Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 5 August 2006
Accepted 19 November 2006

Direct Bragg-peak phase retrieval by a hybrid-input–output algorithm with proper intensity normalization

Ning Lei

Center for Advanced Radiation Sources, The University of Chicago, 5640 S. Ellis Ave, Chicago, IL 60637, USA. Correspondence e-mail: ninglei@yahoo.com

A hybrid-input–output algorithm is applied to reconstruct hypothetical carbon cluster crystals in both two and three dimensions. It is shown that normalizing the Bragg peaks to those from a cluster of solid spheres or discs with uniform electron density can often lead to faithfully reconstructed objects at both atomic and low resolutions. It is shown that, even without the central peak, low-resolution structures can still be reconstructed with good fidelity. The effect of Bragg-peak noise on object reconstruction is examined throughout the paper. Successful reconstructions of 500- and 1000-atom carbon cluster crystal structures at both atomic and low resolutions are given.

1. Introduction

The success of Fourier-transform-based dual-space iteration algorithms, such as the error-reduction (ER) (Gerchberg & Saxton, 1972), the hybrid-input–output (HIO) (Fienup, 1982) and the difference map (DM) (Elser, 2003*a*,*b*), in solving the structures of non-periodic objects from X-ray scattering intensities alone has generated interest in applying such algorithms, including a more recent algorithm of charge flipping (CP), to periodic objects (Elser, 2003*b*; Oszlányi & Sütő, 2004, 2005; Wu *et al.*, 2004; Oszlányi *et al.*, 2006). The intuition in the solvability of the structure of a non-periodic object from its X-ray scattering intensity alone is that the scattering intensity, in principle, can be measured infinitely finely (limited by X-ray spectrometer resolution widths). Therefore, the number of unique equations, each of which relates (to a very good approximation due to Gibbs ripple effect) the scattering intensity at a particular X-ray wavevector transfer to the object's electron densities at equal-spaced grid points, with the grid step size determined by the maxima of the measured wavevector transfers (but not how fine the step size is in reciprocal space), can be much greater than the number of grid points (Miao *et al.*, 2003). In recent years, many non-periodic object structures of one dimension (Zheng *et al.*, 2001), two dimensions (Miao *et al.*, 1999; Marchesini *et al.*, 2003; He *et al.*, 2003) and three dimensions (Miao *et al.*, 2002; Chapman *et al.*, 2006) have been reported to have been solved by such algorithms. In the case of a periodic object, typically what are measured are the Bragg-peak intensities and, therefore, neglecting X-ray absorption, the number of grid points on the object, namely the number of unknowns, exceeds the number of measured Bragg peaks by about a factor of 2, due to the basic relation $I(-\mathbf{Q}_{HKL}) = I(\mathbf{Q}_{HKL})$ for the scattering intensity, where $\mathbf{Q}_{HKL}$is the X-ray wavevector transfer of a Bragg peak with Miller indices (H, K, L). Therefore, it seems impossible to solve the structure of a crystal by applying dual-space iteration algorithms to Bragg-peak intensities alone. Nevertheless, Elser reported (Elser, 2003*b*) successfully solving crystal structures for cases of up to a few hundred non-H atoms in a primitive unit cell with Bragg peaks available up to a crystallographic resolution of about 1 Å or better with his DM algorithm, by applying a real-space constraint of keeping the N largest peaks on the grid points, with N the number of atoms in the primitive unit cell, and setting the rest of the electron densities to zero. The peak-picking procedure was used previously in the well known direct phasing method of Shake-n-Bake (SnB) of Weeks *et al.* (Weeks *et al.*, 1993; Miller *et al.*, 1993). (In addition, SnB uses a minimum function to restrict the phases.) In both Weeks *et al.*'s and Elser's work, the atomic structure factor is normalized to a point-like atom, simply by dividing the original scattering intensity by that of the atom. In the CP algorithm, no such division is carried out. The direct phasing in SnB, DM and CP tend to fail when Bragg peaks are not available at atomic resolutions and, in addition, in the case of SnB, typically when the number of non-H atoms in a primitive unit cell exceeds 2000.

This work, using the HIO algorithm, tries to phase Bragg peaks directly with data available up to any resolution. The keys to this work are a proper normalization of the scattering intensities and setting an upper limit to the electron density in the HIO procedure. Instead of normalizing to a point atom, the atoms are normalized to solid spheres of uniform electron density of one electron per unit volume, with a proper choice of the solid-sphere radius R. Examples of solving hypothetical two- and three-dimensional carbon cluster crystals of up to 1000 atoms in a primitive unit cell are given. It will be shown that a crystal structure can still be found even if the central Bragg peak is missing. The simulated Bragg peaks are created

with Poisson noises, with adjustable noise levels. The unit cell in this work always means a primitive unit cell and lattice symmetries are not taken into account. Resolution, in this paper, if not specified otherwise, always means crystallographic resolution, $2\pi/Q_{\max}$ with $Q_{\max}$ the largest wavevector transfer for the available Bragg peaks.

2. Algorithm details

For crystals composed of atoms of type A, the structure factor amplitude is normalized to

$$|F_{\rm norm}(\mathbf{Q}_{HKL})| = \frac{|F(\mathbf{Q}_{HKL})|}{|F_A(\mathbf{Q}_{HKL})|}\,|F_{\rm sphere}(\mathbf{Q}_{HKL}, R_0/\beta_S)|, \qquad (1)$$

where $|F(\mathbf{Q}_{HKL})|$ is the unit-cell X-ray scattering structure factor amplitude at Bragg peak $\mathbf{Q}_{HKL}$, $F_A(\mathbf{Q}_{HKL})$ is the scattering factor for atom A and $F_{\rm sphere}(\mathbf{Q}_{HKL}, R_0/\beta_S)$ is the scattering factor for a solid sphere of radius R_0/β_S with uniform electron density of one electron per unit volume. R_0 is conveniently set as the atom radius. Choosing a proper value for β_S can be $Q_{\max}$ dependent, as will be shown. Equation (1) transforms the question of finding the electron-density profile for a cluster of atoms of type A into that for a cluster of uniform solid spheres of radius R_0/β_S. In the first step of the HIO algorithm, random positive electron densities are generated and the phases of the Fourier transform of the random densities are obtained. The next iteration couples the normalized structure factor amplitudes with the phases and carries out an inverse Fourier transform to generate an electron-density profile. The electron-density profile is then subjected to real-space constraints to generate the output electron-density profile from the second iteration. The procedure is best described by the following mathematical expression:

$$\begin{cases} \text{density } = \mathrm{FT}^{-1}(|F_{\rm norm}(\mathbf{Q}_{HKL})|\exp(i\phi_n)) \\ \text{if density } (n_a, n_b, n_c) > \rho_H,\ \ \rho_{n+1}(n_a, n_b, n_c) \\ \quad = \beta_H * \text{density}(n_a, n_b, n_c) \\ \text{elseif density } (n_a, n_b, n_c) < \rho_L,\ \ \rho_{n+1}(n_a, n_b, n_c) \\ \quad = \rho_n(n_a, n_b, n_c) - \beta_L * \text{density } (n_a, n_b, n_c) \\ \text{else } \rho_{n+1}(n_a, n_b, n_c) = \text{density } (n_a, n_b, n_c) \\ \phi_{n+1} = \text{angle}[\mathrm{FT}(\rho_{n+1})] \\ \rho_1 = 0, \phi_1 = \text{angle}[\mathrm{FT}(\text{random positive density})]. \end{cases} \qquad (2)$$

In equation (2), ϕ_n are the phases at the nth iteration, $\rho_n(n_a, n_b, n_c)$ is the electron density at grid (n_a, n_b, n_c) at the nth iteration and angle[...] means to obtain the phase angles. In equation (2), FT(...) and $\mathrm{FT}^{-1}(\ldots)$ mean Fourier and inverse Fourier transforms, respectively. β_H and β_L are two positive numbers, typically chosen between zero and one. In this paper, all HIO procedures are started with ($\beta_H = 0.2$, $\beta_L = 0.9$) (large correction) and after a certain number of iterations, typically as a quantum of 50, ($\beta_H = 0.9$, $\beta_L = 0.2$) (small correction) is used for the rest of the iterations. In equation (2), ρ_H is positive and often chosen around one. Throughout this paper, $\rho_L = 0$.

3. Examples

3.1. Three-dimensional Lennard-Jones carbon clusters

Lennard-Jones carbon cluster atom coordinates are obtained from the Cambridge Cluster Database. Each carbon cluster crystal is formed by packing the clusters in close contact in all three dimensions in an orthorhombic lattice. A 500-atom carbon cluster is chosen. The view along the z axis of the 500-atom carbon cluster is shown in Fig. 1(a), with the dots indicating atom centers. The atom–atom closest distance is 1.12 Å, and thus $R_0 = 0.56$ Å for the solid sphere is taken. The orthorhombic lattice parameters are found to be $a = 9.06$, $b = 10.23$ and $c = 10.44$ Å, corresponding to an occupancy ratio of 0.380. To mimic the noise level in the scattering data, $\text{Poisson}(\alpha_{\rm noise} I(\mathbf{Q}_{HKL}))/\alpha_{\rm noise}$ is used for the simulated Bragg peak intensity. Poisson(...) represents a Poisson distribution random-number generator with the argument as the mean. $\alpha_{\rm noise}$ is a positive number. The number of grids in the unit cell along the three cell dimensions are N_a, N_b and N_c, which are set to be odd positive numbers. The largest wavevector transfer is set to be $Q_{\max} = \pi(N_a - 1)/a$, beyond which the Bragg-peak intensities are deliberately set to be unavailable and the corresponding structure factor amplitudes are to float in the HIO procedure. This means that about half of the structure factors are to float.

Now take $N_a = N_b = N_c = 37$, which gives $Q_{\max} = 12.48$ Å^{-1} corresponding to a resolution of 0.50 Å. The Bragg peak at $\mathbf{Q}_{HKL} = 0$ is available [the case when $I(0)$ is unavailable will be discussed later in this paper]. At $\beta_S = 2.2$, $\rho_H = 1.2$ and 200×2 iterations, which means 200 iterations with ($\beta_H = 0.2$, $\beta_L = 0.9$) followed by 200 iterations with ($\beta_H = 0.9, \beta_L = 0.2$), Fig. 1($b$) shows a view along the z axis of a reconstructed 500-atom carbon cluster structure. In fact, the reconstructed structure is for the corresponding uniform solid-sphere cluster and this meaning is implied throughout this paper. The color of each dot in Fig. 1(b) represents the value of the electron density at that location. The individual solid spheres (composed of many dots) are clearly visible and their relative coordinates closely resemble those of the original object shown in Fig. 1(a). Translational shift and inversion of the obtained structure in relation to the original one can happen since both give the same Bragg peak intensities. To make Fig. 1(b) more appealing, the electron-density values at half of a grid are computed and shown through typical Fourier expansion formula for a periodic object. In Fig. 1(b), only those electron densities larger than 0.4 Å^{-3} are shown. The reconstructed 500-atom carbon cluster has a crystallographic R-factor of 0.158, defined by $\sum_{HKL} ||F|_{\rm HIO} - |F|_{\rm calc}| / \sum_{HKL} |F|_{\rm calc}$, where $|F|_{\rm calc}$ and $|F|_{\rm HIO}$ are the calculated and HIO-procedure-obtained structure factor amplitudes, respectively. Throughout this paper, R-factor always means the crystallographic R-factor. The signal-to-noise ratio (SNR) for Fig. 1(b) is 84.9.

To measure how much the reconstructed structure matches the original one, mean phase errors between the HIO-procedure-obtained phases and the phases calculated from the original solid-sphere cluster coordinates are computed for

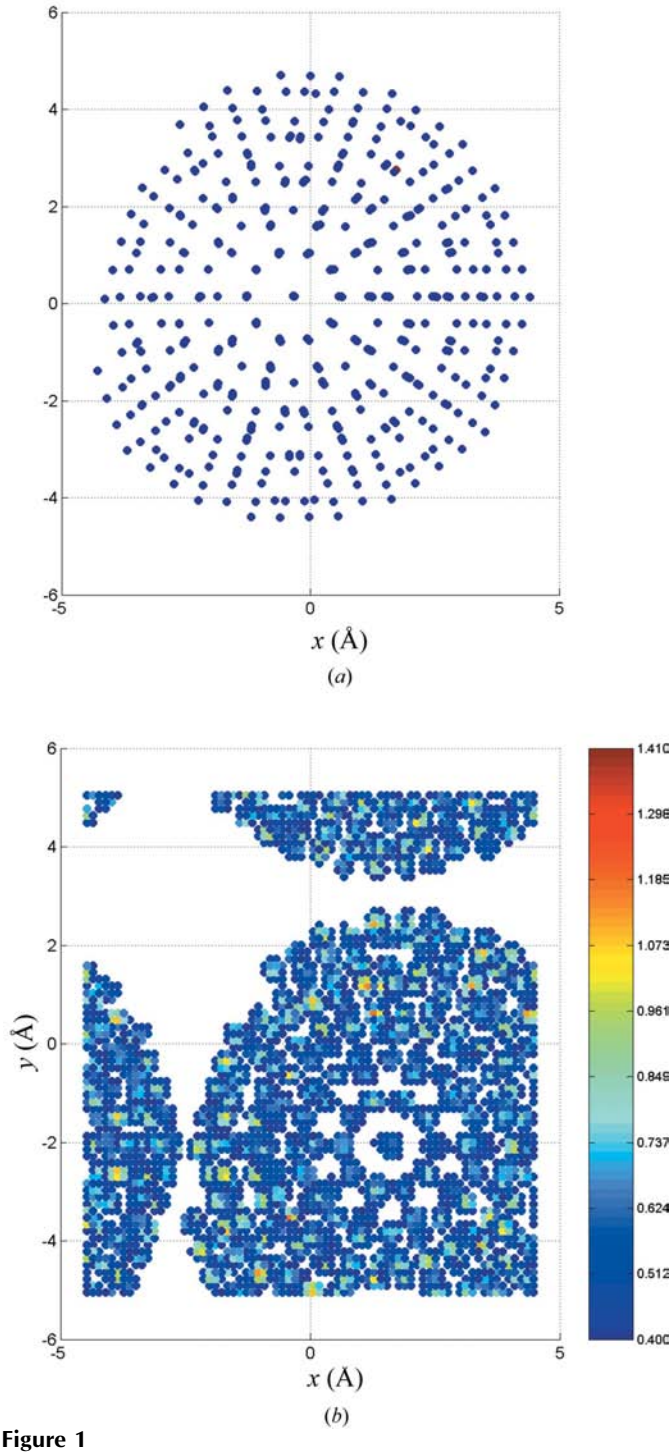

Figure 1
(*a*) View along the z axis of the 500-atom carbon cluster obtained from the Cambridge Cluster Database. Each dot represents an atom center. Cell dimensions are $a = 9.06$, $b = 10.23$, and $c = 10.44$ Å. (*b*) A reconstructed 500-atom carbon (actually uniform solid sphere) cluster viewed along the z axis, obtained from a HIO procedure with 200×2 iterations, at a resolution of 0.50 Å ($N_a = N_b = N_c = 37$), with $\beta_S = 2.2$ and $\rho_H = 1.2$. The first 200 iterations use ($\beta_H = 0.2$, $\beta_L = 0.9$) and the next 200 iterations use ($\beta_H = 0.9$, $\beta_L = 0.2$). The electron densities at half of a grid step are also computed through the standard Fourier transform and shown. Electron densities less than 0.4 Å^{-3} are not shown. Each small spot represents an electron density at a grid or half grid point. SNR = 84.9, R-factor = 0.158, $e\phi2$ = 0.312 and $e\phi3$ = 0.397. The central peak is available. The color bar indicates the electron-density values.

($|H| \leq 2, |K| \leq 2, |L| \leq 2$) and ($|H| \leq 3, |K| \leq 3, |L| \leq 3$) (computing the mean phase error for all Bragg peaks can be very time consuming). In calculating the mean phase errors, to take care of object translational shift, the obtained object is shifted one grid step size at a time to cover all the grid points (this can also be done at a half grid step size to gain more accuracy). In addition, the mean phase errors are computed for both the original object and its inversion. Among all the translational shifts and the inversion, the minimum mean phase error is selected to be the mean phase error between the obtained and original objects. The mean phase errors for Fig. 1(*b*) are quite small at 0.312 and 0.397 for ($|H| \leq 2, |K| \leq 2, |L| \leq 2$) and ($|H| \leq 3, |K| \leq 3, |L| \leq 3$), respectively. To simplify notation, in this paper, $e\phi2$ and $e\phi3$ are used to denote the mean phase errors for ($|H| \leq 2, |K| \leq 2, |L| \leq 2$) and ($|H| \leq 3, |K| \leq 3, |L| \leq 3$), respectively. That a higher Q_{HKL} creates a larger phase error is understandable since at a higher Q_{HKL} a small deviation in atom location generates a large phase shift due to the nature of the Fourier transform. In addition to their relatively small values, the translational shifts for $e\phi2$ and $e\phi3$ in Fig. 1(*b*) are the same, indicating that the obtained structure matches the original one well.

It is not the case that every run will end up with a very good picture like Fig. 1(*b*). The obtained structures from some runs do not resemble the original 500-atom carbon cluster at all, and large R-factors and mean phase errors occur. To measure the success rate, 20 HIO runs are generated at $\beta_S = 2.0$ with an SNR of 72.7. The rest of the HIO input parameters take the same values as those for Fig. 1(*b*) and the central peak is available. The R-factors and mean phase errors for the 20 runs are shown in Fig. 2. The figure shows that $e\phi2$ and $e\phi3$, as well as the R-factor, after most runs, are of satisfactorily small values, and the corresponding obtained structures look very similar to the original one. The occasionally large mean phase

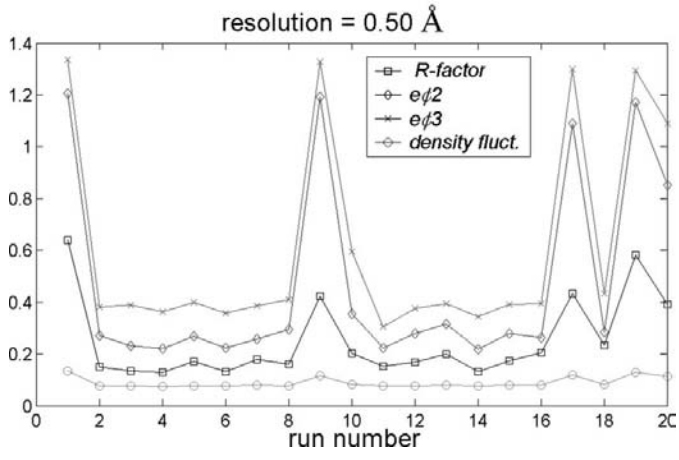

Figure 2
R-factors (square), $e\phi2$ (diamond), $e\phi3$ (cross) and density fluctuations (circle) for each of 20 HIO runs, for the 500-atom carbon cluster crystal, at a resolution of 0.50 Å ($N_a = N_b = N_c = 37$), SNR = 72.7, $\beta_S = 2.0$ and $\rho_H = 1.2$, with 200×2 iterations. The central peak is available.

errors are close to those from a random positive electron-density profile. In fact, $e\phi 2$ and $e\phi 3$ between a random positive electron-density structure and the 500-solid-sphere cluster are about 1.14 ± 0.06 and 1.31 ± 0.02, respectively. In Fig. 2, the density fluctuation is meant to test its usefulness in gauging how good the final reconstructed structure is, and is defined as the average difference between the electron density at one pixel to its six nearest neighbors. Mathematically,

$$\begin{aligned}\text{density fluctuation} &= \Bigg\{ \sum_{n_a=1,n_b=1,n_c=1}^{N_a,N_b,N_c} \sum_{m=-1,1} [|\rho(n_a, n_b, n_c) - \rho(n_a + m, n_b, n_c)| \\ &+ |\rho(n_a, n_b, n_c) - \rho(n_a, n_b + m, n_c)| + |\rho(n_a, n_b, n_c) \\ &- \rho(n_a, n_b, n_c + m)|] \Bigg\} (6N_aN_bN_c)^{-1}. \qquad (3)\end{aligned}$$

It is thought that when the obtained structure is faithful and the grid step size is not that large, namely, the solid sphere can cover many grid points, the density fluctuation tends to be small since the major density difference occurs only at the boundary between the sphere and the non-sphere region. It is obvious from the figure that the density fluctuation does behave synchronously to the *R*-factor and mean phase errors and has a value of about 0.047 for the faithfully reconstructed objects. Therefore, the density fluctuation does, as expected, in addition to the *R*-factor, offer a measure for the goodness of the obtained structure, at least for the case when high-resolution data are available.

To demonstrate how easy it is to apply the HIO algorithm to reconstruct the original carbon cluster crystal structure when high-resolution X-ray scattering data are available, Fig. 3(*b*) shows, in the view along the *z* axis, a reconstructed 1000-atom carbon cluster, at $\beta_S = 2.0$. The original 1000-atom carbon clusters are set to form a close-contact primitive orthorhombic lattice with unit-cell dimensions $a = 12.39$, $b = 12.02$ and $c = 12.55$ Å. In Fig. 3(*b*), $N_a = N_b = N_c = 37$, which gives a resolution of 0.52 Å. The SNR is about 200. The HIO-related parameters take the same values as those for Fig. 1(*b*), with 200×2 iterations. The central peak is available. The reconstructed carbon cluster (only densities larger than 0.4 are shown) clearly resembles the original one shown in Fig. 3(*a*). In Fig. 3(*b*), the *R*-factor is 0.103, $e\phi 2 = 0.185$, $e\phi 3 = 0.289$, and the density fluctuation is 0.062.

Now, let us study the cases when only low-resolution data are available. For the 500-atom carbon cluster crystal, take $N_a = N_b = N_c = 9$, corresponding to a resolution of 2.27 Å. The generated Bragg peaks have an SNR of 31. At $\beta_S = 0.94$, with central peak available, and the rest of the HIO-procedure-related parameter values the same as those for Fig. 1(*b*), a reconstructed 500-atom carbon cluster is shown in Fig. 4(*a*). The obtained object has an *R*-factor of 7.7×10^{-12}, $e\phi 2$ of 0.677, $e\phi 3$ of 0.839 and density fluctuation of 0.219. The small *R*-factor and mean phase errors do indicate a good reconstructed object. Visually, the shape of the reconstructed object resembles well the original one shown in Fig. 4(*b*). Note that $\beta_S < 1$ corresponds to expanding the spheres and thus the nearest neighbors overlap a little bit, creating a small region which has an electron density of 2 Å^{-3}. However, since the largest wavevector transfer is not that high at a value less than 12.5 Å^{-1}, the inverse Fourier transform will not show the high density but a much smoother electron-density profile between the neighboring spheres.

To measure the success rate for the low-resolution case, Fig. 5(*a*) plots the mean phase errors and density fluctuations (multiplied by 6) for each of 20 HIO runs, with $\beta_S = 0.94$, SNR = 31, and the rest of the HIO input parameter values the same as those for Fig. 4(*a*). Fig. 5(*a*) shows that, out of the 20 runs, quite a few reconstructed objects have small mean phase

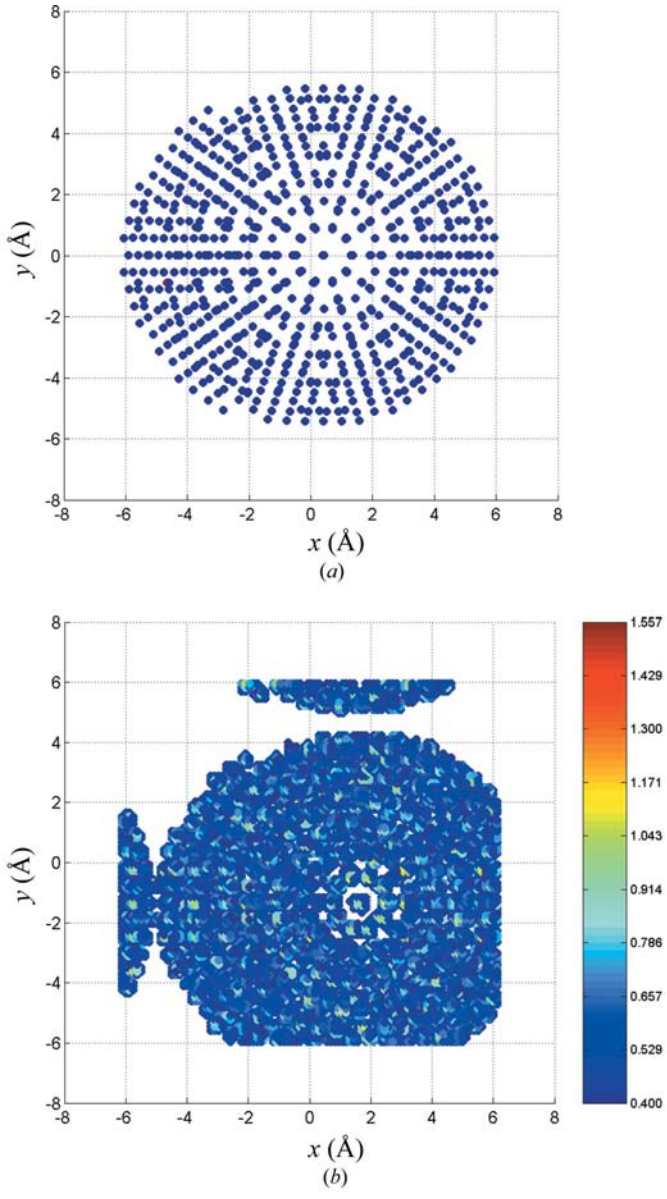

Figure 3
Like Fig. 1, structures viewed along the *z* axis, for the 1000-atom carbon cluster. The cell dimensions are $a = 12.39$, $b = 12.02$ and $c = 12.55$ Å. Resolution = 0.52 Å ($N_a = N_b = N_c = 37$), SNR = 200, $\beta_S = 2.0$, *R*-factor = 0.103, $e\phi 2 = 0.185$ and $e\phi 3 = 0.289$. The central peak is available.

errors, although the success rate is much less than that for the case when high-resolution data are available. Fig. 5(*a*) also indicates that the density fluctuation parameter does not behave monotonically with the mean phase errors (the *R*-factors are extremely small, close to zero). The intuition is that the grid step size is simply too large, relative to the atom diameter, to make the density fluctuation parameter a good measure. Note that the reconstructed object shown in Fig. 4(*a*) is from run 15 in Fig. 5(*a*). The extremely small *R*-factors for the case of $\beta_S = 0.94$ suggests that there are many ways to arrange the electron densities so that the computed Bragg-peak intensities match the measured ones, and therefore selecting the faithfully reconstructed objects is challenging. In order to make the electron-density fluctuation a good measure for selecting a faithfully reconstructed object, the solid spheres can be expanded a little more with $\beta_S = 0.87$, which actually results in increased electron-density smoothness between nearest spheres after the inverse Fourier transform. Fig. 5(*b*) plots the *R*-factors, mean phase errors and density fluctuations (multiplied by 6) from 20 HIO runs at $\beta_S = 0.87$ and the rest of the HIO input parameters take the same values as those for Fig. 5(*a*). We note that runs 9 and 15 in Fig. 5(*b*) have the smallest mean phase errors, and also have the smallest density fluctuations, although the differences from the others are small, suggesting that when data are available up to a resolution of 2.27 Å, by expanding the spheres with $\beta_S = 0.87$, the density fluctuation parameter may provide a good measure to select faithfully reconstructed objects.

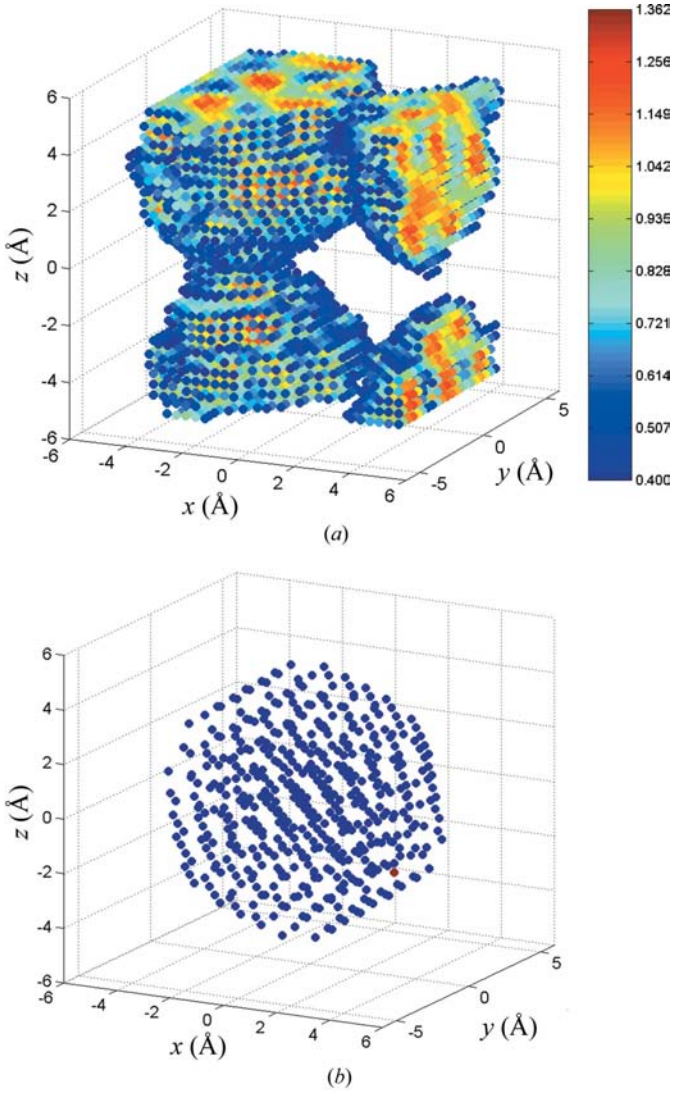

Figure 4
(*a*) A tilted view of a reconstructed 500-atom carbon cluster at a resolution of 2.27 Å ($N_a = N_b = N_c = 9$), SNR = 31 and $\beta_S = 0.94$. The central peak is available. The other HIO-related parameters have the same values as those for Fig. 1(*b*). *R*-factor = 7.7×10^{-12}, $e\phi2 = 0.677$ and $e\phi3 = 0.839$. The color bar indicates the electron-density values. (*b*) A tilted view of the original 500-atom carbon cluster, with the dots representing atom centers.

3.2. Two-dimensional crystals of carbon atom clusters

The scattering structure factor normalization and HIO procedure described in the last subsection can also be applied

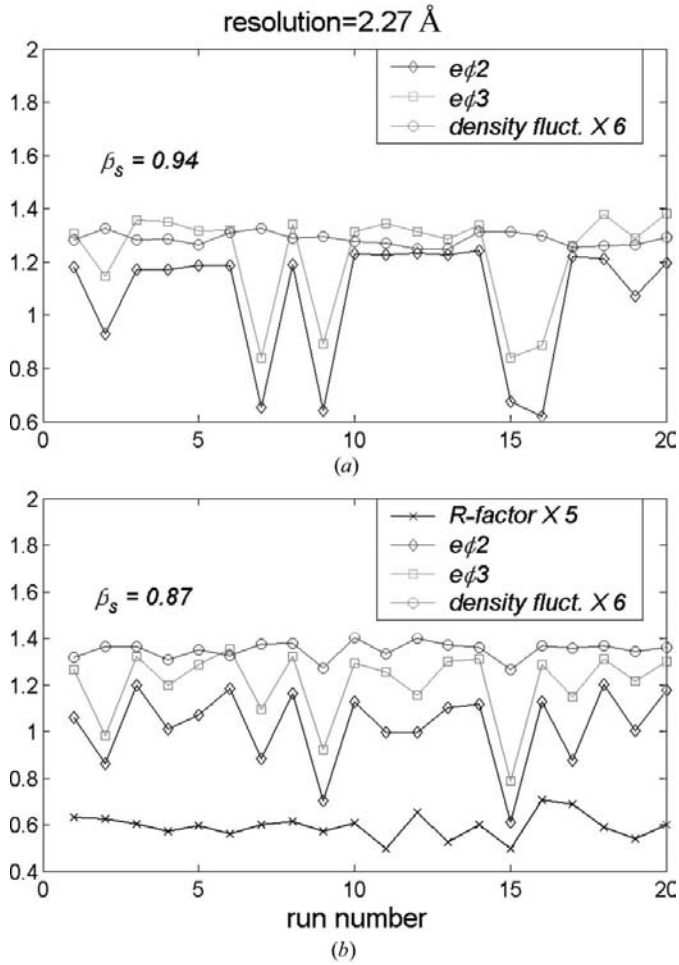

Figure 5
The mean phase errors $e\phi2$ (diamond) and $e\phi3$ (square), and density fluctuations (circle) from each of 20 HIO runs at $N_a = N_b = N_c = 9$ (resolution 2.27 Å), for the 500-atom carbon cluster crystal, with $\rho_H = 1.2$ and 200×2 iterations. The central peak is available. (*a*) SNR = 31 and $\beta_S = 0.94$. (*b*) SNR = 27 and $\beta_S = 0.87$. The density fluctuations shown are the real ones multiplied by 6.

to reconstruct two-dimensional crystal objects. As an example, a crystal of rings each of which is formed by 40 C atoms is studied. The rings form a close contact hexagonal lattice with unit-cell lengths of $a = b = 15.4$ Å and the angle between the unit-cell base vectors is $\gamma = \pi/3$ rad. The electron density of the original object before normalizing the C atoms to uniform discs is shown in Fig. 6(*a*). Fig. 6(*b*) shows the reconstructed ring with a resolution of 0.53 Å, namely, $Q_{\max} = 11.8$ Å^{-1}. The central peak is available. For Fig. 6(*b*), $N_a = N_b = 59$, the SNR is 31, $\beta_S = 2$, $\rho_H = 1.2$ and the number of iterations is 200×2. Fig. 6(*c*) shows the reconstructed object at a resolution of 1.9 Å ($N_a = N_b = 17$ and $Q_{\max} = 3.3$ Å^{-1}), with an

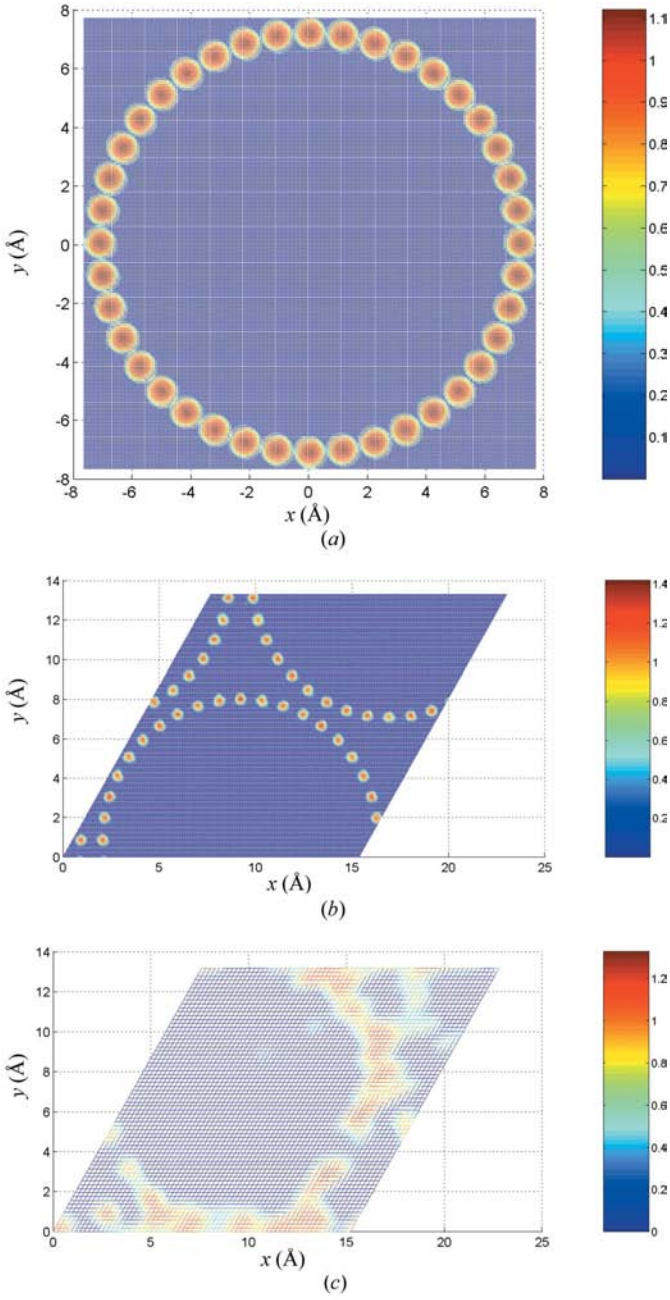

Figure 6
(*a*) Electron density for a hypothetical 40-atom carbon cluster ring. (*b*) A reconstructed object for the hexagonally packed 40-atom carbon rings (actually the corresponding 40 uniform discs), at a resolution of 0.53 Å. $N_a = N_b = 59$, SNR = 31, $\beta_S = 2$, $\rho_H = 1.2$ and the number of iterations is 200×2. (*c*) A reconstructed ring at $N_a = N_b = 17$, SNR = 16.1, $\beta_S = 0.867$ and the other parameters have the same values as those in (*b*). The central peak is available for both (*b*) and (*c*). For both (*b*) and (*c*), electron densities less than 0.2 Å^{-3} are shown as 0. Color bars indicate electron-density values.

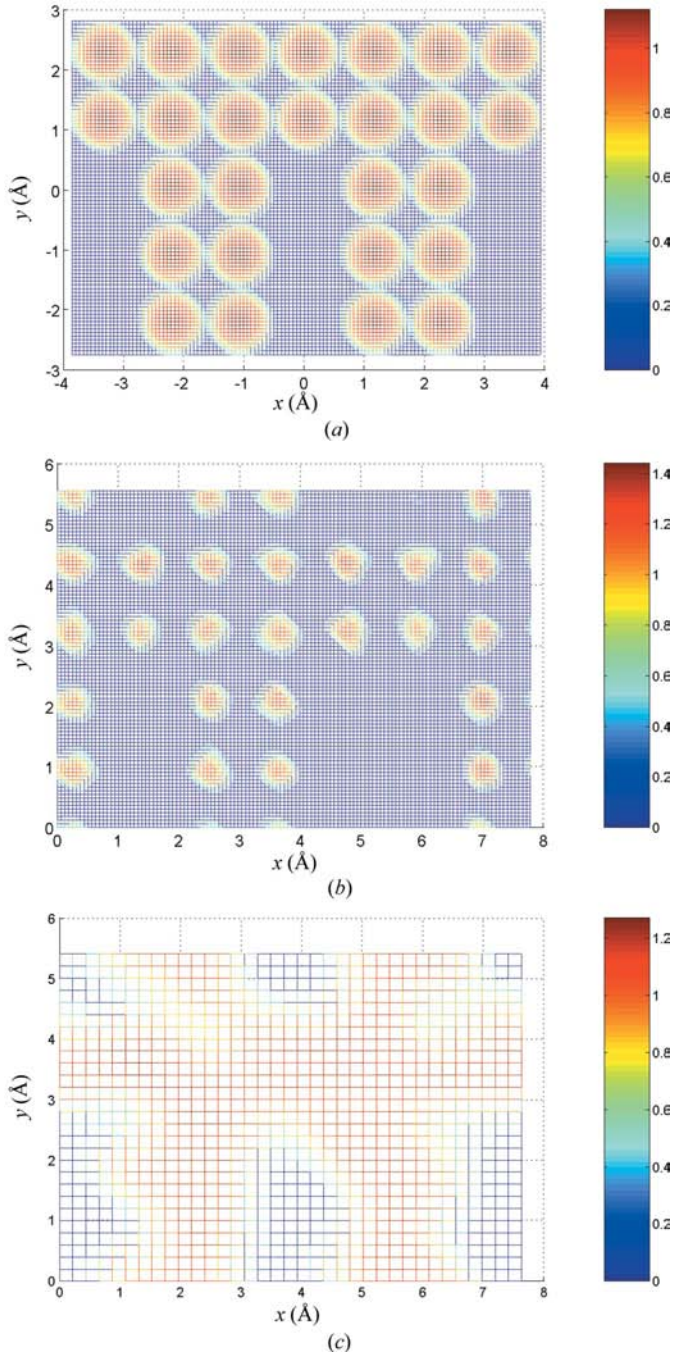

Figure 7
(*a*) Electron density for the original object: a Π-shaped C-atom cluster. (*b*) A reconstructed object, at a resolution of 0.52 Å, SNR = 32, $N_a = 31$, $N_b = 23$ and $\beta_S = 1.8$. (*c*) A reconstructed object at a resolution of 1.9 Å, SNR = 24, $N_a = 9$, $N_b = 7$ and $\beta_S = 0.867$. Both (*b*) and (*c*) use $\rho_H = 1.2$ and 200×2 iterations. Color bars indicate electron-density values. Electron densities less than 0.4 Å^{-3} are shown as 0.

SNR of 16.1. In Fig. 6(*c*), $\beta_S = 0.867$, and the rest of the HIO-related input parameters take the same values as those for Fig. 6(*b*). In Figs. 6(*b*) and 6(*c*), electron densities less than 0.2 Å^{-3}are shown as 0. The crystallographic *R*-factors for Figs. 6(*b*) and 6(*c*) are 0.077 and 0.0046, respectively.

Another two-dimensional crystal object studied is shown in Fig. 7. The original non-normalized Π-shaped object of C-atom cluster, shown in Fig. 7(*a*), form a rectangular lattice with unit-cell lengths of $a = 7.84$ and $b = 5.60$ Å. For Fig. 7(*b*), SNR = 31.9, a resolution of 0.52 Å ($N_a = 31$ and $N_b = 23$), $\beta_S = 1.8$, and an *R*-factor of 0.12. Fig. 7(*c*) has an SNR of 23.7, a resolution of 1.9 Å ($N_a = 9$ and $N_b = 7$), $\beta_S = 0.867$, and an *R*-factor of 0.0052. Electron densities less than 0.4 Å^{-3} are shown as 0. Figs. 7(*b*) and 7(*c*) use $\rho_H = 1.2$ and 200 × 2 iterations.

4. Effects of missing central peak and Bragg-peak noise

In a typical protein crystallographic experiment, the central peak is not available and therefore it is important to study the effect of the missing central peak. To investigate the missing-central-peak effect, the 500-atom carbon crystal structure is reconstructed against β_S with a floating $F(0)$. Twenty HIO runs with various levels of resolution with $Q_{\max}$ of 12.5, 8.3, 6.24, 4.16 and 2.77 Å^{-1}, corresponding to the number of grids of $N_a = N_b = N_c = 37$, $N_a = N_b = N_c = 25$,

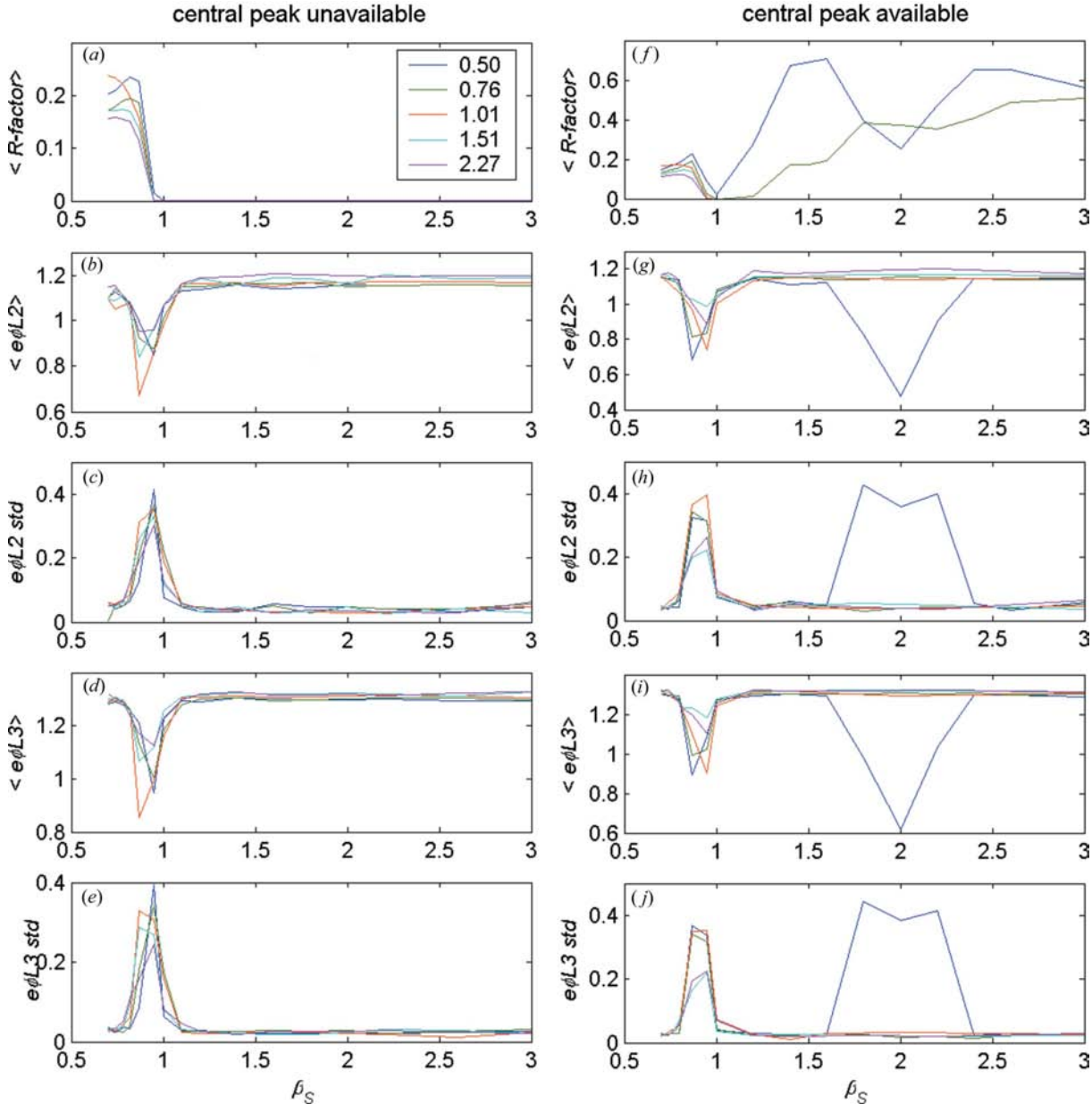

Figure 8
Plots of the averages of the *R*-factor, mean phase errors and mean phase error standard deviations *versus* β_S over 20 HIO runs for the 500-atom carbon cluster crystal at crystallographic resolutions of 0.50 Å (blue), 0.76 Å (green), 1.01 Å (red), 1.51 Å (cyan) and 2.27 Å (magenta). SNRs are around 100. $\rho_H = 1.2$ and 200 × 2 iterations. (*a*)–(*e*) No central peak. (*f*)–(*j*) With central peak.

$N_a = N_b = N_c = 19$, $N_a = N_b = N_c = 13$ and $N_a = N_b = N_c = 9$, are carried out. The SNRs are typically more than 100. In the HIO procedure, $\rho_H = 1.2$ and 200×2 iterations are taken. The results are plotted in Figs. 8(*a*)–(*e*). In the figure, std denotes the standard deviation, $\langle \ldots \rangle$ denotes the average, and the legend shows the line colors representing resolutions in Å. For comparison, runs similar to those in Figs. 8(*a*)–(*e*) are carried out with the central peak available, and the results are shown in Figs. 8(*f*)–(*j*). Two obvious effects are observed. First, the *R*-factors become very small if the central peak is unavailable at $1 < \beta_S \leq 3$, indicating much freedom in selecting an electron-density profile to fit the measured Bragg-peak intensities. Second, for the case of resolution of 0.50 Å, when the central peak is available, $\beta_S = 2$ is an excellent choice to faithfully reconstruct the original object, as indicated by the significant dips of the *R*-factor and mean phase errors. However, those desirable features are lost when the central peak is unavailable, suggesting that, even when very high resolution scattering data are available, without the central peak $\beta_S = 2$ may not yield a good reconstructed object. What is important is that the figures strongly suggest that the probability of obtaining desirable reconstructed objects with a β_S slightly less than one is relatively high and nearly unaffected by the availability of the central peak. Regardless of the central peak availability, with β_S around 0.87, the mean phase errors dip dramatically and the respective standard deviations increase. The large standard deviations show that there is a

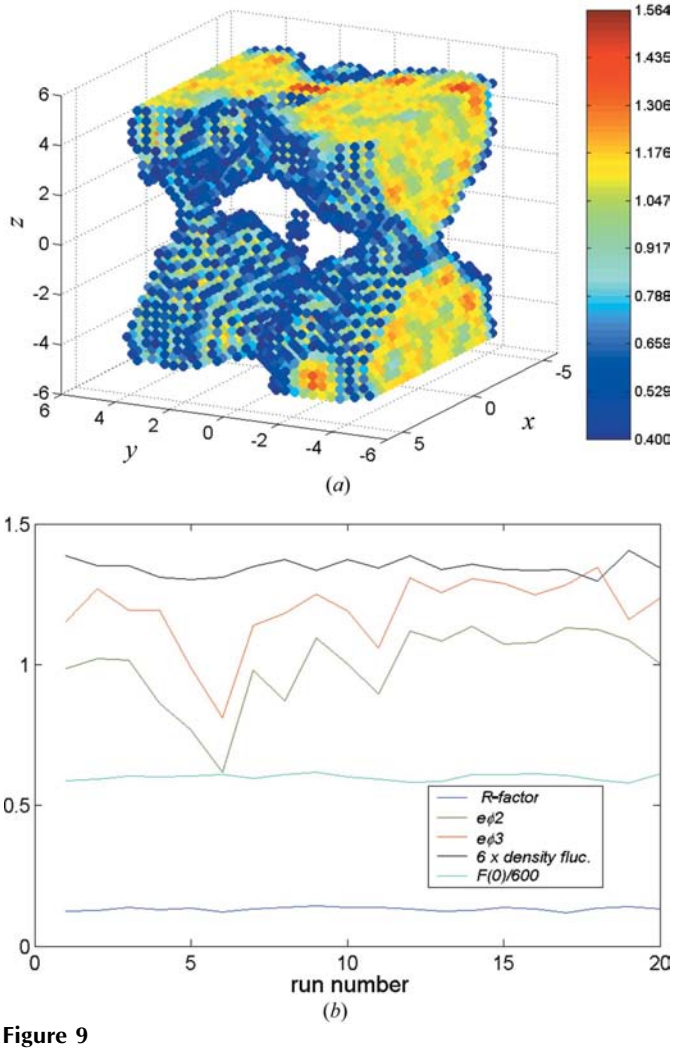

Figure 9
(*a*) A reconstructed 500-atom carbon cluster, corresponding to the minimum mean phase error among 20 HIO runs [see (*b*), run no. 6], at a resolution of 2.27 Å ($N_a = N_b = N_c = 9$), without the central peak. SNR = 37, $\beta_S = 0.87$, $\rho_H = 1.2$ and 200×2 iterations. Each dot represents the electron-density value by the color at that location. The units for all three axes are Å. (*b*) For each of the 20 runs, the *R*-factor (blue), $e\phi 2$ (green), $e\phi 3$ (red), electron-density fluctuation (black, multiplied by 6), and the computed central peak intensity (cyan), which is multiplied by $\beta_S^3/600$, are plotted.

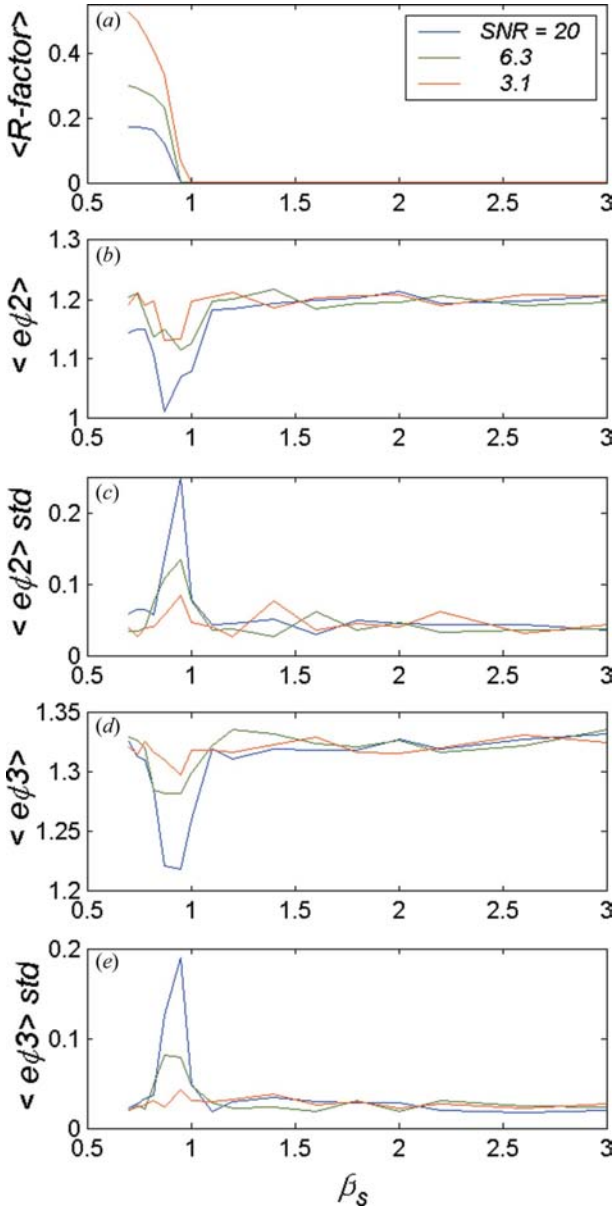

Figure 10
Plots of the averages of the *R*-factors, mean phase errors and mean phase error standard deviations *versus* β_S, for 20 HIO runs, for the 500-atom carbon cluster crystal, at SNR = 20 (blue), 6.3 (green) and 3.1 (red). $N_a = N_b = N_c = 9$ (resolution = 2.27 Å), $\rho_H = 1.2$ and 200×2 iterations.

substantial number of runs whose reconstructed objects have much lower mean phase errors.

To show that, even without the central peak, the HIO algorithm can still yield a faithfully reconstructed object, 20 runs for the 500-atom carbon cluster crystal are carried out, at $N_a = N_b = N_c = 9$ (resolution = 2.27 Å). The computer-generated Bragg peaks have an SNR of 37. The HIO procedure has a floating central structure factor, $\beta_S = 0.87$ and $\rho_H = 1.2$. As usual, 200×2 iterations are taken. Fig. 9(*a*) shows the reconstructed object which has the lowest mean phase errors among the 20 runs. The unit for the *x*, *y*, *z* axes is Å. Electron densities less than 0.4 Å^{-3} are not shown. The mean phase errors are well below those from a random electron-density profile, and the reconstructed object visually resembles well the original one at low resolutions. Fig. 9(*b*) plots some results from the 20 runs of the HIO procedure. The computed central structure factors multiplied by β_S^3 and divided by 600, indicated in the figure as $F(0)/600$, are very close to the theoretical value of 0.613. In fact, run no. 6 whose obtained structure is shown in Fig. 9(*a*) gives β_S^3 $F(0)$ = 365, which is very close to the true value of 368.

In a crystallographic experiment, the Bragg peaks may have a low SNR. To investigate the effect of the noise, especially when the SNR is low, Bragg peaks at several SNR levels for the 500-atom carbon cluster crystal, at $N_a = N_b = N_c = 9$ (resolution = 2.27 Å), are generated, without the central peak. HIO procedures with $\rho_H = 1.2$ and 200×2 iterations are carried out at several β_S. The results are plotted in Fig. 10. Note that each point in the figure is the result of 20 HIO runs. The levels of SNR are shown in the legend. It is clear from the figure that, with decreasing SNR, the *R*-factor increases. More significantly, over the region of β_S where faithfully reconstructed objects can be obtained, with decreasing SNR the mean phase errors increase whereas their respective standard deviations decrease, indicating that with decreasing SNR the HIO procedure is less and less able to reconstruct the original object, as expected. Furthermore, the figure shows that, at an SNR of 20, the HIO procedure is able to reconstruct the original object well, as indicated by the small *R*-factor and mean phase errors, which is not true when the SNR is at a very low value of 3.1. When the SNR is 6.3, the algorithm may marginally be able to faithfully reconstruct the original object.

Care is needed about the noise effect. It is known that, for a non-periodic object, when the SNR is low enough, the *R*-factor from the entire electron distribution (support plus non-support regions but not the support region only) given by the HIO procedure may increase with increasing number of iterations (unstable solution). To see the solution dependence on the iteration numbers with noisy data, Fig. 11 plots, for the 500-atom cluster crystal, at $N_a = N_b = N_c = 9$, the averages of the *R*-factors and $e\phi3$ over 20 runs *versus* β_S at three levels of SNR of 11, 6.1 and 2.9, with several different iteration numbers. $\rho_H = 1.2$ and the central peak is unavailable. In the figure, each column of subplots is for a particular level of SNR. The legends indicate the number of iterations. At all three levels of SNR, there is no clear indication that, with increasing number of iterations, either the *R*-factor or $e\phi3$ becomes larger. To examine the SNR effect on the reconstructed objects when high-resolution data are available, five runs at four levels of SNR of 2.8, 5.1, 9.9 and 20, for the 500-atom carbon cluster crystal with $N_a = N_b = N_c = 37$ were carried out. In the HIO procedure, $\rho_H = 1.2$, $\beta_S = 2$ and the central peak is available. Fig. 12 plots the averages of the *R*-factors *versus* iteration number. At all four levels of SNR, there are cases when increasing number of iterations comes with increasing *R*-factor values. Nevertheless, the *R*-factor oscillation range is small and the general trend is that a larger number of iterations corresponds to a smaller *R*-factor.

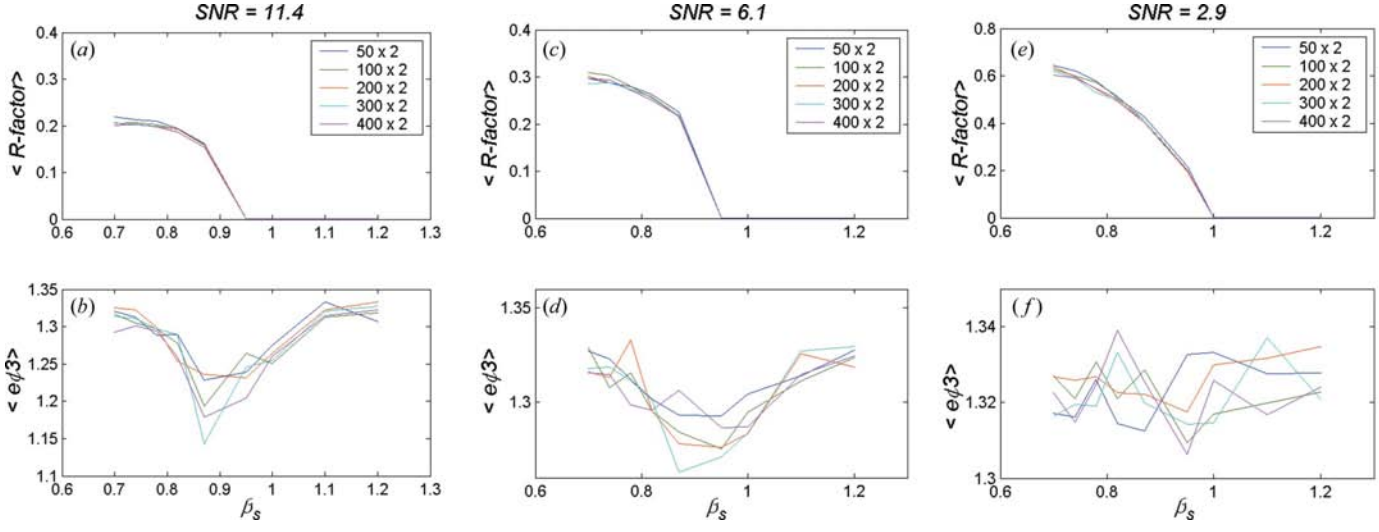

Figure 11
Plots of the averages of the *R*-factors and $e\phi3$ *versus* β_S, at several SNR levels, for 20 HIO runs, for the 500-atom carbon cluster crystal, at a resolution of 2.27 Å ($N_a = N_b = N_c = 9$) and iteration numbers of 50×2 (blue), 100×2 (green), 200×2 (red), 300×2 (cyan) and 400×2 (magenta). $\rho_H = 1.2$ and the central peak is unavailable. std indicates standard deviation. (*a*)–(*b*) SNR = 11. (*c*)–(*d*) SNR = 6.1. (*e*)–(*f*) SNR = 2.9.

5. Structure factor scale effect

The experimentally obtained structure factor amplitudes may be different from the theoretically calculated ones by a scale factor, due to reasons such as illuminated sample volume, detector sensitivity and incoming X-ray flux. To see how the scale factor affects the *R*-factor and the average electron density, 20 runs for the 500-atom carbon cluster with 12 different scale factors ranging from 0.1 to 2 are carried out. Each run uses 200×2 iterations, $N_a = N_b = N_c = 9$, $\beta_S = 1.0$, $\rho_H = 1.0$ and for all the runs the SNR is 680. The central peak is unavailable. The results are plotted in Fig. 13. In the figure, FScale denotes the scale factor so that the HIO input structure factor amplitude is FScale$|F|$, where $|F|$ is the theoretically computed structure factor amplitude. ⟨density⟩ denotes the average electron density over all the grid points. The figure shows that, when the scale factor FScale is much smaller than one, ⟨density⟩ approaches 0.5 while the *R*-factor is close to zero since in this case the upper bound in electron density in the HIO procedure $\rho_H = 1.0$ is quite large, equivalent to there being no upper limit constraint. When FScale approaches 1, ⟨density⟩ approaches the ideal value of 0.38 and starts to become larger, in proportional to FScale once FScale exceeds 1. The *R*-factor jumps once FScale exceeds 1.1.

6. Conclusions

Since a summation of Fourier series does not represent a sharp-featured object well, a natural choice is not to normalize the structure factors to those of point atoms. Normalizing the structure factors to those of solid spheres with uniform electron density has the advantage of establishing an upper limit for electron density over a wide region in real space. This apparently helps in the HIO procedure to faithfully reconstruct the original object, as demonstrated by the examples in this paper. The electron-density bounds limit the value which the phase can take for a newly added term in the Fourier-series summation for the electron density, and force the old phases (already in the series) to adjust to match the density bound criteria. Therefore, the more terms in the summation, the tighter are the phase values. Using 500- and 1000-atom carbon cluster crystals as examples, in the case of the availability of the Bragg peaks up to resolutions of about the atom radius, with the central peak available, the original objects can easily be reconstructed through the HIO procedure, with the solid sphere radius set at about half of the atom radius. In addition, results in this paper suggest that, with Bragg peaks available up to any resolution, including high resolution, even without the central peak, with a proper choice of (β_S, ρ_H), such as $(\beta_S = 0.87, \rho_H = 1.2)$, the original object can still be reconstructed well. In contrast, if the structure factors are not normalized, the HIO procedure gives for the 500-atom carbon cluster crystal (SNR = 35), at a low resolution of 2.27 Å, $\langle e\phi_2 \rangle = 1.16$ and $\langle e\phi_3 \rangle = 1.30$, and, at a high resolution of 0.50 Å, $\langle e\phi_2 \rangle = 1.25$ and $\langle e\phi_3 \rangle = 1.36$. The large mean phase errors with the non-normalized structure factors are not different from those from random positive electron densities. We note that the electron density of a C atom is sharply peaked around the atom center. Normalizing the structure factors to point atoms is equivalent to normalizing to solid spheres with an infinitely large β_S. Fig. 8 suggests that, when $\beta_S \gg 1$, the mean phase errors will be very large. In fact, when normalizing to point atoms, at a resolution of 0.50 Å, for the 500-atom carbon cluster crystal, the HIO procedure gives $\langle e\phi_2 \rangle = 1.18$ and $\langle e\phi_3 \rangle = 1.30$ whereas a normalized-to-solid-sphere procedure with $\beta_S = 2$ gives $\langle e\phi_2 \rangle = 0.48$ and $\langle e\phi_3 \rangle = 0.62$. Examples of reconstructing two-dimensional crystal structures at both high and low resolutions are given as

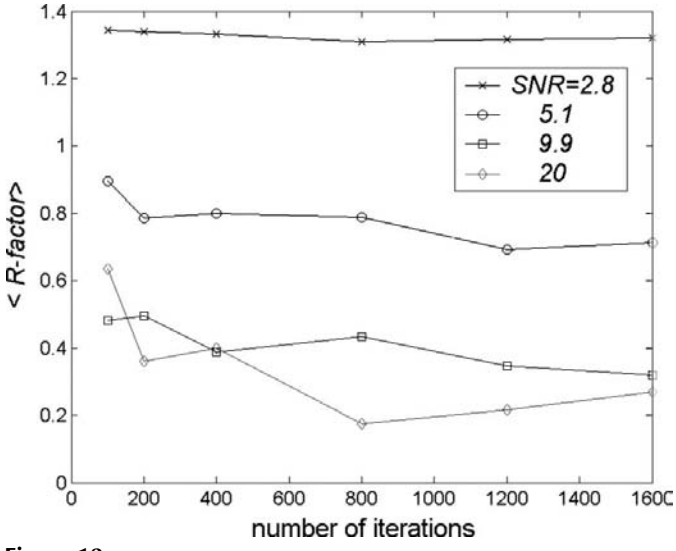

Figure 12
Plots of the averages of the *R*-factors *versus* iteration numbers for five HIO runs, at a resolution of 0.50 Å ($N_a = N_b = N_c = 37$), for the 500-atom carbon cluster crystal, at an SNR of 2.8 (cross), 5.1 (circle), 9.9 (square), and 20 (diamond), respectively. $\beta_S = 2$ and $\rho_H = 1.2$. The central peak is available.

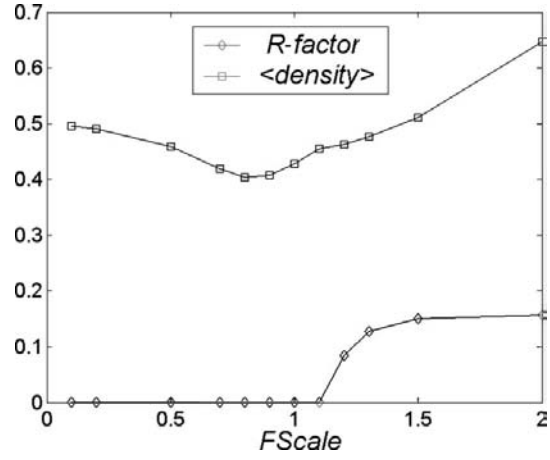

Figure 13
Plots of the *R*-factors (diamond) and average electron densities (square) *versus* the structure factor scale factor, averaged over 20 HIO runs, for the 500-atom carbon cluster, at a resolution of 2.27 Å ($N_a = N_b = N_c = 9$), SNR = 680, $\beta_S = 1.0$, $\rho_H = 1.0$ and 200×2 iterations. The central peak is unavailable.

well. Good results of solving many other two- and three-dimensional crystal structures have also been obtained.

When the real-space grid is fine enough, *i.e.* high-resolution data are available, the crystallographic *R*-factor and electron-density fluctuation are all good indicators of whether a faithful solution is found. However, when only low-resolution data are available, in general, there are not enough terms in the Fourier-series summation to limit the phases to tight regions and, therefore, multiple solutions can occur. We approach this problem by enlarging the solid-sphere radius so that the electron-density fluctuation remains a good measure. In the case of sparsely distributed atoms in a crystal, when only low-resolution data are available, the spheres may need to be enlarged two to three times (and correspondingly the electron-density upper limit ρ_H) of their original size.

We also studied the effects of noise and the scale factor for the Bragg peaks. Our results suggest that, when the SNR is larger than 20, the noise impact on obtaining a solution may be small. We have shown that, when the scale factor is larger than the real one, the obtained average electron density is proportional to the scale factor, and, when the scale factor exceeds the real one by 10%, the *R*-factor starts to become much bigger.

It has been shown (unpublished) that, for a hypothetical two-dimensional crystal composed of two types of discs of the same radii but different electron densities (one type has an electron density of one and the other of two), the algorithm reconstructs the object well when high-resolution data are available, with the electron-density upper limit ρ_H set slightly above two. Whether the algorithm can be successfully applied to solve real crystal structures composed of different types of atoms is unsure. One major reason is that the ratio of the structure factor amplitudes from two different types of atoms is not a constant over the entire available Bragg-peak locations. But the variation may be small enough to give a solution. In the case of a crystal composed of atoms of carbon, nitrogen and oxygen, the denominator in equation (1) can be the average of the structure factor amplitudes of carbon and oxygen.

I thank V. Elser and D. R. Luke for useful communications, and Hongliang Xu for explaining to me his direct phasing method. This research was supported by NIH grant RR07707 to Keith Moffat.

References

Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006). *J. Opt. Soc. Am.* **23**, 1179–1200.
Elser, V. (2003*a*). *J. Opt. Soc. Am.* **20**, 40–55.
Elser, V. (2003*b*). *Acta Cryst.* A**59**, 201–209.
Fienup, J. R. (1982). *Appl. Optics*, **21**, 2758–2769.
Gerchberg, R. W. & Saxton, W. O. (1972). *Optik (Stuttgart)*, **35**, 237–246.
He, H., Marchesini, S., Howells, M., Weierstall, U., Chapman, H., Hau-Riege, S., Noy, A. & Spence, J. C. H. (2003). *Phys. Rev. B*, **67**, 174114.
Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., Weierstall, U. & Spence, J. C. H. (2003). *Phys. Rev. B*, **68**, 140101.
Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.
Miao, J., Ishikawa, T., Anderson, E. H. & Hodgson, K. O. (2003). *Phys. Rev. B*, **67**, 174104.
Miao, J., Ishikawa, T., Johnson, B., Anderson, E. H., Lai, B. & Hodgson, K. O. (2002). *Phys. Rev. Lett.* **89**, 088303.
Miller, R., DeTitta, G. T., Langs, R. J. D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* A**60**, 134–141.
Oszlányi, G. & Sütő, A. (2005). *Acta Cryst.* A**61**, 147–152.
Oszlányi, G., Sütő, A., Czugler, M. & Parkanyi, L. (2006). *J. Am. Chem. Soc.* **128**, 8392–8393.
Weeks, C. M., Detitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* D**49**, 179–181.
Wu, J. S., Spence, J. C. H., O'Keeffe, M. & Groy, T. L. (2004). *Acta Cryst.* A**60**, 326–330.
Zheng, S., Strzalka, J., Ma, C., Opella, S. J., Ocko, B. M. & Blaize, J. K. (2001). *Biophys. J.* **80**, 1837–1850.