Acta Crystallographica Section A
Foundations and Advances
ISSN 2053-2733

Received 16 October 2013
Accepted 30 March 2014

A multiple-common-lines method to determine the orientation of snapshot diffraction patterns from single particles

Liang Zhou, Tian-Yi Zhang, Zhong-Chuan Liu, Peng Liu and Yu-Hui Dong*

Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, 19B Yuquan Road, Shijingshan District, Beijing, People's Republic of China. Correspondence e-mail: dongyh@ihep.ac.cn

With the development of X-ray free-electron lasers (XFELs), it is possible to determine the three-dimensional structures of noncrystalline objects with coherent X-ray diffraction imaging. In this diffract-and-destroy mode, many snapshot diffraction patterns are obtained from the identical objects which are presented one by one in random orientations to the XFEL beam. Determination of the orientation of an individual object is essential for reconstruction of a three-dimensional structure. Here a new method, called the multiple-common-lines method, has been proposed to determine the orientations of high- and low-signal snapshot diffraction patterns. The mean errors of recovered orientations (α, β, γ) of high- and low-signal patterns are about 0.14, 0.06, 0.12 and 0.77, 0.31, 0.60°, respectively; both sets of errors can meet the requirements of the reconstruction of a three-dimensional structure.

1. Introduction

Up to now, X-ray crystallography has been the primary methodology to solve the structures of objects at atomic resolutions. Most of the known structures in the Protein Data Bank (PDB) have been obtained by this way. However, obtaining the high-quality crystals for diffraction is a bottleneck in the X-ray crystallography technique. Most of the viruses, cells and around 40% of protein molecules, especially the membrane proteins, are difficult to crystallize. To determine their structures and understand their functions, a new imaging modality has been developed and that is coherent X-ray diffraction imaging.

It has been proposed that the two-dimensional coherent diffraction patterns of noncrystalline objects could be oversampled and the structures could be recovered by using an oversampling technique (Sayre *et al.*, 1998). Coherent X-ray diffraction imaging has been experimentally demonstrated to determine the two-dimensional structure of noncrystalline inorganic (Miao *et al.*, 1999) and biological specimens (Song *et al.*, 2008). In 2000, from computer simulations, it was found that experiments using very high X-ray dose rates and ultra-short exposures could provide useful structural information before radiation damage destroyed the object (Neutze *et al.*, 2000). With the development of X-ray free-electron lasers (XFELs), such pulses have become a reality and high-quality snapshot diffraction patterns from noncrystalline objects have been achieved (Seibert *et al.*, 2011). When the orientations of all patterns are determined, the three-dimensional structure can be reconstructed by an oversampling technique (Sayre *et al.*, 1998; Miao *et al.*, 1999).

Generally, there are two methods that can solve the orientations, one is a common-line method and the other is an iterative method. Based on the common-line method, Shneerson *et al.* (2008) first proposed a algorithm which can only determine the orientation of a flat Ewald sphere and requires the patterns to have a high signal-to-noise ratio (about 10 photons per pixel); in addition, the mean error of recovered orientations is large, about 3.5°. Bortel & Tegze (2011) proposed a common-arc method, which can determine the orientation of a low-signal pattern with more accuracy by selecting and averaging the $N-1$ candidate orientations (N is the number of patterns), but this method requires relative orientations of all pattern pairs, which will increase computational time by an (N/2)-fold amount. Yefanov & Vartanyants (2013) also proposed a similar algorithm to determine the orientation of a low-signal pattern. Essentially, all these algorithms do not overcome a drawback of the common-line method, which is that the relative orientation between two patterns is only determined by intensity along the common arc, and this is a small part of a diffraction pattern. In principle, this drawback demands high-signal-to-noise patterns, which can be obtained by classifying and averaging the weak-signal patterns (Huldt *et al.*, 2003; Bortel & Faigel, 2007; Bortel *et al.*, 2009; Tokuhisa *et al.*, 2012) or directly increasing the intensity of the XFEL pulse and the molecular weight of the object illuminated (Raines *et al.*, 2010). There are two iterative methods, one is based on the generative topographic mapping (GTM) algorithm (Fung *et al.*, 2009; Schwander *et al.*, 2010)

and the other is based on the expectation maximization compression (EMC) algorithm (Loh & Elser, 2009; Loh *et al.*, 2010). In the two methods, the classification and orientation processes are merged into a single iterative algorithm. It was shown later that the two methods are fundamentally the same (Moths & Ourmazd, 2011). The two iterative methods can work well for weak-signal diffraction patterns, but they are computationally very expensive in realistic cases. Recently, another iterative method, based on correlation maximization, was proposed (Tegze & Bortel, 2012, 2013); this algorithm is a much simplified version of the EMC algorithm and more efficient with respect to the GTM and EMC algorithms. Here we propose a new approach, which can not only solve the orientations of high-signal patterns but also low-signal patterns. Our approach mainly consists of two steps. In the first step, the single-common-line method is used to determine the initial orientations of each pattern. In the second step, the multiple-common-lines method is used to refine the orientations in an iterative way. It has been shown that our method can also solve the effects of twinning (Kirian *et al.*, 2010) in femtosecond X-ray protein nanocrystallography (Zhou *et al.*, 2013). In the following sections, we first describe our method in detail and then we apply it to high- and low-signal patterns, respectively; finally we reconstruct the three-dimensional structure using the high-signal patterns.

2. Determining the orientations using single-common-line method

For a noncrystalline object with orientation (α,β,γ) illuminated by a pulse from an XFEL with a wavelength λ, its snapshot diffraction patterns can be expressed as (Bortel *et al.*, 2009)

$$\mathrm{ccd}(n_1, n_2) = I(\mathbf{k}) = \Phi_{\mathrm{inc}} r_{\mathrm{e}}^2 \left|F(\mathbf{k})\right|^2 \Omega_{\mathrm{pix}}(\mathbf{k}), \tag{1}$$

where $\mathrm{ccd}(n_1,n_2)$ is the diffracted photon flux in the detector pixel (n_1,n_2), r_{e} is the classical electron radius, Φ_{inc} is the flux density of the incoming X-ray pulse, Ω_{pix} is the solid angle of the detector pixel, $F(\mathbf{k})$ is the structure factor, $\mathbf{k}$ is the scattering vector corresponding to the detector pixel (n_1,n_2) in orientation (α,β,γ), α represents roll, β represents pitch and γ represents yaw (Sidi, 1997). For convenience, we can arbitrarily select the orientation of one pattern as the reference orientation and define it as $(0,0,0)$. Here we define $\mathbf{k}_0$ as the scattering vector corresponding to the same pixel (n_1,n_2) in the reference orientation. The relationships between the reciprocal vectors $\mathbf{k}$ and $\mathbf{k}_0$ can be expressed as

$$\begin{aligned}
\mathbf{k}_0 &= (k_x, k_y, k_z)\\
k_x &= \frac{n_1[(n_1 d)^2 + (n_2 d)^2]^{1/2}}{\lambda(n_1^2 + n_2^2)^{1/2}[(n_1 d)^2 + (n_2 d)^2 + L^2]^{1/2}}\\
k_y &= \frac{n_2[(n_1 d)^2 + (n_2 d)^2]^{1/2}}{\lambda(n_1^2 + n_2^2)^{1/2}[(n_1 d)^2 + (n_2 d)^2 + L^2]^{1/2}}\\
k_z &= \frac{L}{\lambda[(n_1 d)^2 + (n_2 d)^2 + L^2]^{1/2}} - \frac{1}{\lambda}\\
\mathbf{k} &= \mathbf{R}_{\mathbf{x}}(\alpha)\mathbf{R}_{\mathbf{y}}(\beta)\mathbf{R}_{\mathbf{z}}(\gamma)\mathbf{k}_0\\
\mathbf{R}_{\mathbf{z}}(\gamma) &= \begin{pmatrix} \cos\gamma & \sin\gamma & 0\\ -\sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1 \end{pmatrix}\\
\mathbf{R}_{\mathbf{y}}(\beta) &= \begin{pmatrix} \cos\beta & 0 & -\sin\beta\\ 0 & 1 & 0\\ \sin\beta & 0 & \cos\beta \end{pmatrix}\\
\mathbf{R}_{\mathbf{x}}(\alpha) &= \begin{pmatrix} 1 & 0 & 0\\ 0 & \cos\alpha & \sin\alpha\\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix},
\end{aligned} \tag{2}$$

where d is the width of a pixel and L is the distance between the sample and detector. Here we define the Z direction of the particle coordinate as parallel to the incident direction of the X-ray beam and $\mathbf{R}_{\mathbf{z}}(\gamma)$, $\mathbf{R}_{\mathbf{y}}(\beta)$ and $\mathbf{R}_{\mathbf{x}}(\alpha)$ represent the object rotation around the Z, Y and X axes, respectively, in a counterclockwise direction when looking towards the origin.

The amplitudes of structure factors $|F(\mathbf{k})|$ of the identical particles which are used to solve the structure can be easily obtained if the orientations of all snapshot diffraction patterns are known, according to equations (1) and (2). For convenience, $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ is used to represent the amplitudes of the structure factors calculated from the nth pattern in orientation (α,β,γ), $|F_{0,0,0,1}(\mathbf{k})|$ is used to represent the amplitudes of the structure factors corresponding to first pattern (reference pattern) in orientation $(0,0,0)$. There is an intersection arc between $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ and $|F_{0,0,0,1}(\mathbf{k})|$ in reciprocal space as shown in Fig. 1; theoretically the values of $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ and $|F_{0,0,0,1}(\mathbf{k})|$ are equal along this intersection arc, which is called a common line, if (α,β,γ) is the actual orientation of the nth pattern. Therefore, the actual orientation of the nth pattern can be recovered by scanning the angles (α,β,γ) and calculating the average difference among regular grids passed through by the arc of intersection between $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ and $|F_{0,0,0,1}(\mathbf{k})|$. $R_{\mathrm{factor}}(\alpha,\beta,\gamma,n)$ is used to represent the average difference; it can be expressed as

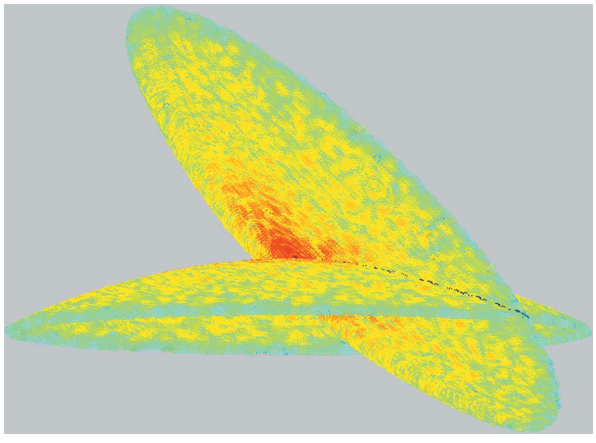

Figure 1
Intersection arc of two patterns with different orientations in reciprocal space.

$$R_{\rm factor}(\alpha,\beta,\gamma,n) = \sum_{hkl} \frac{||F_{\alpha,\beta,\gamma,n}(h,k,l)| - |F_{0,0,0,1}(h,k,l)||}{|F_{0,0,0,1}(h,k,l)|} \Big/ \sum_{hkl} 1, \quad (3)$$

where the integer set *hkl* is the indices of regular grids in reciprocal space passed through by the arc of intersection. The values of $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ and $|F_{0,0,0,1}(h,k,l)|$ can be easily obtained by interpolating $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ and $|F_{0,0,0,1}(\mathbf{k})|$, respectively, onto regular grids in reciprocal space using nearest neighbor interpolation. It should be noted that when the pattern is interpolated onto regular grids in reciprocal space, it is possible that one regular grid in reciprocal space, especially in high resolution, would contain more than one pixel in a detector and, to a certain extent, this will increase the signal-to-noise of $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$. Moreover, the centrosymmetry of the diffraction patterns has been included in our calculation, so there are four intersection arcs (Huldt *et al.*, 2003). It is worth noting that this feature makes calculating $R_{\rm factor}(\alpha,\beta,\gamma,n)$ redundant. By scanning a different set of (α,β,γ), if there is a global minimum of $R_{\rm factor}(\alpha,\beta,\gamma,n)$ found in one orientation (α,β,γ), then this orientation is considered as the orientation of the *n*th pattern.

The searching ranges of α, β and γ are $[-180^\circ, 180^\circ]$, $[-90^\circ, 90^\circ]$ and $[-180^\circ, 180^\circ]$, respectively. It should be noted that an exhaustive search is computationally very expensive, even when the search interval is 1°. In this paper, the highest resolution R is 6.8 Å in all patterns, the diameter O of the sample is about 345 Å and the required angular resolution is $R/2O$ radian (Shneerson *et al.*, 2008), about 0.56°; therefore the smallest search interval should be smaller than 1°. In our work, it is set to be 0.2°. The detailed steps of the searching algorithm are listed below.

(1) Performing the searching in angular steps of 5° initially. Calculating and sorting the $R_{\rm factor}(\alpha,\beta,\gamma,n)$ values in each orientation (α,β,γ), then obtaining N_1 candidate orientations corresponding to the smallest $R_{\rm factor}(\alpha,\beta,\gamma,n)$ values. In this paper, we performed the searching in whole orientation space for 20 patterns and found that setting the value of N_1 to 150 is enough to include the correct orientation. Owing to the computational capabilities in our lab, the search range of α, β and γ is limited to $40 \times 40 \times 40^\circ$ near the actual orientation and N_1 is set to 10.

(2) Then performing the searching near the N_1 candidate orientations, in angular steps of 1°. The search ranges are $[\alpha - 4^\circ, \alpha + 4^\circ]$, $[\beta - 4^\circ, \beta + 4^\circ]$ and $[\gamma - 4^\circ, \gamma + 4^\circ]$, where (α,β,γ) is one of the N_1 candidate orientations. Then, N_2 new candidate orientations, also corresponding to the smallest $R_{\rm factor}(\alpha,\beta,\gamma,n)$, are obtained after sorting; N_2 is usually set to 10–15.

(3) Finally, performing the searching near the N_2 new candidate orientations in angular steps of 0.2°. The search ranges are $[\alpha - 0.8^\circ, \alpha + 0.8^\circ]$, $[\beta - 0.8^\circ, \beta + 0.8^\circ]$ and $[\gamma - 0.8^\circ, \gamma + 0.8^\circ]$, where (α,β,γ) is one of the N_2 candidate orientations. The minimum value of $R_{\rm factor}(\alpha,\beta,\gamma,n)$ and its corresponding orientation are selected as the solution.

The computational time of the single-common-line method is directly proportional to the number of searches S_{N_1,N_2}. When the search is performed in the whole orientation space, S_{N_1,N_2} is about 220 000, where N_1 and N_2 are 150 and 15, respectively; it takes about 75 min on a desktop computer with a 3.40 GHz Inter(R) Core i7-3770 processor. The values of $N_1 \sim 150$ and $N_2 \sim 15$ are obtained according to simulated patterns; in a real case, they may not be accurate and should be adjusted according to the experimentally recorded patterns. S_{N_1,N_2} is much smaller than the number of searches $360 \times 360 \times 180$ (23 328 000), where (α,β,γ) varies sequentially with an angular step of 1°. Therefore, this searching algorithm can greatly reduce the computational time.

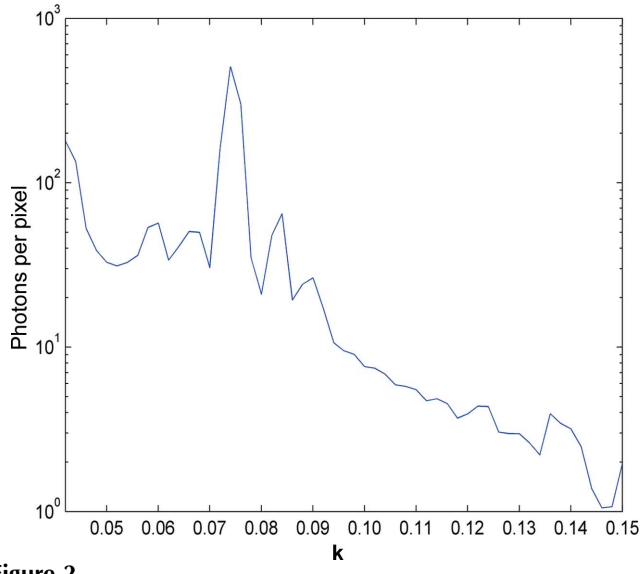

Figure 2
Average photon counting per detector pixel *versus* vector $\mathbf{k}$ (Å^{-1}) in reciprocal space; it is about 1.0 in the outer pixels.

3. Simulations

We downloaded from the Protein Data Bank a virus capsid (2tbv; Hopper *et al.*, 1984), which has a large molecular weight, as the model structure. The molecular weight of the capsid is about 5000 kDa and the diameter O is about 345 Å. The virus capsid is embedded into a three-dimensional array of $64 \times 64 \times 64$ voxels after Gaussian broadening; we take this matrix as our simulated object. The structure factors on each detector pixel are calculated from the electron densities in the voxels by Fourier transformation. The snapshot diffraction patterns can be calculated according to equation (1). We simulate an XFEL pulse with a wavelength of 5.0 Å and 5×10^{13} photons per pulse that was focused to a 100 nm spot. The detector contains 600×600 pixels with a width of 100 µm, the sample-to-detector distance is 27.6 mm (providing a sampling frequency of about 4 in low resolution). In total, 1000 snapshot diffraction patterns with different orientations (including the reference orientation) are simulated. The orientation of each pattern is generated by uniform random distribution. The intensities at the central 7×7 pixels are removed and Poisson noises are added to these patterns.

The average photon counts per pixel in different resolutions are shown in Fig. 2; the photon count in the outer pixels is about 1.0. The values of $|F_{\alpha,\beta,\gamma,n}(\mathbf{k})|$ and $|F_{0,0,0,1}(\mathbf{k})|$ are interpolated into a three-dimensional array of 256 × 256 × 256 voxels and $R_{\rm factor}(\alpha,\beta,\gamma,n)$ is calculated according to equation (3). As shown in Fig. 3, there is a minimum $R_{\rm factor}(\alpha,\beta,\gamma,21)$ nearby the actual orientation of the 21st pattern. However, there is a relatively large error in the $\alpha\gamma$ plane (Fig. 3*b*). For our sample of the 21 patterns, a typical comparison of the recovered orientations with the actual orientations is shown in Table 1. For a total of 1000 patterns, the mean errors of the orientations (α,β,γ) are about 0.66, 0.14 and 0.51°. According to the angular resolution function (Shneerson *et al.*, 2008), the required angular resolution is about 0.56° and so the recovered orientations are basically accurate.

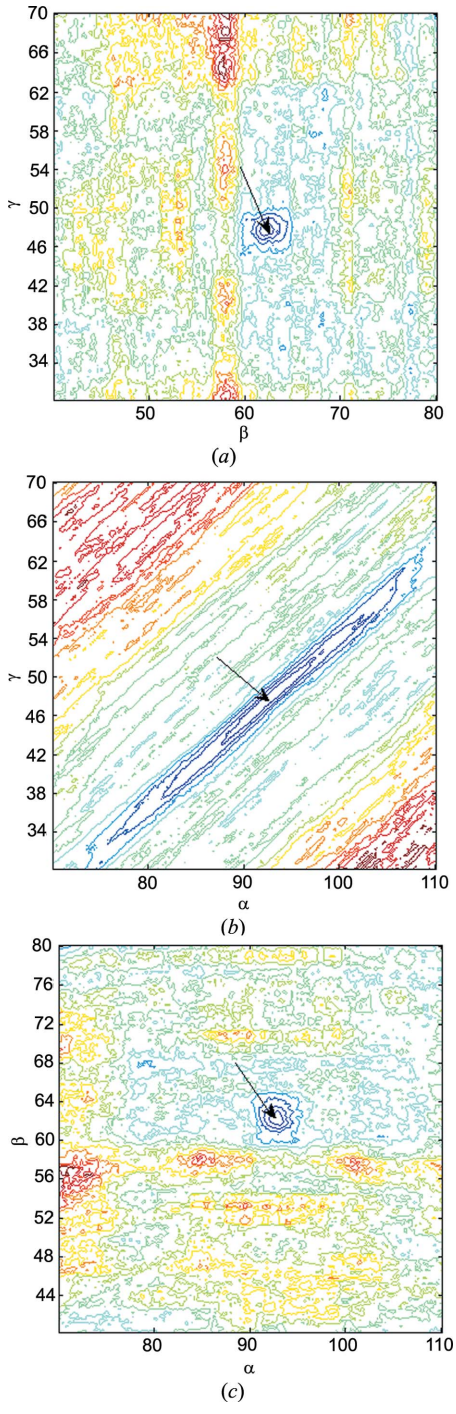

Figure 3
Contour maps of $R_{\rm factor}(\alpha,\beta,\gamma,21)$ between the 21st pattern and the reference diffraction pattern (first pattern) in the vicinity of the orientation (α = 92.6, β = 62.1, γ = 47.6°), which is the actual relative orientation of the 21st pattern. (*a*) Contour map along the $\beta\gamma$ plane, where the arrow points to the minimum in the orientation (α = 92.6, β = 62.2, γ = 47.4°). (*b*) Contour map along the $\alpha\gamma$ plane where the arrow points to the minimum in the orientation (α = 93.8, β = 62.1, γ = 48.6°). (*c*) The contour map along the $\alpha\beta$ plane where the arrow points to the minimum in the orientation (α = 92.8, β = 62.2, γ = 47.6°).

4. Determining the orientations using multiple-common-lines method

From the above simulation, we find that there are some diffraction patterns whose azimuth error is relatively large, such as the 11th, 13th, 14th, 20th and 21st patterns in Table 1. The main reason for the large errors is that the number of regular grids passed through by the four intersection arcs which are used to calculate $R_{\rm factor}(\alpha,\beta,\gamma,n)$ is small, as shown in Table 1. The number of regular grids passed through by the intersection arc between two Ewald spheres is related to their relative orientation. As shown in Fig. 4, when the relative orientation α belongs to (−180°, −150°] or [−30°, 30°] or [150°, 180°] and β belongs to [−30°, 30°], where γ is random, the number of regular grids passed through by the intersection arc is relatively large. Therefore, when the recovered orientation (α,β,γ) by the single-common-line method is located in this special range, it can be considered more accurate, such as the second, third, fourth, fifth and sixth patterns in Table 1.

In order to improve the accuracy of orientation determinations of other patterns, the diffraction patterns whose determined orientations are located in this special range and the original reference pattern can be used together to determine the orientations of the other patterns. Therefore, we can divide all the patterns into two subsets for convenience, the patterns whose orientations were determined by the single-common-line method as belonging to the special range and the original reference diffraction pattern are classified as subset 1 (for example, the first to sixth patterns in this paper), the other diffraction patterns are classified as subset 2 (for example, the

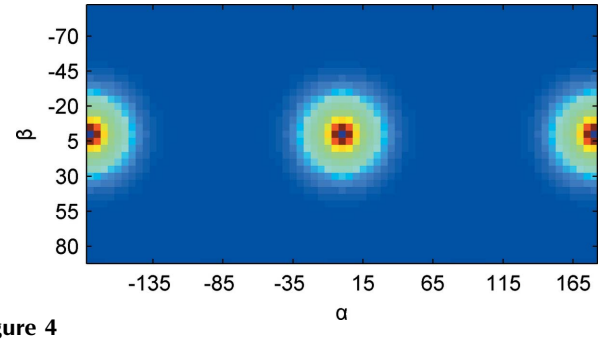

Figure 4
Number of regular grids passed through by the intersection arcs between two Ewald spheres along the $\alpha\beta$ plane, where γ is random, the range of β is −90–90°, the range of α is −180–180°.

Table 1
Results for determining the relative orientations of the 20 diffraction patterns using the single-common-line method.

Snapshot patterns	α (actual/recovered) (°)	β (actual/ recovered) (°)	γ (actual/recovered) (°)	Number of regular grids passed through by the four intersection lines
Second pattern	18.6/18.6	21.2/21.2	67.6 /67.6	2757
Third pattern	18.4/18.6	−34.8/−34.8	−32.2/−32.4	1775
Fourth pattern	−20.8/−20.8	−28.2/−28.4	122.8/122.8	2108
Fifth pattern	−31.7/−32.0	−27.7/−27.8	97.1/97.2	1546
Sixth pattern	8.8/8.6	12.4/12.2	−98.3/−98.6	3646
Seventh pattern	−116.6/−116.8	−34.4/−34.4	144.4/144.6	1446
Eighth pattern	−117.5/−118.2	−46.2/−46.0	−14.3/−13.8	1041
Ninth pattern	−153.5/−153.2	−69.3/−69.4	15.9/15.6	1004
Tenth pattern	−129.4/−129.6	73.9/73.8	−110.3/−110.4	958
11th pattern	62.2/61.2	62.6/62.2	123.4/122.4	1025
12th pattern	148.3/148.8	43.8/43.6	−113.5/−113.2	1274
13th pattern	18.7/17.8	73.3/72.8	67.7/67.0	922
14th pattern	−68.9/−71.2	−82.2/−82.4	−149.8/−147.6	881
15th pattern	24.1/24.2	44.4/44.4	82.5/82.6	1267
16th pattern	140.4/140.4	−54.9/−54.8	−89.3/−89.4	1075
17th pattern	56.9/57.0	9.6/9.6	−99.9/−99.8	1156
18th pattern	157.7/157.6	43.7/43.6	63.3/63.2	1323
19th pattern	59.9/59.6	9.8/9.8	148.5/148.4	1088
20th pattern	−106.4/−105.4	−76.6/−76.6	138.7/137.6	942
21st pattern	92.6/93.6	62.1/62.0	47.6/48.4	1077
Mean error	0.47	0.13	0.42	

seventh to 21st patterns in this paper). The detailed steps of the multiple-common-lines method are listed below.

(I) Refining the orientations of patterns in subset 1.

(1) The nth pattern whose orientation is to be refined is interpolated onto regular grids in reciprocal space with a random orientation (α,β,γ), $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ is used to represent this pattern.

(2) The reference patterns, which consist of all others patterns in subset 1, are also interpolated onto the regular grids in reciprocal space; a three-dimensional data set $\langle|F(h,k,l)|\rangle$ can be assembled from these patterns. Generally more than one pixel in the new reference patterns contributes to a given index of regular grid hkl, the pixel values are averaged to get a more accurate value at hkl. The initial orientations of the new reference patterns are given by the single-common-line method.

(3) Scanning the angles (α,β,γ) and calculating the average difference among regular grids passed through by intersection arcs between the $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ and three-dimensional data set $\langle|F(h,k,l)|\rangle$. Because $\langle|F(h,k,l)|\rangle$ is constituted from more than one pattern, there will be several intersection arcs between $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ and $\langle|F(h,k,l)|\rangle$. Therefore, compared with the single-common-line method, more diffraction intensity can be used to calculate the average difference, which will increase the accuracy of calculated average difference and recovered orientations. The parameter multi-$R_{\rm factor}(\alpha,\beta,\gamma,n)$ is used to represent the average difference; it can be expressed as

$$\text{multi-}R_{\rm factor}(\alpha,\beta,\gamma,n)=\sum_{hkl}\frac{\big||F_{\alpha,\beta,\gamma,n}(h,k,l)|-\langle|F(h,k,l)|\rangle\big|}{\langle|F(h,k,l)|\rangle}\Big/\sum_{hkl}1, \quad (4)$$

where the integer set hkl is the indices of regular grids passed through by all the intersection arcs between $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ and $\langle|F(h,k,l)|\rangle$. Then, using the searching algorithm in §2, a new recovered orientation of the nth pattern can be obtained and it will be used in next iteration.

(4) Repeat steps (1)–(3) for the rest of the patterns in subset 1, except the original reference pattern, until the orientations of all patterns reach convergence. In this paper, after 2–3 cycles the orientation of each pattern (high signal-to-noise) in subset 1 no longer changes.

(II) Redetermining the orientations of patterns in subset 2.

Using the searching algorithm in §2, the orientations of each pattern in subset 2 can be redetermined by scanning the angles (α,β,γ) and calculating average difference multi-$R_{\rm factor}(\alpha,\beta,\gamma,n)$ among regular grids passed through by all of the intersection arcs between the pattern and the reference patterns which consist of all the patterns in subset 1.

(III) Refining the orientations of all patterns.

The orientations of all the patterns can be redetermined by the same algorithm in step (I), all orientations will no longer change after a few cycles (2–3 cycles for high-signal-to-noise patterns in this paper); it means that the orientations of all patterns are self-consistent.

In step (III), if there are enough patterns, $\langle|F(h,k,l)|\rangle$ can be considered as an approximate three-dimensional structure factor. Using the intersection arcs between $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ and $\langle|F(h,k,l)|\rangle$, $|F_{\alpha,\beta,\gamma,n}(h,k,l)|$ can be located in an appropriate orientation to fit $\langle|F(h,k,l)|\rangle$, then $\langle|F(h,k,l)|\rangle$ can be updated. Repeating this for all patterns, after a few cycles, the best fitting orientations of patterns and the ultimate three-dimensional structure factor $\langle|F(h,k,l)|\rangle$ can be obtained. Since all of the common lines between one pattern and the other patterns are included in the calculation, the method is called the multiple-common-lines method. In this method, the

whole diffraction patterns will be exploited to determine the orientation of each of the patterns.

For the multiple-common-lines method, the most time-efficient computation is step (III). The time for these computations is proportional to $N_c(N_{p-1})S_{N_1,N_1}$, where N_p is the number of patterns, N_c is the number of cycles, S_{N_1,N_2} is the number of searches in the single-common-line method. Generally the larger the value of N_p, the more cycles (N_c) there will be. Therefore, in this paper, to save computational time for the 1000 patterns, we divided them into sets of about 20 patterns each; the orientations of patterns in each set can be determined by the multiple-common-lines method. When S_{N_1,N_2} is equal to 220 000, for a total of 1000 patterns, it will take about four days on a 50-node computing cluster (each node with a 3.40 GHz Inter(R) Core i7-3770 processor). However, it is still time consuming, and in order to reduce the amount of computation, a compromise solution can be established. After determination of the orientations of 20 or more patterns using the multiple-common-lines method, these patterns can be taken as the new reference patterns; then the orientations of the other patterns can be determined by the new reference patterns using the algorithm step (II). Using the compromise algorithm, it will take about one day on the same computing cluster.

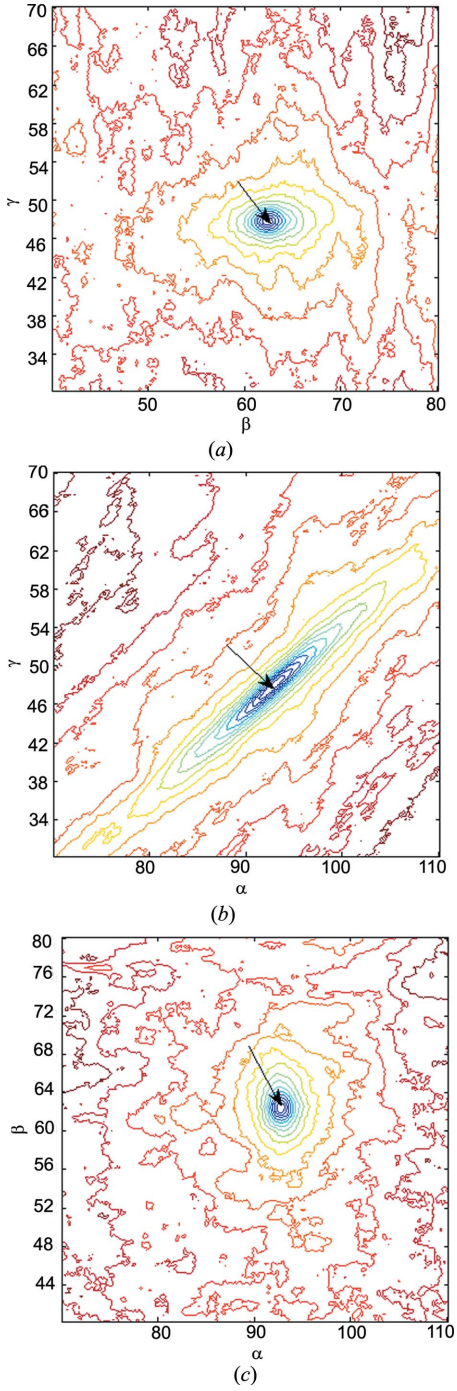

Figure 5
Contour maps of multi-$R_{factor}(\alpha, \beta, \gamma, 21)$ between the 21st pattern and the reference patterns which consist of the other 20 patterns. (*a*) The contour map along the $\beta\gamma$ plane, where the arrow points to the minimum in the orientation (α = 92.6, β = 62.2, γ = 47.4°). (*b*) The contour map along the $\alpha\gamma$ plane where the arrow points to the minimum in the orientation (α = 93.0, β = 62.1, γ = 47.8°). (*c*) The contour map along the $\alpha\beta$ plane where the arrow points to the minimum in the orientation (α = 92.8, β = 62.2, γ = 47.6°).

In order to compare with the single-common-line method, we employ the multiple-common-lines method to determine the orientations of the same patterns, second to 21st in Table 1. As shown in Fig. 5, there is a more obvious minimum multi-$R_{factor}(\alpha,\beta,\gamma,n)$ nearby the real orientation of the 21st pattern and the error in the $\alpha\gamma$ plane is also significantly reduced. As can be seen in Table 2, the azimuth errors of the 11th, 13th, 14th, 20th and 21st patterns are significantly reduced and for all 1000 patterns the mean errors of the orientation (α,β,γ) are about 0.14, 0.06 and 0.12°. Thus the accuracy of the orientation is significantly improved and sufficiently accurate. Using the compromise algorithm, the mean errors of the orientation (α,β,γ) are about 0.17, 0.05 and 0.09°.

The multiple-common-lines method can also be used to determine the orientations of low-signal patterns. A total of 81 patterns (including the original reference pattern) are simulated, whose parameters are the same as those in §3 except the pulse flux (10^{12} photons per 100 × 100 nm), the photon counts in outer pixels and outer Shannon–Nyquist pixels are about 0.015 and 0.1, respectively, where the solid angle of a Shannon–Nyquist pixel is $(\lambda/2O)^2$. These patterns are divided into two sets which consist of 40 patterns each. Using the single-common-line method and multiple-common-lines method (after 8–9 cycles), the orientations of patterns in each set can be recovered (all orientations no longer change). In the case of low-signal patterns, the determined orientations of patterns in subset 1 are not accurate enough by the single-common-line method. Therefore, after obtaining the initial orientations by the single-common-line method, step (III) in the multiple-common-lines method is directly used to recover the orientations in each set.

The results of the single-common-line method and multiple-common-lines method are shown in Fig. 6 and Fig. 7, respectively. As shown in Fig. 6, most of the determined orientations seriously deviate from the actual orientations. For a total of 80 patterns, the mean errors of the orientation (α,β,γ) are about 7.17, 2.86 and 5.06°. Therefore, the single-common-line method cannot work well for low-signal patterns. As shown in

Table 2
Results for determining the orientations of the 20 diffraction patterns using the multiple-common-lines method.

Snapshot patterns	α (actual/recovered) (°)	β (actual/recovered) (°)	γ (actual/ recovered) (°)	Number of grids passed through by the multiple intersection lines
Second pattern	18.6/18.4	21.2/21.2	67.6/67.4	26638
Third pattern	18.4/18.4	−34.8/−34.8	−32.2/−32.4	22909
Fourth pattern	−20.8/−20.8	−28.2/−28.4	122.8/122.6	25661
Fifth pattern	−31.7/−31.8	−27.7/−27.8	97.1/97.0	26606
Sixth pattern	8.8/8.6	12.4/12.2	−98.3/−98.4	24708
Seventh pattern	−116.6/−116.8	−34.4/−34.4	144.4/144.6	28413
Eighth pattern	−117.5/−117.6	−46.2/−46.2	−14.3/−14.2	29571
Ninth pattern	−153.5/−153.8	−69.3/−69.4	15.9/16.2	33500
Tenth pattern	−129.4/−129.4	73.9/73.8	−110.3/−110.2	25281
11th pattern	62.2/62.0	62.6/62.6	123.4/123.2	31404
12th pattern	148.3/148.2	43.8/44.0	−113.5/−113.4	27432
13th pattern	18.7/18.4	73.3/73.2	67.7/67.4	33374
14th pattern	−68.9/−68.4	−82.2/−82.4	−149.8/−150.4	29257
15th pattern	24.1/24.0	44.4/44.4	82.5/82.4	28398
16th pattern	140.4/140.2	−54.9/−54.8	−89.3/−89.2	23723
17th pattern	56.9/56.8	9.6/9.6	−99.9/−100.0	29188
18th pattern	157.7/157.6	43.7/43.8	63.3/63.4	27171
19th pattern	59.9/59.8	9.8/9.8	148.5/148.4	28530
20th pattern	−106.4/−106.8	−76.6/−76.8	138.7/139.0	32040
21st pattern	92.6/92.6	62.1/62.2	47.6/47.6	30826
Mean error	0.16	0.09	0.18	

Fig. 7, after the multiple-common-lines method, the misorientation angle is significantly reduced and the mean errors of the orientation (α,β,γ) are about 0.77, 0.31 and 0.60°, which basically meet the requirements of angular resolution. The accuracy of recovered orientations can be further improved by containing more patterns in each set.

5. Interpolating diffraction patterns onto the regular grids in reciprocal space and reconstructing the three-dimensional structure

After recovering the orientations, these patterns can be easily interpolated onto the regular grids in reciprocal space by nearest-neighbor interpolation. The Ewald spheres are embedded into a three-dimensional array of 128 × 128 × 128 voxels, which means that the oversampling degree is two in each dimension in reciprocal space.

After interpolating these 1000 patterns onto the grids in reciprocal space, we can reconstruct the three-dimensional structure by a phase-retrieval algorithm. Currently, phase-retrieval algorithms are well developed, such as the error reduction (ER) algorithm, the hybrid input–output (HIO) algorithm and the oversampling smoothness algorithm (Fienup, 1982; Rodriguez *et al.*, 2013). In this paper, the HIO and the ER algorithms are used to phase the three-dimensional diffraction data, first 1000 iterations of HIO and then 500 iterations of ER are performed. The final $R_{\rm factor}$ by HIO–ER is about 0.14. The definition of the $R_{\rm factor}$ is the same as with X-ray crystallography,

$$R_{\rm factor} = \sum_{hkl} |F_{\rm m}(h,k,l) - F_{\rm r}(h,k,l)| \Big/ \sum_{hkl} F_{\rm m}(h,k,l), \qquad (5)$$

where $F_{\rm m}(h,k,l)$ is the measured amplitude of Fourier transform, $F_{\rm r}(h,k,l)$ is the reconstructed amplitude of Fourier transform. The real structure and the reconstructed structure by HIO–ER are shown in Figs. 8 and 9.

Comparing Figs. 8 and 9, we find that the reconstructed structure is basically consistent with the real structure. But

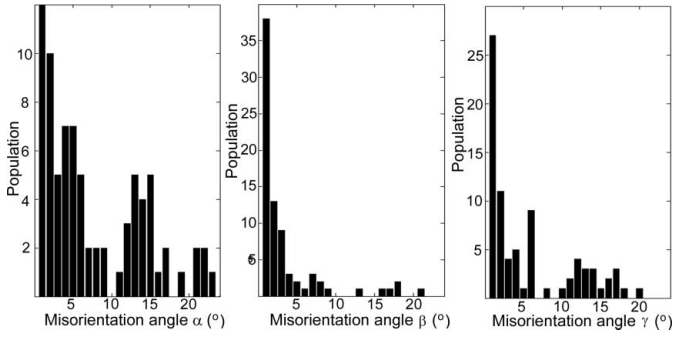

Figure 6
Distribution of misorientation angle of low-signal patterns determined by single-common-line method.

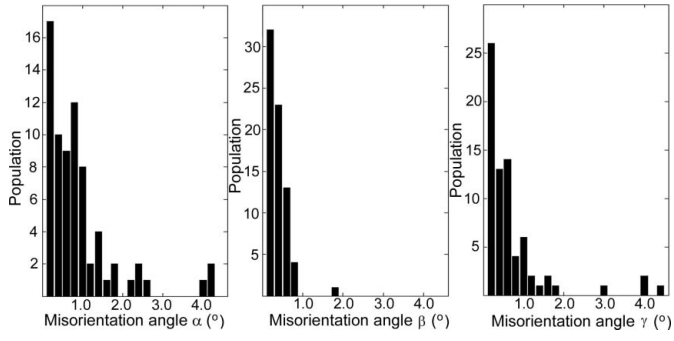

Figure 7
Distribution of misorientation angle of low-signal patterns determined by multiple-common-lines method.

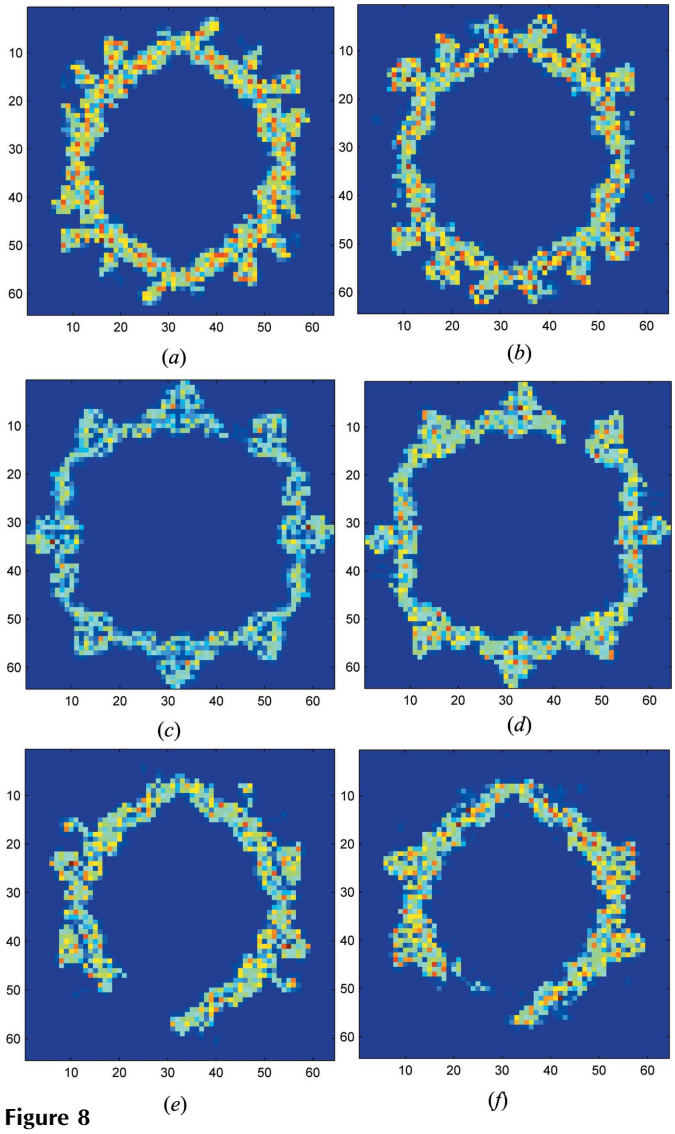

Figure 8
Part of the real structure along the *XY* plane. (*a*) $Z = 21$, (*b*) $Z = 22$, (*c*) $Z = 33$, (*d*) $Z = 34$, (*e*) $Z = 45$, (*f*) $Z = 46$. Pixel size in the image equals $O/64$, about 0.54 nm.

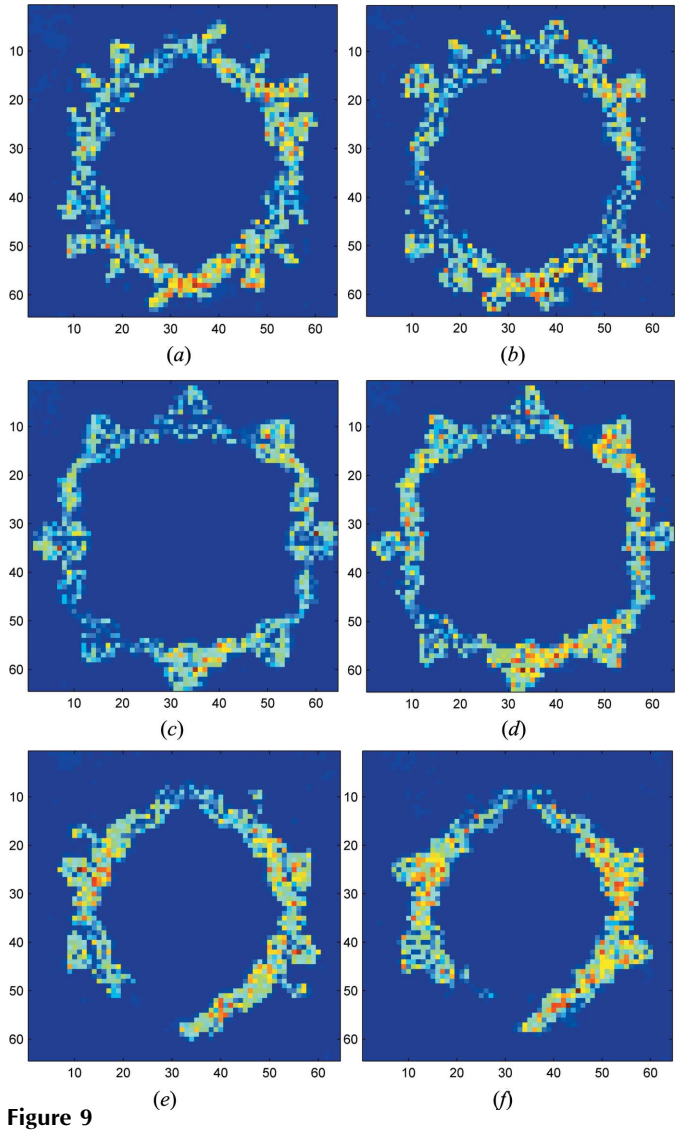

Figure 9
Part of the reconstructed structure along the *XY* plane. (*a*) $Z = 21$, (*b*) $Z = 22$, (*c*) $Z = 33$, (*d*) $Z = 34$, (*e*) $Z = 45$, (*f*) $Z = 46$. Pixel size in reconstructed image equals $O/64$, about 0.54 nm.

there is some noise in the reconstructed structure and this problem is mainly caused by interpolation error, azimuth error and Poisson noise; it can be solved by increasing the number of the diffraction patterns.

6. Conclusion

It has been demonstrated that the multiple-common-lines method can accurately recover the orientations of high- and low-signal diffraction patterns. In this method, the whole of a diffraction pattern is exploited to determine the orientation of each pattern, which can significantly reduce the azimuth error; this mean accuracy can meet the requirements of angular resolution. In essence, the multiple-common-lines method combines the traditional common-line method and an iterative method, and needs no additional classification process, even for low-signal patterns. The calculation time of multiple-common-lines method is proportional to the products of the number of patterns, the number of iterations and the number of searches in orientation space which are easy to implement in parallel computing. After the recovery of orientations, the three-dimensional structure factor $\langle |F(h,k,l)| \rangle$ can be easily obtained by interpolating (nearest-neighbor interpolation or other interpolation methods) these patterns onto regular grids in reciprocal space, the virus structure can be finally reconstructed by an iterative phasing algorithm. With the development of XFELs, it is hoped that this method can greatly facilitate the three-dimensional structure determination of noncrystalline materials and nanocrystals.

References

Bortel, G. & Faigel, G. (2007). *J. Struct. Biol.* **158**, 10–18.
Bortel, G., Faigel, G. & Tegze, M. (2009). *J. Struct. Biol.* **166**, 226–233.
Bortel, G. & Tegze, M. (2011). *Acta Cryst.* A**67**, 533–543.
Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.

Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. (2009). *Nat. Phys.* **5**, 64–67.
Hopper, P., Harrison, S. C. & Sauer, R. T. (1984). *J. Mol. Biol.* **177**, 701–713.
Huldt, G., Szoke, A. & Hajdu, J. (2003). *J. Struct. Biol.* **144**, 219–227.
Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., Spence, J. C. H., Hunter, M., Fromme, P., White, T., Chapman, H. N. & Holton, J. (2010). *Opt. Express*, **18**, 5713–5723.
Loh, N.-T. D. *et al.* (2010). *Phys. Rev. Lett.* **104**, 225501.
Loh, N.-T. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.
Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.
Moths, B. & Ourmazd, A. (2011). *Acta Cryst.* A**67**, 481–486.
Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
Raines, K. S., Salha, S., Sandberg, R. L., Jiang, H., Rodríguez, J. A., Fahimian, B. P., Kapteyn, H. C., Du, J. & Miao, J. (2010). *Nature*, **463**, 214–217.
Rodriguez, J. A., Xu, R., Chen, C.-C., Zou, Y. & Miao, J. (2013). *J. Appl. Cryst.* **46**, 312–318.
Sayre, D., Chapman, H. N. & Miao, J. (1998). *Acta Cryst.* A**54**, 232–239.
Schwander, P., Fung, R., Phillips, G. N. & Ourmazd, A. (2010). *New J. Phys.* **12**, 035007.
Seibert, M. M. *et al.* (2011). *Nature*, **470**, 78–81.
Shneerson, V. L., Ourmazd, A. & Saldin, D. K. (2008). *Acta Cryst.* A**64**, 303–315.
Sidi, M. J. (1997). *Spacecraft Dynamics and Control.* Cambridge University Press.
Song, C., Jiang, H., Mancuso, A., Amirbekian, B., Peng, L., Sun, R., Shah, S. S., Zhou, Z. H., Ishikawa, T. & Miao, J. (2008). *Phys. Rev. Lett.* **101**, 158101.
Tegze, M. & Bortel, G. (2012). *J. Struct. Biol.* **179**, 41–45.
Tegze, M. & Bortel, G. (2013). *J. Struct. Biol.* **183**, 389–393.
Tokuhisa, A., Taka, J., Kono, H. & Go, N. (2012). *Acta Cryst.* A**68**, 366–381.
Yefanov, O. M. & Vartanyants, I. A. (2013). *J. Phys. B At. Mol. Opt. Phys.* **46**, 164013.
Zhou, L., Liu, P. & Dong, Y.-H. (2013). *Chin. Phys. C*, **37**, 028101.