional nature of fXS data, rather than the one-dimensional data used in SAXS/WAXS. The reduction of ambiguities sometimes present in the interpretation of SAXS/WAXS data is likely to result in an improved understanding of the structural and dynamic properties of macromolecules in solution.

APPENDIX *A*
Fast Fourier alignment of three-dimensional Zernike polynomial-based models

Expressing $\rho_{\cdot}(\mathbf{r})$ as a sum of three-dimensional Zernike polynomials,

$$\rho_{\cdot}(\mathbf{r}) = \sum_{n=0}^{n_{max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} c_{nlm}^{\cdot} R_{nl}(r) Y_{lm}(\omega), \qquad (16)$$

one obtains for the mean:

$$\langle \rho_{\cdot}(\mathbf{r}) \rangle = \int\limits_{|\mathbf{r}| \leq \mathbf{1}} \rho_{\cdot}(\mathbf{r})\, \mathrm{d}\mathbf{r} = |c_{000}^{\cdot}|. \tag{17}$$

Similarly, the variance of $\rho_{\cdot}(\mathbf{r})$ can be obtained as

$$\begin{aligned} \sigma^2(\rho_{\cdot}) &= \int\limits_{|\mathbf{r}| \leq \mathbf{1}} \left[\rho_{\cdot}(\mathbf{r}) - \langle \rho_{\cdot}(\mathbf{r}) \rangle\right]^2 \mathrm{d}\mathbf{r} \\ &= \sum_{n>0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{+l} (c_{nlm}^{\cdot} [c_{nlm}^{\cdot}]^*). \end{aligned} \tag{18}$$

The cross term $\langle \rho_{\text{fixed}}(\mathbf{r}) \rho_{\text{moving}}(\mathbf{r}|\alpha, \beta, \gamma) \rangle$ is evaluated with a fast Fourier transform (FFT)-based method for all possible orientations. Following Trapani & Navaza (2006), this so-called rotation function takes the form

$$\mathcal{R}[\alpha, \beta, \gamma] = \sum_{l=0}^{l_{\max}} \sum_{m,m'=-l}^{l} C_{m,m'}^{l} D_{m,m'}^{l}(\alpha, \beta, \gamma) \tag{19}$$

with $D_{m,m'}^{l}(\alpha, \beta, \gamma)$ defined in equation (24) and

$$C_{m,m'}^{l} = \frac{4\pi}{3} \int\limits_{0}^{1} c_{lm}^{\text{fixed}}(r) [c_{lm'}^{\text{moving}}(r)]^* r^2\, \mathrm{d}r, \tag{20}$$

where the functions $c_{lm}^{\text{fixed}}(r)$ and $c_{lm}^{\text{moving}}(r)$ are radial dependent expansion coefficients of the fixed and moving three-dimensional models, respectively:

$$\rho(\mathbf{r}) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} c_{lm}(r) Y_{lm}(\omega_r). \tag{21}$$

From equation (16) it is easy to see that

$$c_{lm}(r) = \sum_{n=0}^{n_{\max}} c_{nlm} R_{nl}(r). \tag{22}$$

The orthogonality in the radial part of the three-dimensional Zernike polynomials quickly leads to

$$C_{m,m'}^{l} = \frac{4\pi}{3} \sum_{n}^{n_{\max}} c_{nlm}^{\text{fixed}} [c_{nlm'}^{\text{moving}}]^*. \tag{23}$$

Using

$$D_{m,m'}^{l}(\alpha, \beta, \gamma) = d_{m,m'}^{l}(\beta) \exp[i(m\alpha + m'\gamma)], \tag{24}$$

the familiar fast rotation function in sections of β appears:

$$\begin{aligned} \mathcal{R}_{\beta}[\alpha, \gamma] &= \sum_{m=-l_{\max}}^{l_{\max}} \sum_{m'=-l_{\max}}^{l_{\max}} S_{m,m'}(\beta) \exp[i(m\alpha + m'\gamma)], \\ S_{m,m'}(\beta) &= \sum_{l} C_{m,m'}^{l} d_{m,m'}^{l}(\beta), \end{aligned} \tag{25}$$

where $d_{m,m'}^{l}(\beta)$ is the Wigner small-d function. It can be computed using standard recurrence relations or more advanced FFT methods as described by Trapani & Navaza (2006). After computing the rotation function in sections of fixed β, a subsequent peak picking can provide a list of approximate solutions that can be further optimized using local search methods.

HL, BKP and PHZ were 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. JCHS and HL acknowledge funding from the Human Frontier Science Program (HFSP) award No. 024940. DKS acknowledges support from NSF grant No. MCB-1158138 and the Research Growth Initiative (RGI) of the University of Wisconsin–Milwaukee. We thank Dr N. Zatsepin for stimulating discussions. The authors express gratitude to their peers who made experimental data available to the general public *via* the CXIDB.

References

Aquila, A. *et al.* (2012). *Opt. Express*, **20**, 2706–2716.
Barty, A. *et al.* (2012). *Nature Photonics*, **6**, 35–40.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bogan, M. J. *et al.* (2008). *Nano Lett.* **8**, 310–316.
Boutet, S. *et al.* (2012). *Science*, **337**, 362–364.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
Chen, G., Modestino, M. A., Poon, B. K., Schirotzek, A., Marchesini, S., Segalman, R. A., Hexemer, A. & Zwart, P. H. (2012). *J. Synchrotron Rad.* **19**, 695–700.
Elser, V. (2011). *Ultramicroscopy*, **111**, 788–792.
Emma, P. *et al.* (2010). *Nature Photonics*, **4**, 641–647.
Franke, D. & Svergun, D. I. (2009). *J. Appl. Cryst.* **42**, 342–346.
Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. (2009). *Nature Physics*, **5**, 64–67.
Giannakis, D., Schwander, P. & Ourmazd, A. (2012). *Opt. Express*, **20**, 12799–12826.
Kam, Z. (1977). *Macromolecules*, **10**, 927–934.
Kam, Z., Koch, M. & Bordas, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 3559–3562.
Kassemeyer, S. *et al.* (2012). *Opt. Express*, **20**, 4149–4158.
Kern, J. *et al.* (2012). *Proc. Natl Acad. Sci. USA*, **109**, 9721–9726.
Kirian, R. A., Schmidt, K. E., Wang, X., Doak, R. B. & Spence, J. C. (2011). *Phys. Rev. E*, **84**, 011921.
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
Krukenberg, K. A., Street, T. O., Lavery, L. A. & Agard, D. A. (2011). *Q. Rev. Biophys.* **44**, 229–255.
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.
Liu, H., Poon, B. K., Janssen, A. J. E. M. & Zwart, P. H. (2012). *Acta Cryst.* A**68**, 561–567.
Loh, N. D. *et al.* (2010). *Phys. Rev. Lett.* **104**, 239902.
McGreevy, R. L. & Pusztai, L. (1988). *Mol. Simul.* **1**, 359–367.
Mak, L., Grandison, S. & Morris, R. J. (2008). *J. Mol. Graph. Model.* **26**, 1035–1045.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.
Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.
Novotni, M. & Klein, R. (2003). *Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications*, SM '03, pp. 216–225. New York, NY, USA: ACM.
Pendry, J. B. (1974). *Low Energy Electron Diffraction: the Theory and its Application to Determination of Surface Structure.* London: Academic Press.
Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. (2007). *Q. Rev. Biophys.* **40**, 191–285.
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., Poon, H. C., Schwander, P., Uddin, M. & Schmidt, M. (2011). *Opt. Express*, **19**, 17318–17335.
Saldin, D. K., Poon, H. C., Shneerson, V. L., Howells, M., Chapman, H. N., Kirian, R. A., Schmidt, K. E. & Spence, J. C. H. (2010). *Phys. Rev. B*, **81**, 174105.
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, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Schwander, P., Giannakis, D., Yoon, C. H. & Ourmazd, A. (2012). *Opt. Express*, **20**, 12827–12849.
Seibert, M. M. *et al.* (2011). *Nature (London)*, **470**, 78–81.
Shneerson, V. L., Ourmazd, A. & Saldin, D. K. (2008). *Acta Cryst.* A**64**, 303–315.
Spence, J. C. H., Weierstall, U. & Chapman, H. (2012). *Rep. Prog. Phys.* **75**, 102601.
Svergun, D. I. (1999). *Biophys. J.* **76**, 2879–2886.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. (2001). *Biophys. J.* **80**, 2946–2953.
Trapani, S. & Navaza, J. (2006). *Acta Cryst.* A**62**, 262–269.
Weierstall, U., Spence, J. C. & Doak, R. B. (2012). *Rev. Sci. Instrum.* **83**, 035108.
Yoon, C. H. *et al.* (2011). *Opt. Express*, **19**, 16542–16549.