Three-dimensional reconstruction for coherent diffraction patterns obtained by XFEL

Miki Nakano,[a] Osamu Miyashita,[a] Slavica Jonic,[b] Changyong Song,[c] Daewoong Nam,[c] Yasumasa Joti[d] and Florence Tama[a,e,f]*

[a]Advanced Institute of Computational Science, RIKEN, 6-7-1 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan, [b]IMPMC, Sorbonne Universités – CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206, Paris 75005, France, [c]Department of Physics, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Republic of Korea, [d]XFEL Utilization Division, Japan Synchrotron Radiation Research Institute (JASRI), 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5198, Japan, [e]Department of Physics, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan, and [f]Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan. *Correspondence e-mail: florence.tama@riken.jp

The three-dimensional (3D) structural analysis of single particles using an X-ray free-electron laser (XFEL) is a new structural biology technique that enables observations of molecules that are difficult to crystallize, such as flexible biomolecular complexes and living tissue in the state close to physiological conditions. In order to restore the 3D structure from the diffraction patterns obtained by the XFEL, computational algorithms are necessary as the orientation of the incident beam with respect to the sample needs to be estimated. A program package for XFEL single-particle analysis based on the *Xmipp* software package, that is commonly used for image processing in 3D cryo-electron microscopy, has been developed. The reconstruction program has been tested using diffraction patterns of an aerosol nanoparticle obtained by tomographic coherent X-ray diffraction microscopy.

1. Introduction

Biological molecules, such as proteins, nucleic acids and their complexes, are responsible for many vital cellular functions, including gene transcription and protein synthesis, and their dysfunctions result in severe diseases (Watson & Crick, 1953; Aloy & Russell, 2002; Berman *et al.*, 2003). Information on the structure and dynamics of biomolecules and biomolecular complexes is helpful for understanding their functional mechanisms, and X-ray crystallography is currently the most widely used technique to determine the three-dimensional (3D) structure of biomolecules (Drenth, 2007; Rupp, 2009). However, this technique requires molecules to be crystallized, and it is difficult to apply to insoluble molecules or intrinsically disordered proteins. Although cryo-electron microscopy (cryo-EM) does not require crystallization and can observe heterogeneous samples (Kühlbrandt, 2014), applications to observe the inner structure of thick objects (more than 500 nm) can be challenging due to multiple-scattering events (Cheng *et al.*, 2015).

Single-particle imaging using femtosecond X-ray pulses from X-ray free-electron lasers (XFELs) could allow the inner structure of biological molecules to be observed in a state close to nature without crystallization (Neutze *et al.*, 2000; Huldt *et al.*, 2003; Chapman *et al.*, 2006; Gaffney & Chapman, 2007). Radiation damage is a serious problem in high-reso-

experimental diffraction patterns (angle unknown)
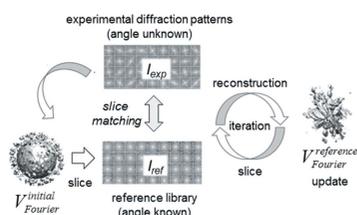
lution microscopy as it reduces the spatial resolution of the molecular structure (Henderson, 1995; Kirz *et al.*, 1995; Cheng *et al.*, 2015). XFELs can significantly relax the resolution barrier imposed by radiation damage by recording the diffraction pattern before specimen destruction, due to femtosecond-short pulse duration (Neutze *et al.*, 2000; Chapman *et al.*, 2011; Hirata *et al.*, 2014; Suga *et al.*, 2014). XFEL experimental data are becoming increasingly available and several low-resolution structures from single-particle approaches have been reported (Seibert *et al.*, 2011; Galla-gher-Jones *et al.*, 2014; Kimura *et al.*, 2014; Xu *et al.*, 2014; Ekeberg *et al.*, 2015; Takayama *et al.*, 2015; van der Schot *et al.*, 2015). It has also been shown, theoretically, that high-resolu-tion 3D structures could be obtained using millions of diffraction patterns (Loh & Elser, 2009; Tegze & Bortel, 2012; Tokuhisa *et al.*, 2012; Hosseinizadeh *et al.*, 2014) and that dynamics could be directly extracted from two-dimensional (2D) data (Tokuhisa *et al.*, 2016).

However, there are still challenging problems in obtaining high-resolution 3D structures of biomolecules from XFEL experimental data. Because the diffraction intensity from biological molecules is too weak, an insufficient photon count is a serious problem, especially at high-wavenumber pixels which determine the resolution in real space. On the other hand, the diffraction intensity at low-wavenumber pixels is too strong and saturates the detection range, which hinders the determination of the overall shape of the system by phase-retrieval procedures. In addition, 3D imaging requires the assembly of diffraction patterns from many identical copies of a reproducible object. Therefore, 2D diffraction images should be obtained from structurally homogeneous samples, but it is generally difficult for sub-micrometer systems, which is the typical target size for XFEL single-particle analysis.

Along with these challenges, one of the critical algorithmic problems to be solved in order to reconstruct 3D structures from diffraction patterns obtained in XFEL experiments is the estimation of the orientation of single particles with respect to the laser beam (three Euler angles) (Loh & Elser, 2009; Philipp *et al.*, 2012; Ekeberg *et al.*, 2015). This is also a common problem in cryo-EM single-particle analysis. Several open-source software packages have been developed to reconstruct 3D molecular structures from cryo-EM projection images (Grigorieff, 2007; Tang *et al.*, 2007; Shaikh *et al.*, 2008; Scheres, 2012; de la Rosa-Trevín *et al.*, 2013). There are currently two main strategies for estimating the 3D orientation of particles. One is called projection matching (Penczek *et al.*, 1994; Sorzano *et al.*, 2004; Grigorieff, 2007; Yang & Penczek, 2008), where the best angular parameters are estimated by finding an image which has a maximum correlation coefficient in the reference library and is used to construct a new volume. Another widely used strategy is the maximum-likelihood-based algorithm (Scheres *et al.*, 2007; Sigworth *et al.*, 2010; Scheres, 2012; Lyumkis *et al.*, 2013). In these approaches, a number of angular assignments are considered for each target image, and are concurrently used for reconstruction with relative weights based on the similarities between the target image and reference images.

While a cryo-EM electron density image contains structural information in real space, an XFEL diffraction pattern contains structural information in Fourier space and is related to a slice (Ewald sphere) of the 3D diffraction intensity. Thus, 3D reconstruction from XFEL data can be performed using similar procedures to those used in cryo-EM. For instance, 'slice matching' can be performed to determine the orienta-tion of the diffraction patterns. One of the algorithms successfully applied for 3D reconstruction from XFEL data is EMC (Ekeberg *et al.*, 2015), which uses a 'maximum like-lihood' approach. A 'maximum cross correlation' algorithm for 3D reconstruction was previously tested with synthetic data using a large number (8000–100000) of diffraction patterns (Tegze & Bortel, 2012). Here, we aim to reconstruct a 3D model employing a 'maximum cross correlation' algorithm with experimental data and to examine necessary parameter calibrations in detail. Our maximum cross-correlation algo-rithm is similar to that of Tegze & Bortel, but there are also some differences between the two algorithms as follows: we reconstructed the 3D structure factor amplitude instead of the 3D diffraction intensity by using a weight function based on the Kaiser–Bessel window. A phase recovery procedure is needed to obtain a 3D model in real space from the assembled 3D model in Fourier space (assembled from the diffraction patterns at the determined orientations). However, this task is also not trivial and its result may depend on the choice of the phase recovery algorithm (Sekiguchi *et al.*, 2016); therefore we here focus on the angular assignment process only.

Our program was implemented on top of *Xmipp*, which is an image-processing software package primarily aimed at single-particle 3D cryo-EM (de la Rosa-Trevín *et al.*, 2013). The program suite contains many useful tools for image analysis that could be used for analyzing XFEL data. We tested our reconstruction program using experimental diffraction data of an aerosol nanoparticle obtained by tomographic coherent X-ray diffraction microscopy (CXDM) (Miao *et al.*, 2006; Barty *et al.*, 2008; Jiang *et al.*, 2010) instead of data from XFEL. These data are very similar to those from XFEL experiments, the main difference being that XFEL is 'single-shot' while X-ray diffraction microscopy allows (weaker but) repeating exposure. The speckle patterns from CXDM were employed to simulate XFEL single-particle data, *i.e.* a set of diffraction patterns obtained from a non-crystalline nanoparticle with different sample orientations. Sample orientations were estimated using the approach demonstrated here, and the assigned angles were later compared with the angles actually used for the data collection. We discuss the choice of parameters and protocols for a successful estimation of the incident beam angles.

2. Theory: reconstruction of 3D structure in Fourier space from diffraction patterns

We performed the reconstruction of the structure factor amplitude distribution (hereafter called 'volume'), from diffraction patterns, based on the 'slice matching' protocol
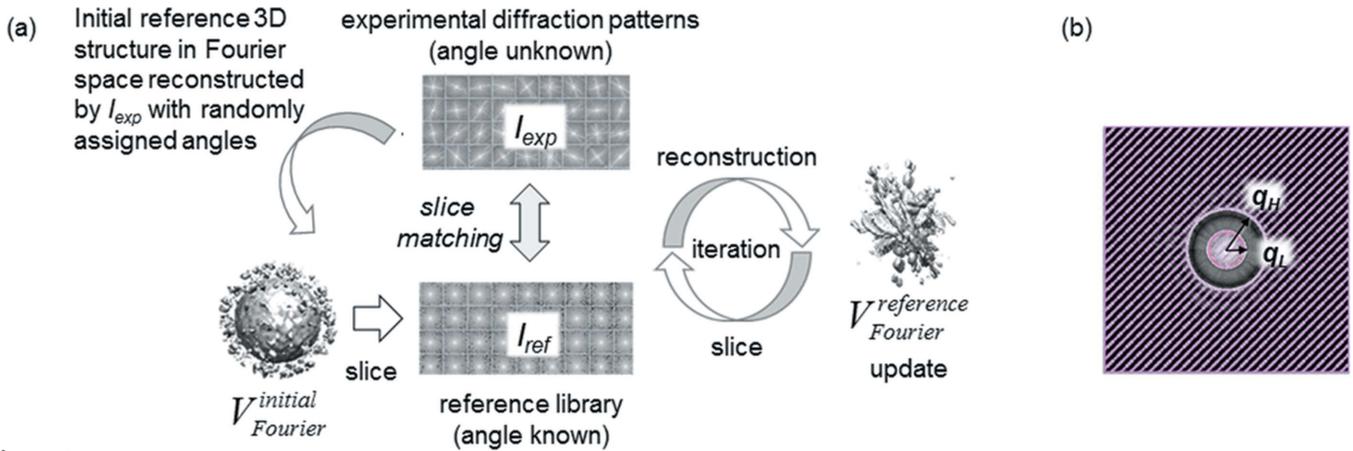
Figure 1
(a) Schematic view of our program for volume reconstruction from experimental diffraction patterns. (b) Matching region in the diffraction pattern used for calculation of cross correlation. Center and outer regions of the diffraction pattern are masked (filled with pink stripes).

(Fig. 1a). While others reconstruct 3D diffraction patterns by assembling 2D diffraction patterns in 3D space (Tegze & Bortel, 2012; Ekeberg *et al.*, 2015), we here reconstruct 3D amplitudes, because the 3D amplitude can be directly used for phase retrieval. More precisely, we convert 2D diffraction patterns to structure factor amplitude and then assemble into 3D space (comparison results between 3D diffraction amplitude distribution and 3D diffraction intensity are shown in §4). The protocol is as follows:

(i) Create an initial volume, $V_{\text{Fourier}}^{\text{initial}}$. In general, we create $V_{\text{Fourier}}^{\text{initial}}$ using a set of 2D diffraction patterns with incident beam angles that are randomly assigned. Each 2D diffraction pattern has been centro-symmetrized, and the square roots of the intensities are calculated to obtain structure factor amplitude before calculation. In the first iteration, $V_{\text{Fourier}}^{\text{initial}}$ is used as the reference volume, $V_{\text{Fourier}}^{\text{reference}}$.

(ii) Create a reference library of N_{ref} diffraction patterns by extracting, using cubic spline interpolation, slices from $V_{\text{Fourier}}^{\text{reference}}$ at orientations (angles φ and θ) distributed over a sphere evenly, using a given angular sampling (discretization) step (Bunge & Baumgardner, 1995). Then, square slice pixel values to obtain the corresponding diffraction patterns.

(iii) Calculate the zero-mean normalized cross correlation, CC, between each experimental diffraction pattern and all reference patterns rotated by angles ψ in plane (with a given angular discretization step) using the following equation,

$$ \text{CC}_{pq} = \frac{(1/N_{\text{pix}}) \sum\limits_{i}^{N_{\text{pix}}} \left\{ \left[I_{\text{exp},p}(i) - \bar{I}_{\text{exp},p} \right] \left[I_{\text{ref},q}^{\psi}(i) - \bar{I}_{\text{ref},q} \right] \right\}}{\sigma_{I_{\text{exp},p}} \sigma_{I_{\text{ref},q}}}, \quad (1) $$

where $I_{\text{exp},p}(i)$ and $I_{\text{ref},q}(i)$ are the diffraction intensity at pixel i of the pth experimental and qth reference diffraction patterns, respectively ($p = 1$ to N_{exp}, $q = 1$ to N_{ref}). N_{pix} is the number of pixels in each diffraction pattern. $\bar{I}_{\text{exp},p}$ and $\bar{I}_{\text{ref},q}$ are the average intensities of the pth experimental and qth reference diffraction patterns, and $\sigma_{I_{\text{exp},p}}$ and $\sigma_{I_{\text{ref},q}}$ are their standard deviations, respectively. $I_{\text{ref},q}^{\psi}(i)$ is the diffraction intensity of the qth reference pattern rotated by angle ψ in the plane. The

reference pattern and ψ resulting in the maximum CC coefficient (CC^{max}) are denoted by $M_{\text{ref}}^{\text{opt}}$ and ψ^{opt}, respectively. The incident beam angles φ^{opt} and θ^{opt} used to create $M_{\text{ref}}^{\text{opt}}$ and the in-plane angle ψ^{opt} assigned to the experimental image are the Euler angles set which determines the particle orientation. For the calculation, we apply masks as described below to enhance the sensitivity of the angular assignment.

(iv) Reconstruct a volume, V_{Fourier}, using the experimental images with the Euler angles assigned as described above. The diffraction amplitude at voxel k in the reconstructed volume, $A(k)$, is obtained as the weighted average of the square roots of the diffraction intensities, $[I_{\text{exp},p}(i)]^{1/2}$, in the experimental diffraction images [equation (2)]. To calculate $A(k)$, a weight function based on the Kaiser–Bessel window is used (Lewitt, 1990; Abrishami *et al.*, 2015), which depends on the distance d_{kj} between the position k and j within V_{Fourier}: k is the center position of the voxel k and j is the position of the pixel i within the 2D diffraction image, p, in a 3D volume [equation (3)],

$$ A(k) = \sum_{p=1}^{N_{\text{exp}}} \sum_{i=1}^{N_{\text{pix}}} w(d_{kj}) \left[I_{\text{exp},p}(i) \right]^{1/2} \Big/ \sum_{p=1}^{N_{\text{exp}}} \sum_{i=1}^{N_{\text{pix}}} w(d_{kj}), \quad (2) $$

$$ w(d_{kj}) = \frac{\xi(\alpha, \eta)}{I_0(\pi\alpha)} I_0 \left\{ \pi\alpha \left[1 - \left(\frac{2d_{kj} - \eta}{\eta - 1} - 1 \right)^2 \right]^{1/2} \right\}, \quad (3) $$

$$ 0 \le d_{kj} \le \eta. $$

I_0 is the zeroth-order modified Bessel function, η is the maximum radius for interpolation, and α is a variable which determines the decreasing rate of $w(d_{kj})$. $\xi(\alpha,\eta)$ is the normalization factor determined by α and η. With a large η, the diffraction intensity would be interpolated using the pixels farther in the mapped position in the reconstructed volume. With a large α, the weight for the interpolation would be decreased quickly as d_{kj} increases. These parameters need careful calibration for successful reconstructions and are discussed later in detail. Finally, we centro-symmetrized the reconstructed volume in Fourier space.

(v) Update $V_{\text{Fourier}}^{\text{reference}}$ using the volume reconstructed in the previous step. The reference library sampling steps will be made smaller for the refinement of angle assignment in the next iteration. The method allows the correlation coefficient between the experimental diffraction image and the reference diffraction images to be calculated in the vicinity of the currently assigned angles. This is an option of the method that should not be used in the beginning of the iterative assignment process.

(vi) Iterate from (ii) to (v) until convergence of $V_{\text{Fourier}}^{\text{reference}}$. The final 3D structure is denoted as $V_{\text{Fourier}}^{\text{final}}$.

In the calculation of CC, we excluded the center and outer regions of the diffraction patterns to improve the sensitivity of the matching. The center region is excluded in the calculation since these very large intensity values reduce the sensitivity of the CC calculation. Also, in practice, these diffraction intensities are often too strong to be measured and are thus missing in the data, especially in XFEL applications. In the outer region of diffraction patterns, intensities are usually too weak to be detected and are uncertain because of noise. Therefore, we only calculated CC for the annular regions defined by the inner and outer radii, q_L and q_H, as shown in Fig. 1(b).

The 'slice matching' approach described above [steps (i)–(vi)] was implemented based on a projection matching protocol included in *Xmipp*, which uses a Fourier-space representation of the reference volume for library creation as well as 3D interpolation in Fourier space for 3D reconstruction.

3. Results

3.1. Pre-processing of experimental diffraction patterns

To test our 'slice matching' protocol, we used tomographic CXDM data of aerosol nanoparticles. The similarity between CXDM and single-particle XFEL resides in the absence of sample crystallization and in the sample irradiation from different incident beam orientations. However, different diffraction patterns collected in a CXDM experiment correspond to different orientations of the same sample while each single-particle XFEL diffraction pattern corresponds to an orientation of a different sample. Additionally, in tomographic CXDM, the sample is only rotated around one axis (one angle, θ, is known) with respect to the incident beam, whereas all three Euler angles (φ, θ and ψ) are unknown in single-particle analysis using XFELs.

Our final goal is to reconstruct the 3D structure using XFEL experimental diffraction images. The aim of this study is to validate the incident beam angle estimation in our reconstruction program using experimental diffraction images including noise and obtain insight into the calibration of the necessary parameters. Diffraction images obtained by tomographic CXDM are ideal data for our purpose, since incident beam angles with respect to the sample are known (the tilt angles were set experimentally). Pretending that we do not know the incident beam angles, we used our program to estimate the orientation of each diffraction pattern, and
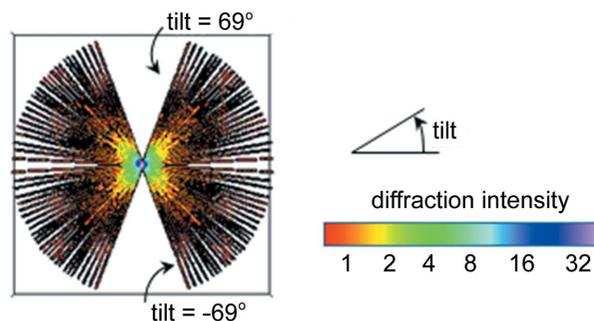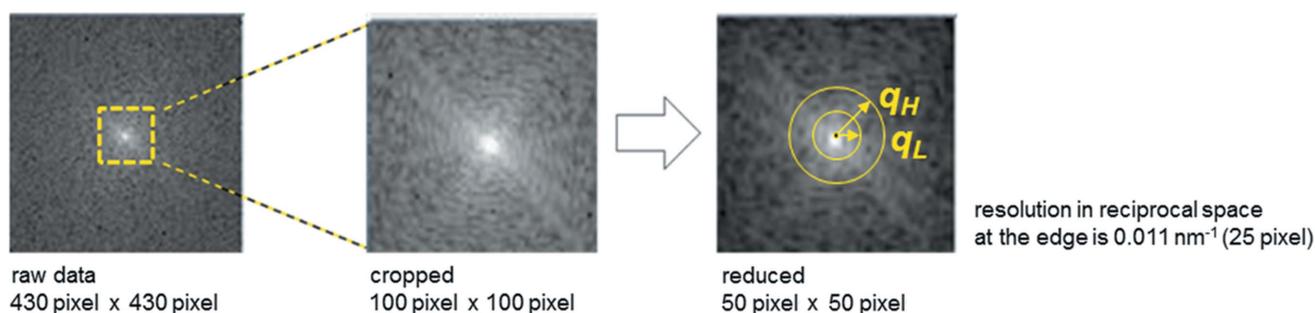

Figure 2
Cross section of the volume $V_{\text{Fourier}}^{\text{true}}$ that was assembled from experimental diffraction patterns arranged in 3D Fourier space using the ground-truth angles. The tilt axis is perpendicular to the cross section.

compared against the actual orientation. A total of 53 diffraction patterns at various sample orientations (tilting angles from $-69°$ to $+69°$ in $1°$ to $4°$ increments in an equal slope scheme) were measured using a CCD camera. The specimen size used in this study was about 1.5 μm. The data were all 3×3 binned and resized to 430 pixel $\times$ 430 pixel. All diffraction patterns were centro-symmetrized.

Fig. 2 shows a cross-section view of the arrangement of the full-size experimental diffraction patterns in 3D Fourier space using the ground-truth incident beam angles, $V_{\text{Fourier}}^{\text{true}}$, with the interpolation parameters $\eta = 1$ pixel and $\alpha = 15$. The tilt axis is perpendicular to the cross section shown in Fig. 2. The empty regions in this 3D space are related to the limitations in the maximum tilt angle achievable with the given CXDM apparatus, which is identical to the missing wedge problem in cryo-electron tomography (Lučić *et al.*, 2005). Also, there is one common line shared by all experimental diffraction patterns.

Before applying the 3D reconstruction algorithm, we cropped the outer region of the diffraction patterns to remove the pixels which do not practically carry diffraction information (the values of pixels much farther from the central pixel are usually zero or too small to be distinguished from noise). The resolution in reciprocal space at the edge of the cropped image is approximately 0.011 nm^{-1} ($= q_{\text{max}}$), which corresponds to 91 nm resolution in real space. Furthermore, we reduced the image size from 100 pixel $\times$ 100 pixel to 50 pixel $\times$ 50 pixel by binning (Fig. 3). These reductions were necessary in order to cover a sufficient number of voxels to ensure that nearby 2D images in the assembled 3D volume are close enough to detect the consistency (and inconsistency) of the 3D reconstruction. Cropping the outer region of a diffraction pattern corresponds to reducing the resolution in real space, while diffraction pattern binning (reducing the oversampling in Fourier space) corresponds to denoising in Fourier space (without binning, the diffraction intensity fluctuates more due to noise, including Poisson noise). The curvature of the Ewald sphere can be ignored, taking into account the current resolution limit. Regarding searching orientations of diffraction patterns by 2D slice matching of a 3D structure in Fourier space, diffraction pattern cropping is important because the search is focused on the central region that contains information about the object's shape in real space.

Figure 3
Pre-processing of the experimental diffraction patterns.

3.2. Adjustments of beam intensity variations

To reconstruct 3D structures, variations of beam intensity embedded in the diffraction patterns need to be normalized. Fig. 4(a) shows the diffraction intensities averaged over the cropped region reduced to 50 pixel × 50 pixel. In Fig. 4(a), the average diffraction intensity over all ground-truth tilt angles was 32.59 ± 4.15. It is expected that the average diffraction intensity is not the same for different ground-truth tilt angles and that a continuous change is possible. In this study, we normalized (scaled) the cropped diffraction patterns (50 pixel × 50 pixel) so that the average intensity is the same for all ground-truth tilt angles [Figs. 4(a) and 4(d)]. Note that, in all scaled diffraction patterns, the sum of diffraction intensities at low wavenumbers ($q_L < 5$ pixel) is 95% of the total sum. This scaling smoothens the variations of the intensity averaged

over the annular regions defined by q_L and q_H, as shown in Figs. 4(b), 4(c), 4(e) and 4(f).

3.3. Adjustment of the interpolation parameters

To find optimal interpolation parameters for 3D reconstruction, we reconstructed volumes using experimental images (pre-processed as described in the previous sections) and their ground-truth orientations (ground-truth tilt angles), with various interpolation parameters for 3D reconstruction [Figs. 5(a)–5(e)]. The interpolation radius, η, needs to be sufficiently large to fill the reciprocal space where experimental data are missing. The parameter α controls the relative weight for the interpolation; with a smaller α, data points are more equally counted, regardless of the distance, and the reconstructed volume becomes blurred. The maximum
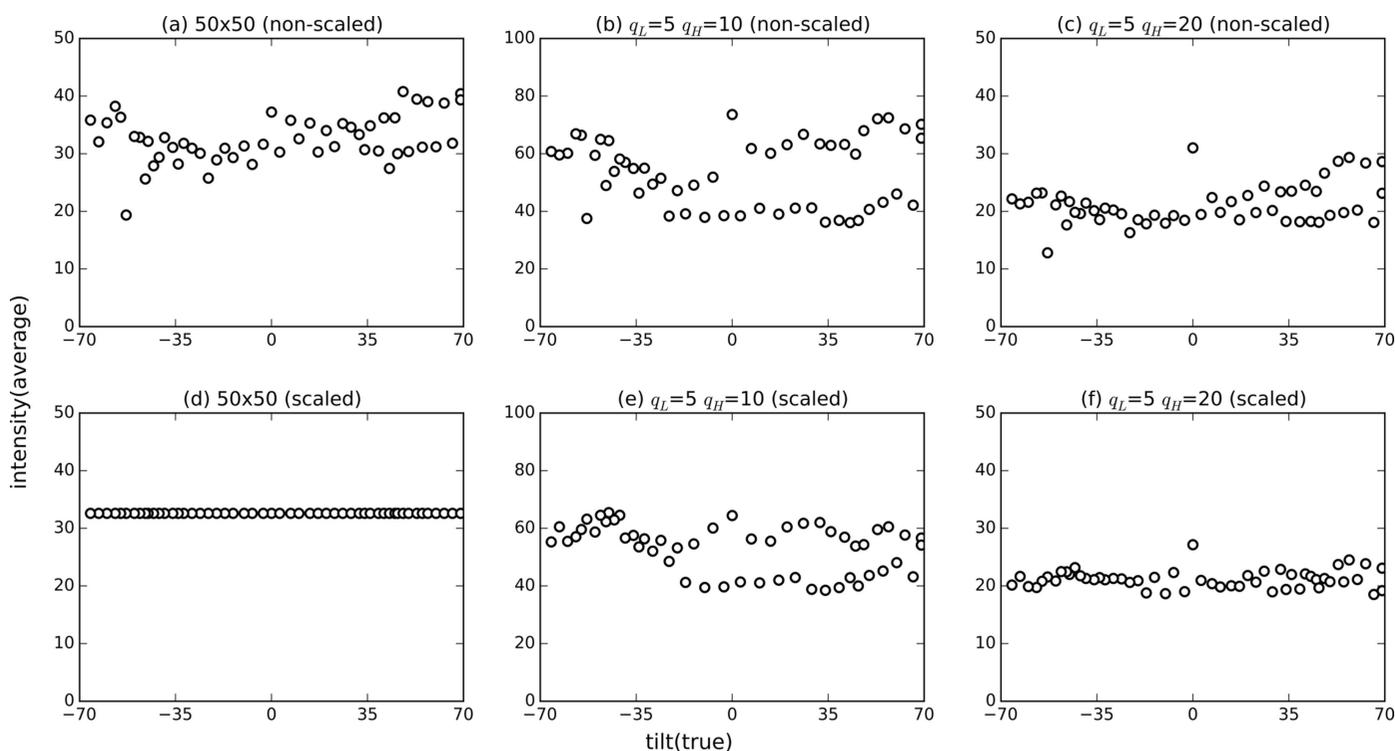
Figure 4
Intensity of each diffraction pattern for non-scaled data [(a)–(c)] and for scaled data [(d)–(f)] averaged over the cropped image regions of size reduced to 50 pixels × 50 pixels (a,d), within annular region $q_L = 5$ pixels and $q_H = 10$ pixels (b,e), and within annular region $q_L = 5$ pixels and $q_H = 20$ pixels (c,f). The horizontal-axis tilt(true) represents the ground-truth incident beam angles.
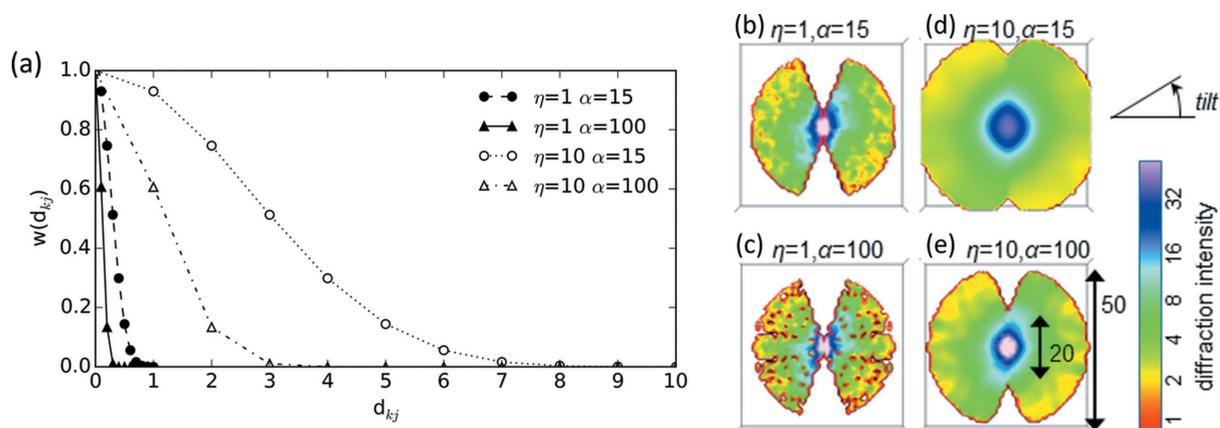
Figure 5
(a) Normalized interpolation weight function based on the Kaiser–Bessel window. d_{kj} is the distance between the center of the voxel k within the volume and the pixel i within the 2D image mapped to the position in the volume. The values of $w(d_{kj})$ are normalized at the values of $d_{kj} = 0$ for each parameter set. (b–e) Cross sections of the Fourier volumes reconstructed using the pre-processed experimental diffraction patterns, their ground-truth angles and various interpolation parameters (the same cross-section view is shown for four volumes). The tilt axis is perpendicular to the provided cross sections.

diffraction pattern frequency used for the 3D reconstruction was set to 0.8 q_{max} (= 20 pixels) to avoid the protrusion of the intensity outside the volume box when using large values of η.

The volumes reconstructed with the interpolation parameter, $\eta = 1$, show discontinuity inside [Figs. 5(b) and 5(c)]. In the 3D reconstruction algorithm, angles are assigned to maximize the consistency between neighboring slices. With such a small η, each voxel is determined merely by the nearest slice (diffraction pattern); in other words, neighboring slices have no influence on each other, and thus there is no inconsistency to be reduced. Indeed, when angular assignment is started from a set of random angles, convergence is often reached within a few iteration steps without any improvement (results not shown). Thus, we need to use a sufficiently large η. With $\eta = 10$, the interpolated 3D volume shows continuity [Figs. 5(d) and 5(e)], indicating that diffraction patterns sufficiently influence each other in an assembled 3D space.

We then optimized other interpolation parameters for angular assignment. We performed iterations starting from the initial reference volumes obtained from images (pre-processed as explained above) with ground-truth orientations with the different interpolation parameters, and examined the resulting tilt angles. A total of 139 reference diffraction patterns were created from −69° to +69° (which was the incident beam angle range used in the tomography experiment), with 1° tilt angle intervals with cubic spline interpolation. It should be noted that the angular range for reference diffraction patterns would not be limited in the case of single-particle XFEL data because these reference patterns would need to be computed in all orientations in such case. Two different matching regions were defined as earlier, i.e. one having $q_L = 5$ and $q_H = 10$ pixels and the other having $q_L = 5$ and $q_H = 20$ pixels. Fig. 6 shows the assigned tilt angles as a result of this test.

With the combination of $\eta = 10$ and $\alpha = 15$, the tilt angles are not well estimated [Figs. 6(a) and 6(b)], presumably because the reconstructed volume becomes too blurred as shown in Fig. 5(d). The use of large values of α compensates for this

effect by emphasizing the nearby images for interpolation (Fig. 5e). By increasing α to 100 with $\eta = 10$, the agreement between the estimated and ground-truth angles greatly improved [Figs. 6(c) and 6(d)], and angles similar to ground-truth angles are obtained.

As we described in §2, we excluded the center region of the experimental diffraction patterns ($q < 5$ pixel) for CC calculations. The diffraction intensities in this region are very high and have large fluctuations such that the CC measure is much less discriminative if the pixels in this region are included in the calculations. The results with $q_H = 10$ were slightly better than those with $q_H = 20$ [Figs. 6(c) and 6(d)]. The use of too
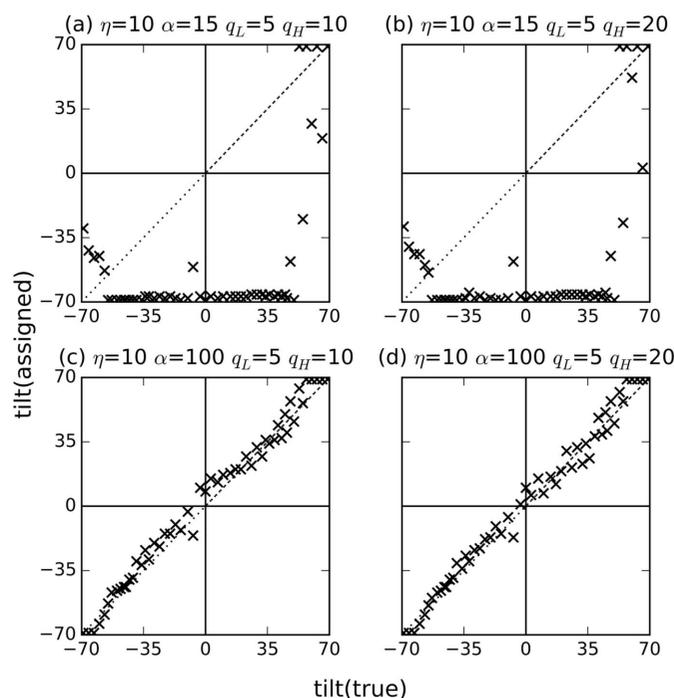
Figure 6
Estimated tilt angles of experimental diffraction patterns using the alignment initiated with ground-truth angles and the 3D reconstruction with various interpolation parameters.

large matching regions would not be beneficial since the pixels farther from the image center are more prone to noise. Therefore, all further experiments were performed using scaled cropped experimental diffraction patterns of size reduced to 50 pixel × 50 pixel, $\eta = 10$ pixel and $\alpha = 100$ for volume reconstruction, and $q_L = 5$ pixel and $q_H = 10$ pixel for orientation estimation (these q_L and q_H correspond to 0.004 nm^{-1} and 0.002 nm^{-1} in reciprocal space, respectively).

3.4. Estimation of the incident beam angle from random initial reference volumes

Using the parameters calibrated in the preliminary analysis presented above, we performed the tilt angle estimation using an initial reference volume computed from experimental images by assigning random orientations. We created five reference initial volumes using five sets of random angles (in the range $[-69°, 69°]$) for each experimental image. It should be again noted that there would not be such an angular range limit when using single-particle XFEL data (the initial volumes would be generated from experimental images at any random orientation in the case of single-particle XFEL data).

Fig. 7 shows the results of the tilt angle estimation. The parameters converged within 20 to 50 iterations. We obtained good results in two of five trials (random#2 and random#5), meaning that the estimated angles were similar to the ground-truth angles in these two trials. For the other three trials, although the estimated tilt angles largely deviated from the correct angles, the relative orientations of nearby diffraction patterns (the diffraction patterns with similar orientations) were partially recovered. For example, in the trial random#3, it appears as if [assigned 35, 70] should be wrapped to $[-100, -70]$, but this is because this reconstructed volume is made from two disconnected groups of images; in one group, the images from [true -35, 70] are correctly aligned and in another the images from [true -65, -35] are aligned, but the connection around the image with [true -35] was not recovered during this iteration. Thus, two groups exchanged their positions in this iteration, and appear 'swapped'. It should be noted, however, that such optimization issues may be less severe for data from a typical XFEL single-particle experiment. In such experiments, different pairs of diffraction patterns may share different common lines, unlike the single-

tilt-axis tomographic CXDM data in which all patterns share just one common line. In the above particular example, if we had a few diffraction patterns that are (near) perpendicular to the tilt-axis, i.e. transverse to the other images, such remaining inconsistencies could be identified and possibly resolved. These results show that our 3D reconstruction method can successfully identify relative orientations of similarly oriented images although the results depend on their initial volume.

Fig. 8 shows the cross sections of the five initial random volumes and the corresponding final reconstructed 3D volumes. For all the trials, the cross sections of the initial and final volumes are significantly different. For the random#2 and random#5 trials, which showed good matching results in Fig. 7, final cross sections were similar to the cross sections of the volume reconstructed using the ground-truth angles and the same interpolation parameters ($\eta = 10$ pixel and $\alpha = 100$) (Fig. 5e). In the random#3 trial, slices were not assigned from 20° to 45° in both Fig. 7(c) and Fig. 8(h).

3.5. Cross-correlation map among diffraction patterns

To further assess the results of the five trial runs, we examined the correlation coefficients between experimental images and all reference images which were sliced from the final volume. We created the cross-correlation map (CC map) as shown in Figs. 9(a)–9(e). In Figs. 9(a)–9(e), horizontal axes represent the image number, which were sorted by the assigned tilt angle at the end of the refinement to check the CC coefficient between assigned and all reference tilt angles. For the random#2 and random#5 trials, which showed good matching results, the maximum CC coefficient values mostly follow diagonal lines. This is more the case for random#2 than for random#5, which is consistent with the matching results as shown in Fig. 7. This means that the diffraction patterns are more smoothly arranged in the final volume for random#2 and random#5 than for other trials.

On the other hand, for unsuccessful reconstructions such as random#1, correlations between the experimental image and the reference slices are not continuously distributed. This is more evident in Figs. 9(f)–9(j), which show the maximum CC coefficient (CC^{max}) between each experimental diffraction pattern and their optimal reference images from the final reconstruction. There are multiple gaps in random#1,
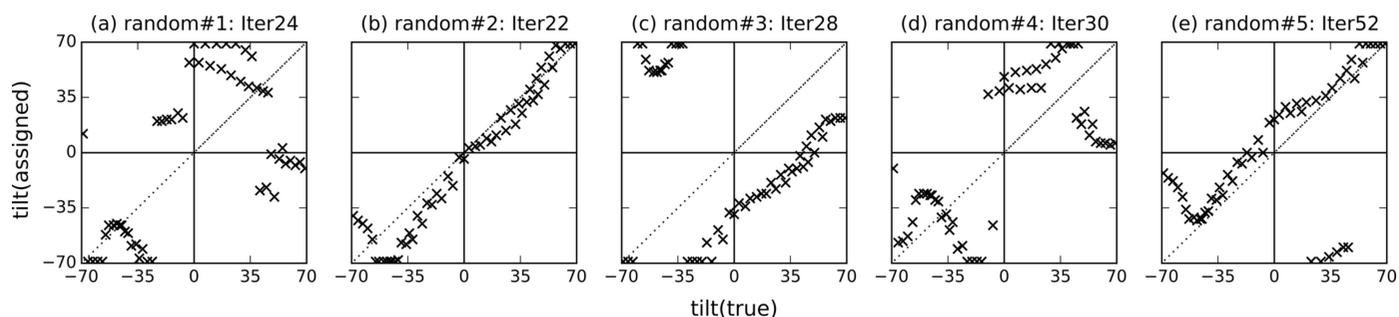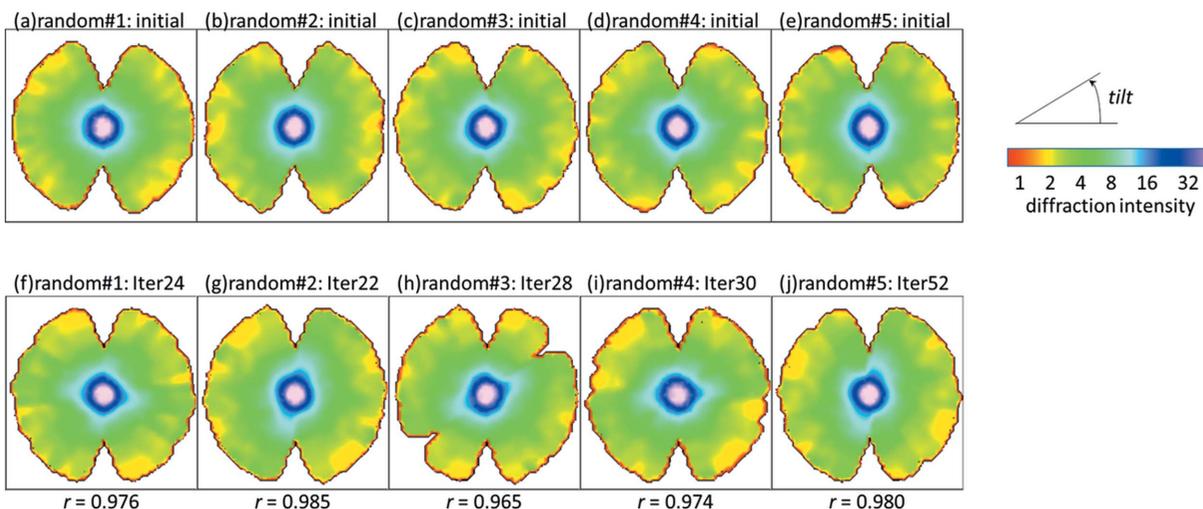

Figure 7
Tilt angles assigned to the experimental images by the proposed method initiated with five random initial volumes. The horizontal and vertical axes represent the ground-truth incident tilt angles and the tilt angles assigned by the proposed method, respectively. Iter is the number of iterations until convergence.
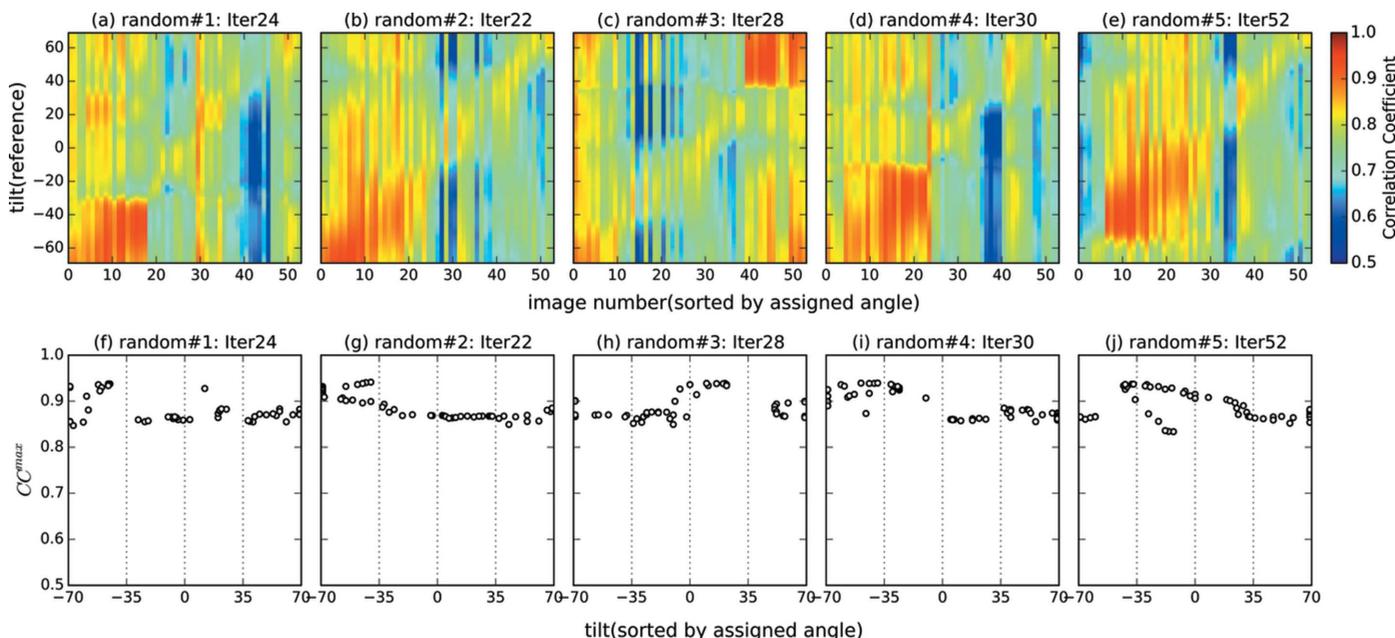
Figure 8
Cross section of (a–e) random initial volumes and (f–j) finally reconstructed volumes. The tilt axis is perpendicular to the cross section. The same cross-section view is shown for all volumes. The r-values shown below panels (f)–(j) are the correlation coefficients between finally reconstructed volumes and ground-truth volume (Fig. 5e).

random#3 and random#4, for example from $-45°$ to $-30°$, from $-20°$ to $-10°$, from $0°$ to $20°$ and from $25°$ to $35°$ in random#1, i.e. these tilt angles were not assigned in their final reconstruction. In some cases, multiple images are assigned to the same angles. For random#2, a set of images with ground-truth angles from $-69°$ to $-45°$ are assigned to the same angles (Fig. 9g). A similar result was obtained for random#5 (from $-45°$ to $-20°$) as shown in Fig. 9(j). This is likely to be due to similarities in the diffraction patterns at these angles. In addition, CC^{max} values are discontinuous before and after the gaps, indicating the fragmented matching. These results show that angular assignments are initial value dependent. Geometrical relations between similar images are identified

by the algorithm and nearby images are sorted correctly; however, discontinuities in the image alignment occasionally occur, resulting in a set of independent blocks of correctly ordered images. However, these artifacts should be less significant for typical single-particle experiments as the rotation angles are expected to be distributed evenly.

4. Discussion

Our slice matching and 3D reconstruction protocol is similar to the cross-correlation maximization method previously developed by Tegze & Bortel (2012), in which a 3D reconstruction of NapAB protein molecule using a large number of

Figure 9
(a–e) CC map between the experimental images and the references created from the final volume. CC coefficients were calculated only for the matching regions ($q_L = 5$ pixels, $q_H = 10$ pixels). The horizontal axis represents the image numbers which were sorted by the assigned tilt angle. (f–j) Distribution of the assigned angle for each experimental image and the maximum CC coefficient (CC^{max}) against the reference images created from the final model.
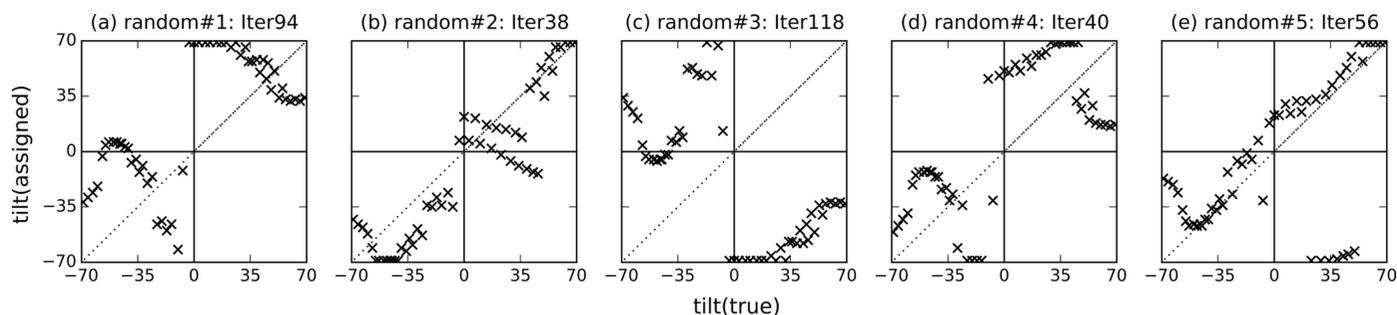
Figure 10
Tilt angles assigned to experimental images by reconstructing the diffraction intensity distribution (using slice matching and starting from the same random angles as in Fig. 7).

simulated diffraction patterns (100000 images) was performed. In this study, we calculate the 3D structure factor amplitude distribution instead of the intensity distribution. Our tests have shown that matching results obtained using the 3D amplitude (Fig. 7) were slightly better than those obtained using the 3D intensity on the system studied here (Fig. 10). This may be attributed to the fact that the structure factor amplitude is smoother than the intensity because the square roots are taken. Also, we consider that it is more useful to reconstruct the 3D amplitude than the 3D intensity because the 3D amplitude can later be directly used to obtain the structure in real space. The results of our study show that such protocols can estimate the orientation of the incident beam angles and the 3D structure factor amplitude with experimental diffraction patterns.

Key parameters of our protocol are the parameters that determine the matching region (q_L and q_H) and the interpolation parameters (η and α). Sometimes, these parameters are best adjusted by trial and error. We found that large q_L and small q_H values give optimal results since the experimental diffraction pattern pixels much further from the image center are unreliable due to noise and the central pixels have too strong intensities which reduce the discriminative power of the CC measure. It is also possible to use a low-wavenumber region first to determine the orientation of diffraction patterns to reconstruct the 3D structure factor amplitude coarsely, and later use a high-wavenumber region to determine the structure with finer details.

In addition, we carefully examined the choice of parameters for the interpolation with the Kaiser–Bessel window. In our test, we assembled the diffraction images of 50 pixel × 50 pixel (the wavenumber at the edge of the diffraction patterns is approximately 0.011 nm^{-1}). Assuming that we evenly distribute 53 images over an angular range of 140°, the distance between the edge pixels on two adjacent images in 3D space is about 1 voxel. Using the combination of $\eta = 10$ and $\alpha = 100$, the weight would decrease to a half value at the voxel about 1.5 units away from the position where the pixel on the 2D diffraction image is mapped within the 3D volume (Fig. 5a). Thus, for each voxel, pixels from three to four images are weight-averaged. These parameter values are appropriate for 3D interpolation from a limited number of 2D diffraction patterns. We have found that the parameters η and α can be

selected in a general case based on the rule that pixels from three to four images should be weight-averaged in 3D space.

We also showed that the reliability of the reconstructed volume could be examined by evaluating the continuity of CC^{max} with respect to the assigned beam angle [Figs. 9(f)–9(j)]. Tegze & Bortel (2016) proposed an approach that evaluates the distribution of CC values between the diffraction patterns and the reference slices [like in Figs. 9(a)–9(e) of our study] assuming that just a few particular diffraction patterns should have high CC values. Our approach is complementary to this approach, focusing on the distribution of assigned angles and CC^{max} continuity.

In single-particle XFEL experiments, the signal-to-noise ratio could be improved by using a large number of diffraction patterns (by averaging many diffraction patterns with similar 3D orientations, as done in single-particle electron microscopy). However, this has not been demonstrated with experimental tomographic CXDM data because a small number of diffraction patterns is usually obtained. Also in our test study (with tomographic CDXM data), only 53 patterns were mapped onto a 140° angular space. Thus, the strategy to improve the signal-to-noise ratio with actual experimental data requires further study.

In addition, in this study, we simply assigned initial orientations randomly, because both the number of images and possibility of assignment angles were small. To create a reliable initial volume could be a challenging problem in single-particle analysis. However, Tegze & Bortel (2012) demonstrated encouraging results that the projection matching of NapAB protein starting from random orientations using 100000 diffraction patterns could converged after 15 iterations. Some optimization methods have been proposed in cryo-EM analysis (Vargas et al., 2014; Sorzano et al., 2015). These studies could help us to create an appropriate initial volume using XFEL diffraction images. By combining with a 3D classification for many final volumes obtained by a large number of trials, we can select the most populated classes as the most plausible volumes.

5. Conclusion

We have developed a protocol for the 3D reconstruction of the absolute value of the structure factor from XFEL diffraction

patterns and software based on *Xmipp* libraries to run it. In this article, we presented this protocol and the results of its tests with experimental diffraction patterns of an aerosol nanoparticle obtained by CXDM. Although the angle estimation for these data is difficult, as only one common line exists between diffraction patterns, encouraging results were obtained. Reconstruction was performed starting from randomly created reference 3D structures, and nearly correct volumes were obtained. We also showed that the plausibility of reconstructed volumes could be evaluated by examining the continuity of CC^{max} with respect to the assigned beam angles. Finally, we showed the sensitivity of the protocol to different parameter values. Additional testing with more experimental data would be necessary to establish more general guidelines to the parameter adjustment.

References

Abrishami, V., Bilbao-Castro, J. R., Vargas, J., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. (2015). *Ultramicroscopy*, **157**, 79–87.

Aloy, P. & Russell, R. B. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 5896–5901.

Barty, A., Marchesini, S., Chapman, H. N., Cui, C., Howells, M. R., Shapiro, D. A., Minor, A. M., Spence, J. C. H., Weierstall, U., Ilavsky, J., Noy, A., Hau-Riege, S. P., Artyukhin, A. B., Baumann, T., Willey, T., Stolken, J., van Buuren, T. & Kinney, J. H. (2008). *Phys. Rev. Lett.* **101**, 055501.

Berman, H., Henrick, K. & Nakamura, H. (2003). *Nat. Struct. Biol.* **10**, 980.

Bunge, H. P. & Baumgardner, J. R. (1995). *Comput. Phys.* **9**, 207–215.

Chapman, H. N., Barty, A., Bogan, M. J., Boutet, S., Frank, M., Hau-Riege, S. P., Marchesini, S., Woods, B. W., Bajt, S., Benner, W. H., London, R. A., Plönjes, E., Kuhlmann, M., Treusch, R., Düsterer, S., Tschentscher, T., Schneider, J. R., Spiller, E., Möller, T., Bostedt, C., Hoener, M., Shapiro, D. A., Hodgson, K. O., van der Spoel, D., Burmeister, F., Bergh, M., Caleman, C., Huldt, G., Seibert, M. M., Maia, F. R. N. C., Lee, R. W., Szöke, A., Timneanu, N. & Hajdu, J. (2006). *Nat. Phys.* **2**, 839–843.

Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B. & Spence, J. C. H. (2011). *Nature (London)*, **470**, 73–77.

Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. (2015). *Cell*, **161**, 438–449.

Drenth, J. (2007). *Principles of Protein X-ray Crystallography*, 3rd ed. New York: Springer.

Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R. N. C., Seltzer, V., Claverie, J. M., Hantke, M., Jönsson, O., Nettelblad, C., van der Schot, G., Liang, M., DePonte, D. P., Barty, A., Seibert, M. M., Iwan, B., Andersson, I., Loh, N. D., Martin, A. V., Chapman, H., Bostedt, C., Bozek, J. D., Ferguson, K. R., Krzywinski, J., Epp, S. W., Rolles, D., Rudenko, A., Hartmann, R., Kimmel, N. & Hajdu, J. (2015). *Phys. Rev. Lett.* **114**, 098102.

Gaffney, K. J. & Chapman, H. N. (2007). *Science*, **316**, 1444–1448.

Gallagher-Jones, M., Bessho, Y., Kim, S., Park, J., Kim, S., Nam, D., Kim, C., Kim, Y., Noh, D. Y., do, Y., Miyashita, O., Tama, F., Joti, Y., Kameshima, T., Hatsui, T., Tono, K., Kohmura, Y., Yabashi, M., Hasnain, S. S., Ishikawa, T. & Song, C. (2014). *Nat. Commun.* **5**, 3798.

Grigorieff, N. (2007). *J. Struct. Biol.* **157**, 117–125.

Henderson, R. (1995). *Q. Rev. Biophys.* **28**, 171–193.

Hirata, K., Shinzawa-Itoh, K., Yano, N., Takemura, S., Kato, K., Hatanaka, M., Muramoto, K., Kawahara, T., Tsukihara, T., Yamashita, E., Tono, K., Ueno, G., Hikima, T., Murakami, H., Inubushi, Y., Yabashi, M., Ishikawa, T., Yamamoto, M., Ogura, T., Sugimoto, H., Shen, J. R., Yoshikawa, S. & Ago, H. (2014). *Nat. Methods*, **11**, 734–736.

Hosseinizadeh, A., Schwander, P., Dashti, A., Fung, R., D'Souza, R. M. & Ourmazd, A. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130326.

Huldt, G., Szőke, A. & Hajdu, J. (2003). *J. Struct. Biol.* **144**, 219–227.

Jiang, H., Song, C., Chen, C. C., Xu, R., Raines, K. S., Fahimian, B. P., Lu, C. H., Lee, T. K., Nakashima, A., Urano, J., Ishikawa, T., Tamanoi, F. & Miao, J. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 11234–11239.

Kimura, T., Joti, Y., Shibuya, A., Song, C., Kim, S., Tono, K., Yabashi, M., Tamakoshi, M., Moriya, T., Oshima, T., Ishikawa, T., Bessho, Y. & Nishino, Y. (2014). *Nat. Commun.* **5**, 3052.

Kirz, J., Jacobsen, C. & Howells, M. (1995). *Q. Rev. Biophys.* **28**, 33–130.

Kühlbrandt, W. (2014). *Science*, **343**, 1443–1444.

Lewitt, R. M. (1990). *J. Opt. Soc. Am. A*, **7**, 1834–1846.

Loh, N. T. D. & Elser, V. (2009). *Phys. Rev.* **80**, 026705.

Lučić, V., Förster, F. & Baumeister, W. (2005). *Annu. Rev. Biochem.* **74**, 833–865.

Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. (2013). *J. Struct. Biol.* **183**, 377–388.

Miao, J., Chen, C. C., Song, C., Nishino, Y., Kohmura, Y., Ishikawa, T., Ramunno-Johnson, D., Lee, T. K. & Risbud, S. H. (2006). *Phys. Rev. Lett.* **97**, 215503.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.

Penczek, P. A., Grassucci, R. A. & Frank, J. (1994). *Ultramicroscopy*, **53**, 251–270.

Philipp, H. T., Ayyer, K., Tate, M. W., Elser, V. & Gruner, S. M. (2012). *Opt. Express*, **20**, 13129–13137.

Rosa-Trevín, J. M. de la, Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M. & Sorzano, C. O. S. (2013). *J. Struct. Biol.* **184**, 321–328.

Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, 1st ed. New York: Garland Science.

Scheres, S. H. W. (2012). *J. Struct. Biol.* **180**, 519–530.

Scheres, S. H. W., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P. B., Frank, J. & Carazo, J. M. (2007). *Nat. Methods*, **4**, 27–29.

Schot, G. van der, Svenda, M., Maia, F. R. N. C., Hantke, M., DePonte, D. P., Seibert, M. M., Aquila, A., Schulz, J., Kirian, R., Liang, M., Stellato, F., Iwan, B., Andreasson, J., Timneanu, N., Westphal, D., Almeida, F. N., Odic, D., Hasse, D., Carlsson, G. H., Larsson, D. S. D., Barty, A., Martin, A. V., Schorb, S., Bostedt, C., Bozek,

J. D., Rolles, D., Rudenko, A., Epp, S., Foucar, L., Rudek, B., Hartmann, R., Kimmel, N., Holl, P., Englert, L., Duane Loh, N., Chapman, H. N., Andersson, I., Hajdu, J. & Ekeberg, T. (2015). *Nat. Commun.* **6**, 5704.

Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., Hantke, M., DePonte, D. P., Barty, A., Schulz, J., Gumprecht, L., Coppola, N., Aquila, A., Liang, M., White, T. A., Martin, A., Caleman, C., Stern, S., Abergel, C., Seltzer, V., Claverie, J., Bostedt, C., Bozek, J. D., Boutet, S., Miahnahri, A. A., Messerschmidt, M., Krzywinski, J., Williams, G., Hodgson, K. O., Bogan, M. J., Hampton, C. Y., Sierra, R. G., Starodub, D., Andersson, I., Bajt, S., Barthelmess, M., Spence, J. C. H., Fromme, P., Weierstall, U., Kirian, R., Hunter, M., Doak, R. B., Marchesini, S., Hau-Riege, S. P., Frank, M., Shoeman, R. L., Lomb, L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Schmidt, C., Foucar, L., Kimmel, N., Holl, P., Rudek, B., Erk, B., Hömke, A., Reich, C., Pietschner, D., Weidenspointner, G., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Schlichting, I., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K., Andritschke, R., Schröter, C., Krasniqi, F., Bott, M., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Chapman, H. N. & Hajdu, J. (2011). *Nature (London)*, **470**, 78–81.

Sekiguchi, Y., Oroguchi, T. & Nakasako, M. (2016). *J. Synchrotron Rad.* **23**, 312–323.

Shaikh, T. R., Gao, H., Baxter, W. T., Asturias, F. J., Boisset, N., Leith, A. & Frank, J. (2008). *Nat. Protoc.* **3**, 1941–1974.

Sigworth, F. J., Doerschuk, P. C., Carazo, J. M. & Scheres, S. H. W. (2010). *Methods Enzymol.* **482**, 263–294.

Sorzano, C. O. S., Jonić, S., El-Bez, C., Carazo, J. M., De Carlo, S., Thévenaz, P. & Unser, M. (2004). *J. Struct. Biol.* **146**, 381–392.

Sorzano, C. O. S., Vargas, J., de la Rosa-Trevin, J. M., Oton, J., Alvarez-Cabrera, A. L., Abrishami, V., Sesmero, E., Marabini, R. & Carazo, J. M. (2015). *J. Struct. Biol.* **189**, 213–219.

Suga, M., Akita, F., Hirata, K., Ueno, G., Murakami, H., Nakajima, Y., Shimizu, T., Yamashita, K., Yamamoto, M., Ago, H. & Shen, J. (2014). *Nature (London)*, **517**, 99–103.

Takayama, Y., Inui, Y., Sekiguchi, Y., Kobayashi, A., Oroguchi, T., Yamamoto, M., Matsunaga, S. & Nakasako, M. (2015). *Plant Cell Physiol.* **56**, 1272–1286.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. (2007). *J. Struct. Biol.* **157**, 38–46.

Tegze, M. & Bortel, G. (2012). *J. Struct. Biol.* **179**, 41–45.

Tegze, M. & Bortel, G. (2016). *Acta Cryst.* A**72**, 459–464.

Tokuhisa, A., Jonic, S., Tama, F. & Miyashita, O. (2016). *J. Struct. Biol.* **194**, 325–336.

Tokuhisa, A., Taka, J., Kono, H. & Go, N. (2012). *Acta Cryst.* A**68**, 366–381.

Vargas, J., Alvarez-Cabrera, A.-L., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. (2014). *Bioinformatics*, **30**, 2891–2898.

Watson, J. D. & Crick, F. H. C. (1953). *Nature (London)*, **171**, 737–738.

Xu, R., Jiang, H., Song, C., Rodriguez, J. A., Huang, Z., Chen, C. C., Nam, D., Park, J., Gallagher-Jones, M., Kim, S., Kim, S., Suzuki, A., Takayama, Y., Oroguchi, T., Takahashi, Y., Fan, J., Zou, Y., Hatsui, T., Inubushi, Y., Kameshima, T., Yonekura, K., Tono, K., Togashi, T., Sato, T., Yamamoto, M., Nakasako, M., Yabashi, M., Ishikawa, T. & Miao, J. (2014). *Nat. Commun.* **5**, 4061.

Yang, Z. & Penczek, P. A. (2008). *Ultramicroscopy*, **108**, 959–969.