JOURNAL OF SYNCHROTRON RADIATION
ISSN 1600-5775

Classification and assessment of retrieved electron density maps in coherent X-ray diffraction imaging using multivariate analysis

Yuki Sekiguchi,[a,b] Tomotaka Oroguchi[a,b] and Masayoshi Nakasako[a,b]*

[a]Department of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan, and [b]RIKEN SPring-8 Center, 1-1-1 Kohto, Sayo, Sayo-gun, Hyogo 679-5148, Japan. *Correspondence e-mail: nakasako@phys.keio.ac.jp

Received 1 July 2015
Accepted 29 September 2015

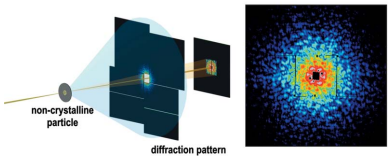
Edited by M. Yamamoto, RIKEN SPring-8 Center, Japan

Keywords: coherent X-ray diffraction imaging; X-ray free-electron laser; structure analysis of non-crystalline particles.

Coherent X-ray diffraction imaging (CXDI) is one of the techniques used to visualize structures of non-crystalline particles of micrometer to submicrometer size from materials and biological science. In the structural analysis of CXDI, the electron density map of a sample particle can theoretically be reconstructed from a diffraction pattern by using phase-retrieval (PR) algorithms. However, in practice, the reconstruction is difficult because diffraction patterns are affected by Poisson noise and miss data in small-angle regions due to the beam stop and the saturation of detector pixels. In contrast to X-ray protein crystallography, in which the phases of diffracted waves are experimentally estimated, phase retrieval in CXDI relies entirely on the computational procedure driven by the PR algorithms. Thus, objective criteria and methods to assess the accuracy of retrieved electron density maps are necessary in addition to conventional parameters monitoring the convergence of PR calculations. Here, a data analysis scheme, named ASURA, is proposed which selects the most probable electron density maps from a set of maps retrieved from 1000 different random seeds for a diffraction pattern. Each electron density map composed of J pixels is expressed as a point in a J-dimensional space. Principal component analysis is applied to describe characteristics in the distribution of the maps in the J-dimensional space. When the distribution is characterized by a small number of principal components, the distribution is classified using the k-means clustering method. The classified maps are evaluated by several parameters to assess the quality of the maps. Using the proposed scheme, structure analysis of a diffraction pattern from a non-crystalline particle is conducted in two stages: estimation of the overall shape and determination of the fine structure inside the support shape. In each stage, the most accurate and probable density maps are objectively selected. The validity of the proposed scheme is examined by application to diffraction data that were obtained from an aggregate of metal particles and a biological specimen at the XFEL facility SACLA using custom-made diffraction apparatus.

1. Introduction

Coherent X-ray diffraction imaging (CXDI) is a lens-less imaging technique for visualizing the structures of non-crystalline particles with dimensions from submicrometers to micrometers at resolutions of several tens of nanometers (Miao *et al.*, 2008). In CXDI experiments, a spatially isolated non-crystalline particle is illuminated by a coherent X-ray beam, and the Fraunhofer diffraction pattern is recorded by an area detector with a sufficiently high sampling frequency to satisfy the oversampling condition (Miao *et al.*, 2003) (Fig. 1*a*). Then, in principle, the electron density map of the specimen particle can be reconstructed by applying phase-retrieval (PR) algorithms (Fienup, 1982) to the oversampled diffraction



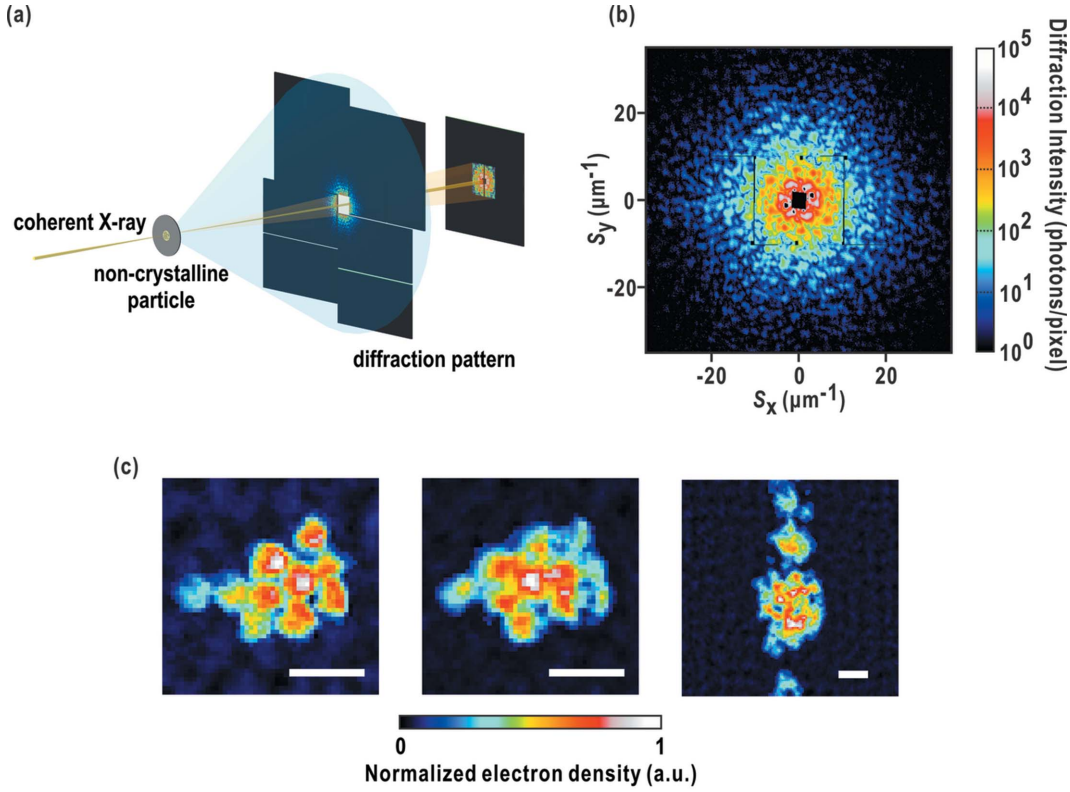


Figure 1
(*a*) Schematic illustration of our CXDI experiment using the KOTOBUKI-1 apparatus and the MPCCD Octal and MPCCD Dual detectors at BL3 of SACLA. (*b*) A representative diffraction pattern from an aggregate of 250 nm gold colloidal particles after merging two MPCCD detectors. (*c*) Three types of electron density map retrieved from the diffraction pattern in panel (*b*). The scale bars indicate 500 nm.

pattern. Because of the large penetration depth of X-rays with short wavelengths, CXDI has the potential to visualize thick specimens larger than 500 nm at a resolution of several tens of nanometers without sectioning or chemical labeling.

Since the first demonstration in 1999 (Miao *et al.*, 1999), many CXDI experiments utilizing synchrotron X-rays have demonstrated the potential to visualize internal structures of non-crystalline particles from materials science and biology (Williams *et al.*, 2003; Shapiro *et al.*, 2005; Miao *et al.*, 2006; Nishino *et al.*, 2009; Takayama & Nakasako, 2012; Nam *et al.*, 2013). CXDI experiments utilizing X-ray free-electron lasers (XFELs) are expected to achieve structure analyses of non-crystalline particles at higher resolutions than those in synchrotron experiments (Seibert *et al.*, 2011; Loh *et al.*, 2012; Nakasako *et al.*, 2013; Hantke *et al.*, 2014; Xu *et al.*, 2014; Kimura *et al.*, 2014). In XFEL-CXDI experiments, the high brilliance and ultra-short duration of XFEL pulses enable us to collect diffraction patterns from non-crystalline particles by single shots in the 'diffraction before destruction' scheme (Neutze *et al.*, 2000; Chapman *et al.*, 2006*a*, 2014). Furthermore, the high repetition rates of XFEL pulses enable us to collect a huge number of diffraction patterns in a short period of time. Thus, fast, automatic and accurate structure analyses are necessary for efficient utilization of XFELs by overcoming inherent problems in PR calculations as described below.

In general, diffraction patterns lose phase information. In X-ray protein crystallography, the phase information is experimentally estimated according to changes in diffraction intensities, for instance, caused by heavy-atom labeling of protein molecules (Drenth, 2007). In CXDI, the phase determination of diffracted X-rays is entirely performed by a computational procedure with little experimental evidence to support the retrieved phases. Mainly owing to the beam stop and saturation of detector pixels, diffraction patterns miss data in very small-angle regions which contain structural information regarding the overall shape and total electrons of specimen particles. In addition, diffraction patterns are affected by Poisson noise in X-ray detection. Because of these factors, it is often difficult to obtain a unique solution for electron density maps of specimen particles.

In this regard, as a representative example, we show electron density maps retrieved from a diffraction pattern of an aggregate of 250 nm gold colloidal particles [Figs. 1(*a*) and 1(*b*)]. We conducted 1000 trials of PR calculations using different initial maps, which we roughly classify into three groups. The maps in the first group almost approximate the shape of the aggregate [left-hand panel in Fig. 1(*c*)], but those in the second group appear as aggregates of degraded or fused particles [middle panel in Fig. 1(*c*)]. For the maps in the third group, there are no electron densities that can be attributed to gold colloidal particles [right panel in Fig. 1(*c*)]. In addition, in the first group, electron density maps display significant fluctuations with almost the same overall shapes. It should be noted that the parameters conventionally used to examine the

correctness of the maps are occasionally better in the second or third group than those in the first group.

As one of the current fashions to obtain plausible electron density maps, researchers perform hundreds of PR-calculation trials for a single diffraction pattern starting from random initial maps. Subsequently, part of retrieved maps, which have similar shapes to each other and display good scores of criteria for evaluation, are averaged. Frequently, retrieved maps are compared with particle images observed in optical microscopy (OM) and/or electron microscopy (EM). However, because the internal fine structures in thick specimens are difficult to observe except by CXDI, high-resolution structures of reconstructed maps cannot be confirmed by other imaging techniques. As previously mentioned, because CXDI itself provides little physical evidence to confirm the accuracy of a retrieved phase set, any scheme for assessing the results of structure analyses is necessary to provide opportunities for more objective selection of retrieved electron density maps.

In the previous studies, we proposed schemes to reconstruct the three-dimensional electron density maps of biological macromolecules in future XFEL-CXDI experiments (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013). We treated the phase-retrieved projection electron density maps composed of J pixels as points in a J-dimensional space, the ith axis of which represents the electron density value of the ith pixel. In this study we also treat a large number of electron density maps retrieved from a diffraction pattern as points in a multidimensional space. Because it is difficult to visualize the distribution of maps in the multidimensional space directly, we attempt to describe the characteristics of the distribution within a small number of dimensions through multivariate analyses. By classifying the maps and calculating parameters to assess their qualities, we evaluate the correctness and accuracy of the maps.

Here we present the proposed scheme, named ASURA, in detail and apply the scheme to experimental diffraction patterns from non-crystalline particles collected at the XFEL facility SACLA.

2. Method

2.1. Outline of the proposed scheme

In many PR calculations, the support shape and internal electron density distribution within the support are simultaneously estimated. In the proposed scheme, we separate the PR process into two stages (Fig. 2): the determination of the most probable overall support shape, and the subsequent estimation of electron density distribution with fine structures inside the support to explain the diffraction pattern.

First, we prepare N (1000 in this study) electron density maps having J pixels (50×50 or 60×60 in this study). The maps are retrieved from a diffraction pattern starting from different initial maps with random electron densities. The electron density of the ith pixel in a map is regarded as the value of the ith axis in the J-dimensional space. Thus, each electron density map can be represented as a point in the J-dimensional space.

Principal component analysis (PCA) is applied to the set of maps in the J-dimensional space (van Heel & Frank, 1981). When the PCA suggests the possibility that the variance among the N electron density maps can be described predominantly by a small number of principal components (PCs) with a minimal loss of information, the k-means clustering method (MacQueen, 1967) is applied to classify the maps in a space spanned by a small number of PCs. By referring to the averaged map and the parameters used to examine the accuracy in each class, we objectively select the best support shape for the subsequent analysis. In the selection, we can refer to the maximum size of the particle estimated by the autocorrelation function of the diffraction pattern (Kobayashi *et al.*, 2014). The particle shapes from OM and/or EM may be helpful for the selection if available in this stage.

In the second stage, the fine structures inside the best support are estimated. We apply PCA to the N electron density maps retrieved under the fixed support condition. The N maps are classified by the same procedure in the first stage. We select the best class which is composed of maps with high accuracy and supported by several parameters defined in the following section. Finally, we obtain a map averaged for the components of the best class. In this stage, OM and/or EM provide little structural information to select the most probable electron density maps with fine internal structures.

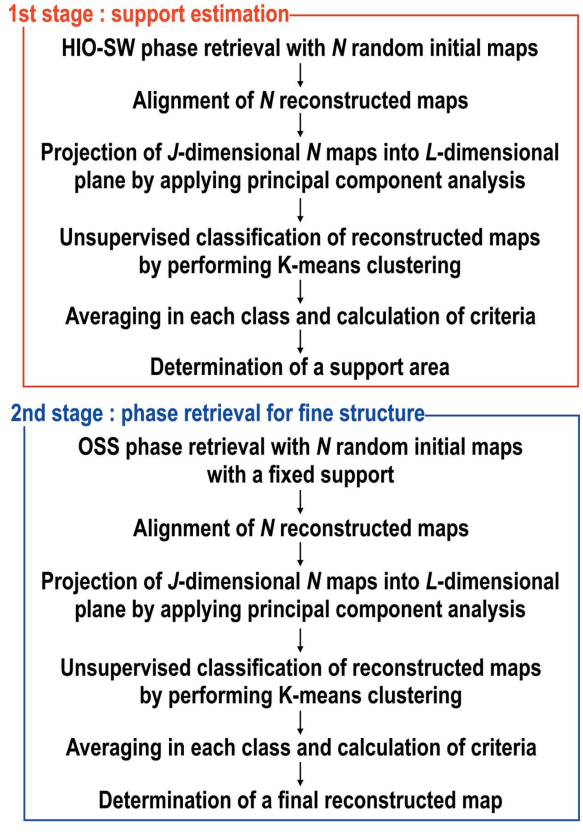


Figure 2
Flowchart of the two-stage structure analysis. The details are described in the *Method* section of the main text.

2.2. PR algorithms used in the proposed scheme

In the first stage, we retrieve the projection electron density maps from a diffraction pattern by combining the hybrid-input–output (HIO) (Fienup, 1982) and shrink-wrap (SW) (Marchesini *et al.*, 2003) algorithms (HIO-SW), which were developed in our previous study (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013; Nakasako *et al.*, 2013; Sekiguchi *et al.*, 2014*a*). The initial support is estimated as an area where the autocorrelation function calculated from the diffraction pattern becomes asymptotically close to zero (Kobayashi *et al.*, 2014). The following equation is used as a real-space constraint in the HIO algorithm,

$$\rho_{k+1}(\mathbf{r}) = \begin{cases} \rho'_k(\mathbf{r}) & \mathbf{r} \in \text{Support and } \rho'_k(\mathbf{r}) \geq 0, \\ \rho_k(\mathbf{r}) - \beta\rho'_k(\mathbf{r}) & \text{otherwise}, \end{cases} \quad (1)$$

where $\rho_k(\mathbf{r})$ is the map at the beginning of the kth HIO-cycle. $\rho'_k(\mathbf{r})$ is the inverse Fourier transformation of a structure factor with the observed amplitude and the phase calculated from $\rho_k(\mathbf{r})$. $\rho_{k+1}(\mathbf{r})$ is an electron density map generated for the next cycle. β is the weight parameter fixed throughout HIO iterations.

After the hth set of HIO iterations of n cycles, the support is updated by the SW algorithm to exclude areas with densities less than a specified threshold after convoluting the following Gaussian to the HIO-retrieved map as a low-pass filter,

$$G_h(\mathbf{r}) = \frac{1}{2\pi{\sigma_h}^2} \exp\left[-|\mathbf{r}|^2/2{\sigma_h}^2\right]. \quad (2)$$

The value of the standard deviation σ_h is changed cycle-dependently. The new support is defined as the area, the electron density of which is higher than a threshold value calculated by multiplying a parameter δ by the maximum density in the low-pass-filtered map.

In the second stage, we used the oversampling smoothness (OSS) algorithm (Rodriguez *et al.*, 2013) to obtain the most probable electron density map inside the best support. The real-space restraint used is

$$\rho''_k(\mathbf{r}) = \begin{cases} \rho'_k(\mathbf{r}) & \mathbf{r} \in \text{Support and } \rho'_k(\mathbf{r}) \geq 0, \\ \rho_k(\mathbf{r}) - \beta\rho'_k(\mathbf{r}) & \text{otherwise}, \end{cases}$$

$$\rho_{k+1}(\mathbf{r}) = \begin{cases} \rho''_k(\mathbf{r}) & \mathbf{r} \in \text{Support}, \\ \mathcal{F}^{-1}\left[G''_k(\mathbf{S})\, W(\mathbf{S})\right] & \mathbf{r} \notin \text{Support}, \end{cases} \quad (3)$$

$$W(\mathbf{S}) = \exp\left[-\frac{1}{2}\left(\frac{\mathbf{S}}{\alpha}\right)^2\right],$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform and $G''(\mathbf{S})$ is the Fourier transform of $\rho''_k(\mathbf{r})$. For maps $j \times j$ pixels in size, parameter α changes from j to $1/j$ every 1000 iterations linearly.

2.3. Multivariate analysis

We classify a large number (N) of electron density maps with J pixels retrieved from a diffraction pattern in each stage. However, it is difficult to visualize the distribution of maps in the J-dimensional space. Thus, using PCA, we examine the possibility that we can characterize the distribution in a low-dimensional space with a minimal loss of information (van Heel & Frank, 1981).

We define a matrix X to express N electron density maps comprising J pixels as

$$X = \begin{pmatrix} x_{11} - \langle x_1 \rangle & x_{12} - \langle x_2 \rangle & \cdots & x_{1J} - \langle x_J \rangle \\ x_{21} - \langle x_1 \rangle & x_{22} - \langle x_2 \rangle & \cdots & x_{2J} - \langle x_J \rangle \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} - \langle x_1 \rangle & x_{N2} - \langle x_2 \rangle & \cdots & x_{NJ} - \langle x_J \rangle \end{pmatrix} \quad (4)$$

where x_{ij} is the electron density at the jth pixel of the ith map, and $\langle x_j \rangle$ is the averaged electron density of the jth pixel among the N maps. Then, eigenvalues and eigenvectors of the covariance matrix $D = X^t X$ are calculated. Eigenvectors represent the directions along which N electron density maps are distributed in the J-dimensional space. Eigenvalues are indices reflecting the variance of the distribution. Each eigenvector has J elements and can be visualized as images with J pixels. When the eigenvectors for the L ($L \ll J$) largest eigenvalues contribute predominantly to the variance among the N maps, the distribution in the J-dimensional space can be represented in an L-dimensional space with minimum loss of information about distribution. Then, the distribution of the maps is displayed as their projection onto the L eigenvectors. The projections onto the L axes are mathematically calculated as the inner product of the J-dimensional maps and the L eigenvectors.

The maps projected onto the L-dimensional space are classified by the k-means clustering method (MacQueen, 1967). Assuming that the maps are classified into M classes, we minimize the sum of squared distances (T) between the maps and the centroids of the classes, which is defined as

$$T = \sum_{m=1}^{M} \sum_{\mathbf{y}_{im} \in m} \left(\mathbf{y}_{im} - \langle \mathbf{y}_m \rangle\right)^2, \quad (5)$$

where $\mathbf{y}_{im}$ is an L-dimensional vector indicating the position of the ith map belonging to the mth class and $\langle \mathbf{y}_m \rangle$ is the centroid of the mth class.

2.4. Parameters characterizing phase-retrieved electron density maps in each class

Here we use four parameters to characterize the phase-retrieved electron density maps: the number of electron density maps in each class defined by the k-means clustering; crystallographic R-factor (Drenth, 2007); zero-angle diffraction intensity; and estimated effective resolution with phase-retrieval transfer function (PRTF) (Chapman *et al.*, 2006*b*). By comparing the parameters and averaged electron density maps in each class, we select the class yielding the most probable electron density map.

2.4.1. Number of maps in each class. The number of maps included in each class can be used as a criterion. This is based on the idea that probable maps appear more frequently than incorrect maps. The most frequently appearing maps are often

accepted as probable maps in XFEL-CXDI analyses (Park *et al.*, 2013; Kimura *et al.*, 2014; van der Schot *et al.*, 2015).

2.4.2. Crystallographic *R*-factor. The crystallographic *R*-factor (Drenth, 2007) is defined as

$$R = \frac{\sum_{\mathbf{S}} \left| \sqrt{I_{\rm obs}(\mathbf{S})} - k\sqrt{I_{\rm cal}(\mathbf{S})} \right|}{\sum_{\mathbf{S}} \sqrt{I_{\rm obs}(\mathbf{S})}}, \quad (6)$$

where $I_{\rm obs}(\mathbf{S})$ and $I_{\rm cal}(\mathbf{S})$ are the observed and calculated diffraction intensities, respectively, at scattering vector $\mathbf{S}$. $I_{\rm cal}(\mathbf{S})$ is calculated using an averaged density map in each class. The scale factor k is defined to equalize the sum of the experimental and calculated diffraction intensity. Thus, the crystallographic *R*-factor measures how the structure amplitude calculated using the structure model is similar to the observed set. The difference between observed and calculated diffraction amplitude is conventionally used as an error function or a convergence criterion from the beginning of development for PR algorithms (Fienup, 1982).

2.4.3. Zero-angle diffraction intensity. In the calculated diffraction intensity, the zero-angle diffraction I_0 (forward scattering) is proportional to the squared total sum of the electron density. When a PR calculation is unsuccessful, I_0 tends to become extremely large (Nishino *et al.*, 2003). Here we use the average of I_0 within each class.

2.4.4. Estimated effective resolution with phase-retrieval transfer function. To monitor the consistency of the retrieved phases (*i.e.* electron density maps) within a class, we define an average of phase terms at scattering vector $\mathbf{S}$ as

$$r(\mathbf{S}) = \left| \sum_{j=1}^{N} \exp\left[i\varphi_j^{\rm cal}(\mathbf{S})\right] \right| \Big/ N, \quad (7)$$

where $\varphi_j^{\rm cal}(\mathbf{S})$ is the phase of the jth map in the class, and N is the number of maps included in the class. The radial average of $r(\mathbf{S})$ is known as the PRTF (Chapman *et al.*, 2006*b*). In CXDI, the PRTF is used to evaluate the effective resolution of an average of reconstructed maps. To yield an effective resolution, typical threshold values of PRTF are 0.5 (Jiang *et al.*, 2010) and $1/e$ (0.368) (Seibert *et al.*, 2011).

Here we consider the relation of $r(\mathbf{S})$ with the figure of merit (FOM) (Blow & Crick, 1959), which is a conventional criterion for examining the reliability of the phase set in X-ray protein crystallography and EM. The FOM is defined as

$$\mathrm{FOM}(\mathbf{S}) = \frac{\left| \sum_{k} P\left[\varphi_k(\mathbf{S})\right] \exp\left[i\varphi_k(\mathbf{S})\right] \right|}{\sum_{k} P\left[\varphi_k(\mathbf{S})\right]}, \quad (8)$$

where $P[\varphi_k(\mathbf{S})]$ is an experimentally determined phase probability distribution function for the kth phase angle at the scattering vector $\mathbf{S}$. When the distribution of the phase angles are measured finely, the sum of $\varphi_j^{\rm cal}(\mathbf{S})$ terms can be approximated as

$$\sum_{j=1}^{N} \exp\left[i\varphi_j^{\rm cal}(\mathbf{S})\right] \approx \sum_{k} P\left[\varphi_k(\mathbf{S})\right] \exp\left[i\varphi_k(\mathbf{S})\right], \quad (9)$$

where $\sum_k P[\varphi_k(\mathbf{S})] = N$. Consequently, we obtain the following approximation,

$$r(\mathbf{S}) \approx \mathrm{FOM}(\mathbf{S}). \quad (10)$$

In single-particle analysis of cryo-electron microscopy, the effective resolution of a reconstructed three-dimensional map is often defined as the resolution where the radial average of the FOM decreases to 0.5 (Rosenthal & Henderson, 2003). In X-ray crystallography, electron density maps with radially averaged FOM of 0.5 are commonly regarded as interpretable, and thus researchers can build molecular models (Lunin & Woolfson, 1993). For instance, the automated model building software *ARP/wARP* is usually applied to electron density maps of radially averaged FOM of around 0.5 at a resolution of 2.5–3.0 Å (Perrakis *et al.*, 1997). Thus we define the effective resolution of an averaged map in the proposed scheme as the resolution at which the radial average of the FOM, *i.e.* the PRTF, decreases to 0.5.

3. Experimental procedure

3.1. Specimen preparation

Gold colloidal particles with a diameter of 250 nm (British Bio Cell International Solutions, UK) were dispersed randomly on silicon nitride membranes (Norcada, Canada) by using a micropipette. Purified chloroplasts from *Cyanidioschyzon merolae* were dispersed on carbon membranes under a humidity-controlled atmosphere, and then rapidly frozen by liquid ethane according to the procedure reported previously (Takayama & Nakasako, 2012; Takayama *et al.*, 2015).

3.2. XFEL-CXDI experiment and data processing

We performed a CXDI experiment at EH3 of BL3 (Tono *et al.*, 2013) at the XFEL facility SACLA. A specimen holder fixing the membranes was set in a diffractometer named KOTOBUKI-1 (Nakasako *et al.*, 2013) and scanned against incident X-ray pulses (Sekiguchi *et al.*, 2014*b*). The intensity and duration of single X-ray pulses with a photon energy of 5.5 keV were approximately 10^{10}–10^{11} photons μm^{-2} pulse^{-1} and 10 fs, respectively. The X-ray pulses were supplied to the experimental hutch at a repetition rate of 1 Hz by using a selector device installed in the beamline. Diffraction patterns were recorded by using multi-port CCD Octal and Dual detectors (Kameshima *et al.*, 2014) placed 1.6 m and 3.2 m downstream from the specimen position, respectively. The central aperture of the Octal detector and the thickness of attenuators in front of the Dual detector were varied depending on the diffraction intensity from specimens.

Subtraction of detector background noise and merging of the diffraction patterns from the two detectors were carried out by using our custom-made data-processing program suite *G-SITENNO* (Sekiguchi *et al.*, 2014*a*,*b*). Each diffraction pattern was binned by summing 2 × 2 pixels into one pixel, and the resulting 128 × 128 pixels were used for PR calculations.

3.3. PR calculations

In the PR calculation in the first stage, we fixed the parameter β in equation (1) to be 0.9. In addition, the standard deviation value of equation (2) in the SW calculation applied after every 100 calculations was varied cycle-dependently. For the hth cycle of the SW, the standard deviation value was given by the following equation,

$$\sigma_h \text{ pixels} = \begin{cases} 2.0 \times 0.98^{h-1} & \sigma_{h-1} \times 0.98 > 0.90, \\ 0.90 & \sigma_{h-1} \times 0.98 < 0.90. \end{cases} \quad (11)$$

The parameter δ defining the threshold value was empirically determined and fixed throughout all SW support updates. After 100 support updates by the SW, 1000 HIO iterations were additionally performed to obtain the final electron density map. Eventually, 11000 HIO iterations were performed with each initial random map.

In the PR calculation in the second stage, we carried out 20000 OSS iterations by varying the parameter α in equation (3) for 20 times with the input of a fixed support.

The retrieved electron density maps have large blank areas outside the support, which originate from the oversampling. Because these blank areas only increase computational cost of the remaining analyses, a large part of them were trimmed away. For example, reconstructed maps with 128 × 128 pixels were reduced to 60 × 60 or 50 × 50 pixels by this procedure.

All data processing by the *G-SITENNO* suite and PR calculations with multivariate analyses were performed on a supercomputer installed in SACLA composed of 960 cores of Intel Xeon CPU X5690 (3.47 GHz per core) (Joti *et al.*, 2015).

3.4. Multivariate analysis

The retrieved electron density maps contain no information on absolute translational positions as diffraction patterns lose phases. Moreover, the maps are ambiguous regarding the π-rotation because of the Friedel's centrosymmetry. Thus, prior to PCA, the electron density maps were superimposed regarding their centroids with accuracy of 1 pixel, and then we selected either the 0-rotation or π-rotation of each map by calculating a correlation coefficient for a selected reference map.

We applied PCA to the π-rotated or 0-rotated electron density maps, and then the maps were projected onto the two-dimensional principal-component space. These maps were classified by the k-means clustering method assuming ten classes through our experiences. Classification by the k-means clustering method strongly depends on the distribution of initial random centroids given for assumed classes. Hence, we performed 100 independent trials and adopted the clustering result with the minimum T.

4. Results

To examine the effectiveness and efficiency of the proposed scheme, we targeted representative diffraction patterns of non-crystalline particles from materials science and biology obtained by the single-shot diffraction experiment. The selected diffraction patterns displayed good signal-to-noise ratios (SNRs) beyond a resolution of 50 nm. Here we describe how the most probable electron density maps were determined through the two-stages analysis of PR calculation and multivariate analysis.

4.1. Diffraction patterns

A diffraction pattern from an aggregate of 250 nm gold particles was recorded in the reciprocal-space resolution range from approximately 1.9 to 68.0 μm^{-1} (corresponding to a real-space resolution of 525–14.7 nm). For the structure analysis, we selected a region with diffraction intensity of approximately 10 photons pixel^{-1} and SNR > 3 (Fig. 3*a*).

The chloroplasts occasionally diffracted X-rays beyond a resolution of 20 μm^{-1}. We extracted a diffraction pattern with

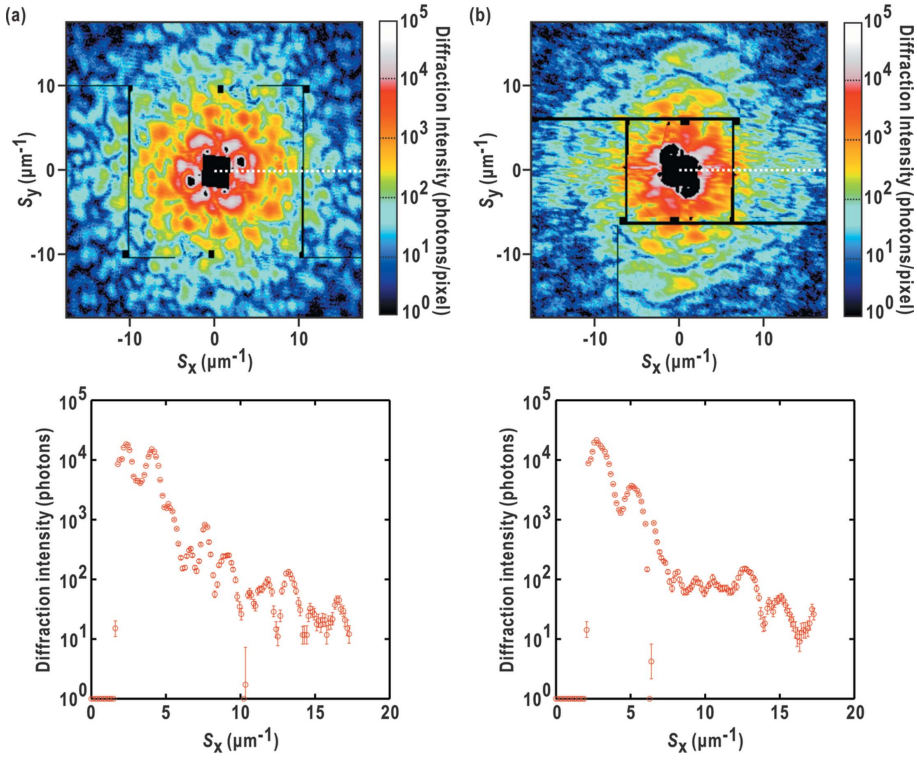


Figure 3
Representative diffraction patterns from an aggregate of 250 nm gold colloidal particles (*a*) and from a chloroplast of *C. merolae* (*b*) shown at resolutions up to 17.4 μm^{-1} at the edge (corresponding to a resolution of 57.4 nm in the real space). The intensity profiles along the dotted lines in the diffraction patterns are shown in the lower panels. Error bars in the line profiles are the square root of the intensity.

a good SNR up to a resolution of 17.4 μm^{-1} (corresponding to 57.4 nm resolution in real space) (Fig. 3*b*). As typically observed in the line profile, the SNR in each resolution shell was almost better than 3 up to the edge of the extracted region. The speckle pattern comprises concentric rings indicating the globular overall shape.

4.2. Structure analysis of an aggregate of gold colloidal particles

4.2.1. Overall shape. Through HIO-SW calculations with the parameter δ set as 0.04 we retrieved 1000 electron density maps. After the PCA, the electron density maps in a 60 × 60-dimensional space were projected onto the plane spanned by the first and second PCs [Figs. 4(*a*) and 4(*b*)], which described 60% of the total variance among the 1000 maps. Most of the maps belonged to one of three clusters designated as I, II and III. For each class, after the *k*-means clustering, we calculated the averaged electron density map and support shape (Fig. 4*c*) together with the parameters used to assess the quality of the class (Table 1).

Approximately half of the 1000 maps belonged to classes 1 and 2 of cluster I [Fig. 4(*b*) and Table 1]. Within each class, the retrieved maps and support shapes were very similar. The electron density maps of ten gold colloidal particles were well separated (Fig. 4*c*). The averaged electron density maps were closely similar between the two classes as well as the averaged support shapes, which exhibit sharp edges. Among all of the classes, the *R*-factors and effective resolutions of classes 1 and 2 were the first and second best, respectively (Table 1).

The averaged electron density maps of clusters II or III look like gold colloidal particles fused (Fig. 4*c*), similar to the map in the middle of Fig. 1(*c*). Both the *R*-factors and effective resolutions were worse than those in cluster I (Table 1). Class 10, displaying large zero-angle diffraction intensity, was an assembly of maps obtained from failed SW calculations [Fig. 4(*c*) and Table 1]. The averaged electron density map spread extremely widely and was poor with regard to the *R*-factor and effective resolution.

Considering the quality of the averaged electron density maps, support

Table 1
Calculated criteria values for the classes in Fig. 4.

Class	Number in class	*R*	$\bar{I}_0$ (photons)	Estimated resolution (nm)
1	285	0.287	5.37 × 10^6	91.3
2	220	0.300	5.88 × 10^6	93.7
3	21	0.328	7.60 × 10^6	104.4
4	111	0.390	1.22 × 10^7	202.9
5	63	0.422	1.48 × 10^7	202.9
6	61	0.431	1.28 × 10^7	202.9
7	25	0.447	1.09 × 10^7	202.9
8	13	0.405	1.33 × 10^7	192.2
9	194	0.423	1.03 × 10^7	192.2
10	7	0.522	8.61 × 10^7	1826.3

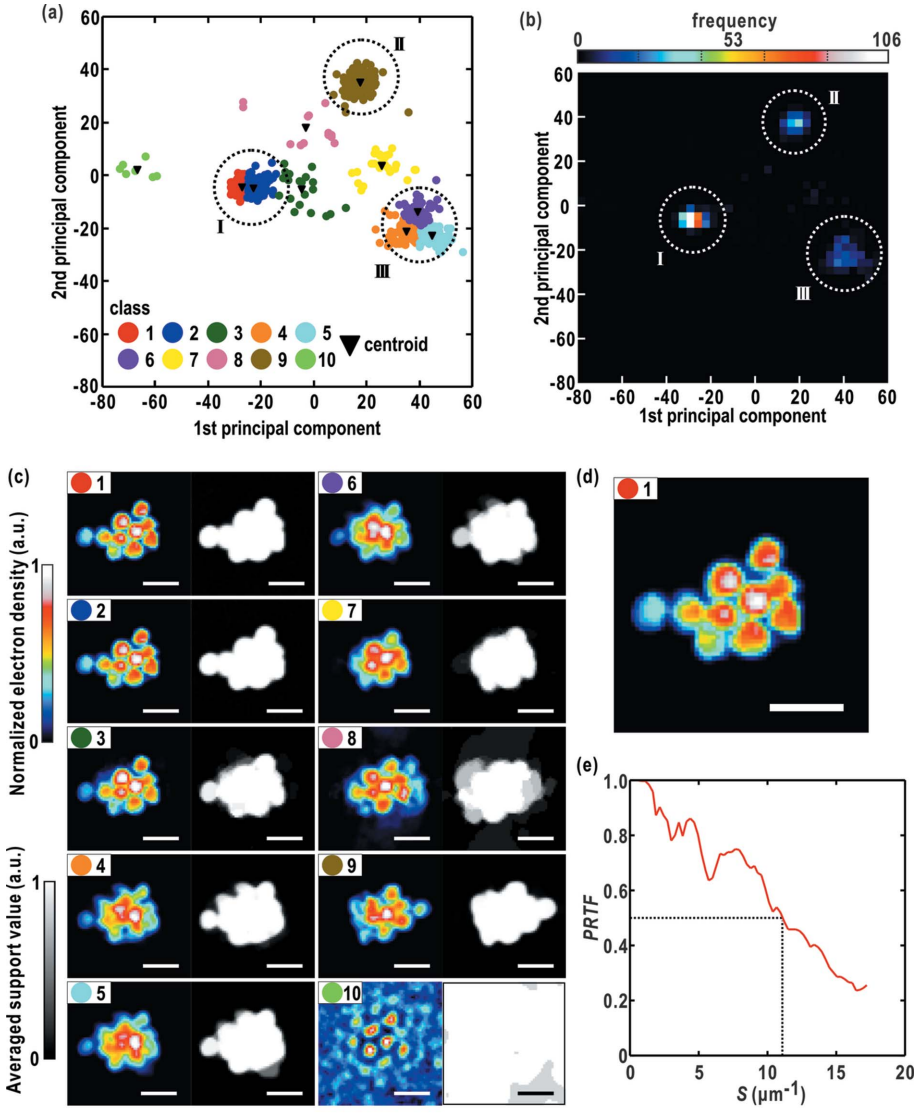


Figure 4
Results of the first stage of estimating the most probable overall support shape of an aggregate of gold colloidal particles from the diffraction pattern in Fig. 3(*a*). (*a*) A projection of 1000 electron density maps in the 3600-dimensional space onto the plane spanned by the first two PCs determined in PCA. The positions of maps are indicated by symbols colored according to the classes determined by the *k*-means clustering. (*b*) The distribution in panel (*a*) expressed as the frequency. (*c*) The averaged electron density map and averaged support shape of each class. The scale bars indicate 500 nm. The values of parameters for each class are compiled in Table 1. (*d*) A magnified view of the averaged electron density map of the selected class 1. (*e*) A PRTF curve calculated from 285 electron density maps belonging to class 1. The dotted lines are used to estimate the effective resolution of the averaged map.

shapes and parameters, we selected class 1 as the most probable support shape in the first stage [Figs. 4(*d*) and 4(*e*), Table 1).

4.2.2. Most probable electron density map with fine structures. The averaged support of class 1, binarized at a threshold level of 0.5, was used as the support in the subsequent OSS calculations. The PCA revealed that 673 of the 1000 OSS-retrieved maps formed two clusters designated as I and II in the plane spanned by the first and second PC vectors, which described 42% of the total variance [Figs. 5(*a*) and 5(*b*)]. Through the *k*-means clustering, classes *A*–*E* in the negative region of the first PC displayed electron density maps closely similar to each other with *R*-factors of 0.19–0.20 [Fig. 5(*c*) and Table 2]. The fine structures in averaged maps suggest that the maps were well classified and separated to avoid smearing by the contamination of other classes of maps.

As the most probable map representing classes *A*–*E*, we selected class *B* with the largest population of distribution and the smallest *R*-factor [Fig. 5(*d*) and Table 2]. The averaged electron density maps of some gold colloidal particles appear as triangular shapes as observed by EM (Nakasako *et al.*, 2013), whereas those in the first stage were approximated as circular shapes. The fine structures would contribute to the improvement of the effective resolution better than 60 nm in PRTF (Fig. 5*e*).

4.3. Structure analysis of a chloroplast

4.3.1. Overall shape. We prepared 1000 HIO-SW maps for the diffraction pattern with a SW parameter δ of 0.035. After PCA was performed, the maps in the 50 × 50-dimensional space were projected onto the plane spanned by the first and second PCs, which described 45% of the total variance among the 1000 maps. Most of the maps were distributed in three clusters, I, II and III [Figs. 6(*a*) and 6(*b*)]. We classified the maps further through *k*-means clustering.

At a glance, classes 1 and 2 forming cluster III seemed to be somewhat better than the other classes regarding their population density and the effective resolution [Fig. 6(*b*) and Table 3]. However, the averaged density map of each class, which comprised one prominent density peak accompanying weak peaks separated by 400 nm, was

Table 2
Calculated criteria values for the classes in Fig. 5.

Class	Number in class	*R*	$\bar{I}_0$ (photons)	Estimated resolution (nm)
A	30	0.195	9.46×10^6	57.4
B	196	0.189	9.60×10^6	57.4
C	124	0.200	9.56×10^6	57.4
D	86	0.203	9.45×10^6	57.4
E	237	0.191	9.56×10^6	57.4
F	84	0.192	9.19×10^6	57.4
G	60	0.231	9.34×10^6	64.1
H	61	0.183	9.02×10^6	57.4
I	54	0.253	8.67×10^6	84.9
J	68	0.216	7.79×10^6	60.9

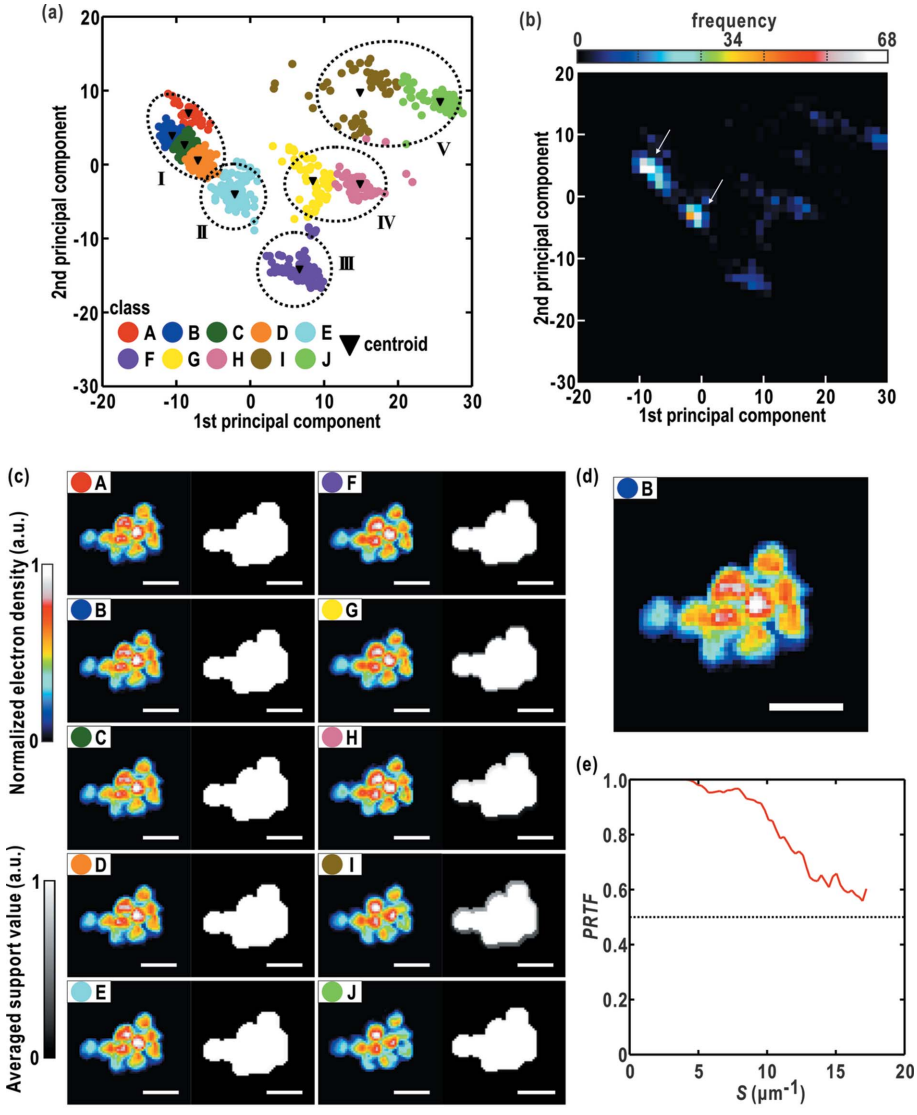


Figure 5
Results of the second stage of estimating the most probable electron density maps inside the support in Fig. 4(*d*). (*a*) A projection of 1000 maps in the 3600-dimensional space onto the plane spanned by the first two PCs determined in PCA. The positions of maps are indicated by symbols colored according to the classes from the *k*-means clustering. (*b*) The distribution in panel (*a*) expressed as the frequency. (*c*) The averaged electron density map and averaged support shape of each class. The scale bars indicate 500 nm. The values of parameters for each class are compiled in Table 2. (*d*) A magnified view of the averaged electron density map of the selected class *B*. (*e*) A PRTF curve calculated from 196 electron density maps belonging to class *B*. The dotted lines are used to estimate the effective resolution of the averaged map.

Table 3
Calculated criteria values for the classes in Fig. 6.

Class	Number in class	R	$\bar{I}_0$ (photons)	Estimated resolution (nm)
1	107	0.522	1.35×10^6	228.3
2	114	0.525	1.57×10^6	202.9
3	62	0.501	4.16×10^6	281.0
4	131	0.531	6.36×10^6	281.0
5	87	0.506	5.15×10^6	281.0
6	137	0.520	6.39×10^6	281.0
7	132	0.547	9.62×10^6	304.4
8	96	0.539	7.32×10^6	304.4
9	48	0.513	4.05×10^6	243.5
10	86	0.542	1.89×10^6	243.5

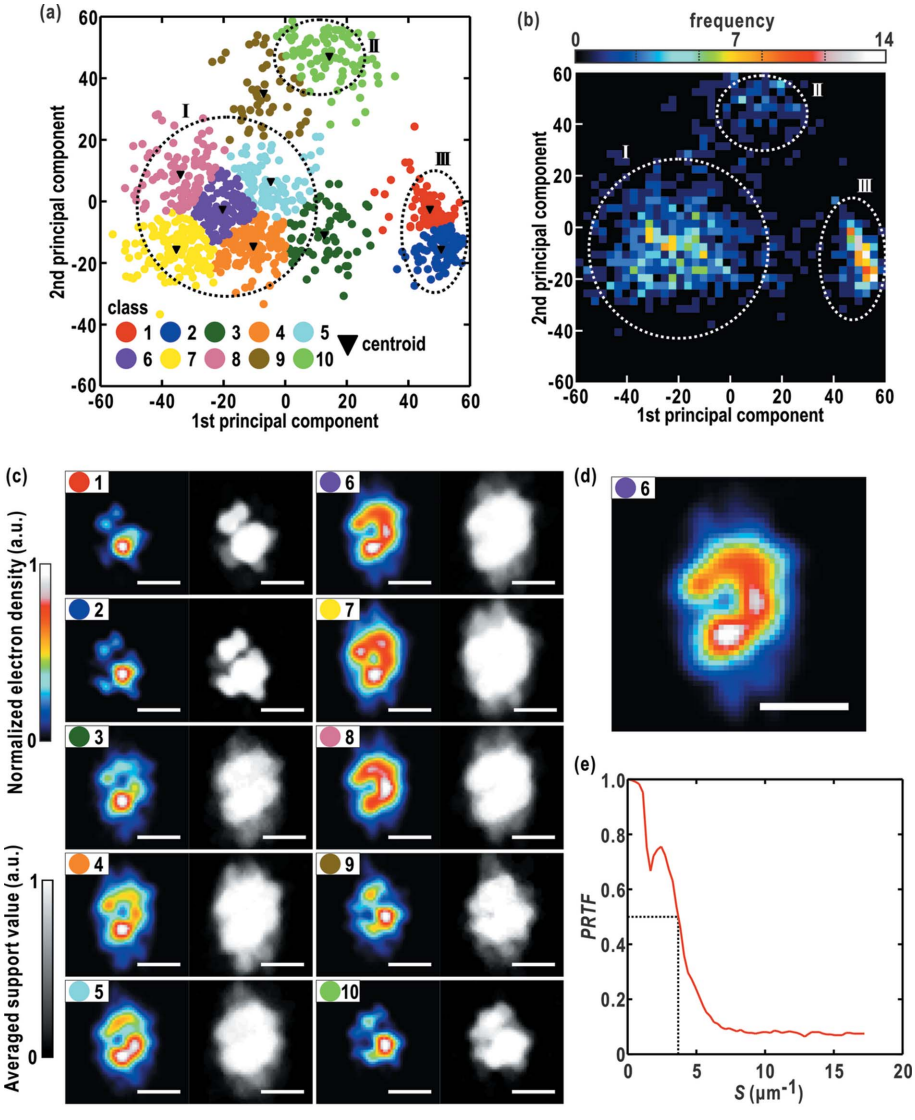


Figure 6
Result of the first stage of estimating the most probable overall support shape of a chloroplast from the diffraction pattern in Fig. 3(*b*). (*a*) A projection of 1000 electron density maps in the 2500-dimensional space onto the plane spanned by the first two PCs determined in PCA. The positions of maps are indicated by symbols colored according to the classes from the *k*-means clustering. (*b*) The distribution in panel (*a*) expressed as the frequency. (*c*) The averaged electron density map and averaged support shape of each class. The scale bars indicate 500 nm. The values of parameters for each class are compiled in Table 3. (*d*) A magnified view of the averaged electron density map of the selected class 6. (*e*) A PRTF curve calculated from 137 electron density maps belonging to class 6. The dotted lines are used to estimate the effective resolution of the averaged map.

inconsistent with the globular shape expected from the concentric interference pattern (Fig. 3*b*).

In X-ray crystallography, if atoms of the model structures are randomly distributed, the *R*-factor becomes 0.59 (Wilson, 1950). *R*-factors of each class were distributed in the range 0.50–0.55, close to 0.59. Class 3 displayed the lowest *R*-factor of 0.50. However, it is uncertain whether the slight difference in *R*-factors between class 3 and a random electron density map can be used as a significant criterion for selecting probable shapes.

Cluster I composed of classes 4–8 contained 583 maps out of 1000 HIO-SW maps, and widely spread in the negative region of the first PC (Fig. 6*b*). The maps had commonly globular C-shapes consistent with the low-resolution images of chloroplasts observed in fluorescence microscopy (Takayama *et al.*, 2015) (Fig. 6*c*). Here, we selected class 6 (Fig. 6*d*), which was the most populated and was located at the center of cluster I. The low effective resolution (Fig. 6*e*) suggested that the electron density distribution in the C-shape was a rough approximation of the internal structure of the chloroplast.

4.3.2. Most probable electron density map with fine structures. Under the constraint of the selected support of class 6 binarized at the threshold of 0.5, we retrieved 1000 electron density maps by using the OSS algorithm. Through applying the PCA to the maps, we found that the first and second PCs described 42% of the total variance of the maps. The maps projected onto the plane spanned by the two PCs were roughly separated into one major cluster (cluster I) and three minor clusters (clusters II–IV) [Fig. 7(*a*) and 7(*b*)]. After the *k*-means clustering, clusters I and IV located in the negative region of the first PC were characterized by maps with prominent peaks in the lower right part (Fig. 7*c*), while cluster III including maps with peaks in the upper left were located in the positive region. Because their averaged maps were inconsistent with the overall C-shape of class 6 (Fig. 6*d*), we rejected the three clusters.

The most populated cluster II composed of classes *A*–*C* displayed their averaged maps approximated as a C-shape (Fig. 7*c*). In contrast to the averaged electron density map of class 6 in the first stage, fine structures appeared in classes *A*–*C*. In particular,

class A composed of 227 maps was most populated, and displayed an *R*-factor of 0.36 (Table 4). Thus, we selected class A as the most probable projection structure of the chloroplast. Owing to the fine structures, the effective resolution (Fig. 7*e*) improved to 8 μm^{-1} (corresponding to 126 nm). Considering that undulations of the interference pattern in 5–15 μm^{-1} originated from the globular shape, the effective resolution may optimistically be better than 12 μm^{-1} (corresponding to 83 nm).

Class G displayed the lowest *R*-factor value of 0.32. However, the averaged electron density map of class G is different from the averaged map of the most probable class A and also from the low-resolution map of class 6. Only considering the conventionally used *R*-factor might be misleading.

Table 4
Calculated criteria values for the classes in Fig. 7.

Class	Number in class	*R*	$\bar{I}_0$ (photons)	Estimated resolution (nm)
A	227	0.363	5.24×10^6	126.0
B	99	0.388	5.44×10^6	126.0
C	69	0.427	4.78×10^6	152.2
D	75	0.476	3.90×10^6	173.9
E	62	0.436	3.72×10^6	182.6
F	40	0.401	5.72×10^6	146.1
G	63	0.322	5.91×10^6	79.4
H	80	0.503	4.28×10^6	243.5
I	137	0.456	4.82×10^6	158.8
J	148	0.399	4.79×10^6	166.0

5. Discussion

Here we have proposed a scheme to identify the most probable electron density maps in CXDI structure analyses. In the first stage of the scheme, the best support shape is first estimated, and then the second stage contributes to visualization of the fine structure inside the support. In each stage, the multivariate analysis helps us select the most probable class of retrieved maps without influence from incorrect maps as demonstrated in the structure analyses of gold colloidal particles and a bacterial chloroplast. In this section we discuss the benefits, limitations and characteristics of the proposed scheme.

5.1. Benefits and limitations of the proposed method

In the structure analyses for the two experimental diffraction patterns (Fig. 3), we exclude incorrect and/or less probable support shapes in the first stage (Figs. 4 and 6), and visualize the fine structures inside the supports to explain the diffraction patterns in the second stage (Figs. 5 and 7). The first stage yields a low-resolution structure, and then the second stage extends the phase to as high resolution as possible. In both of the two stages, the multivariate analysis is powerful in describing the characteristic distribution of maps in the multi-dimensional space.

In this study, we used the HIO-SW algorithm in the first stage and the OSS algorithm in the second stage. Other PR

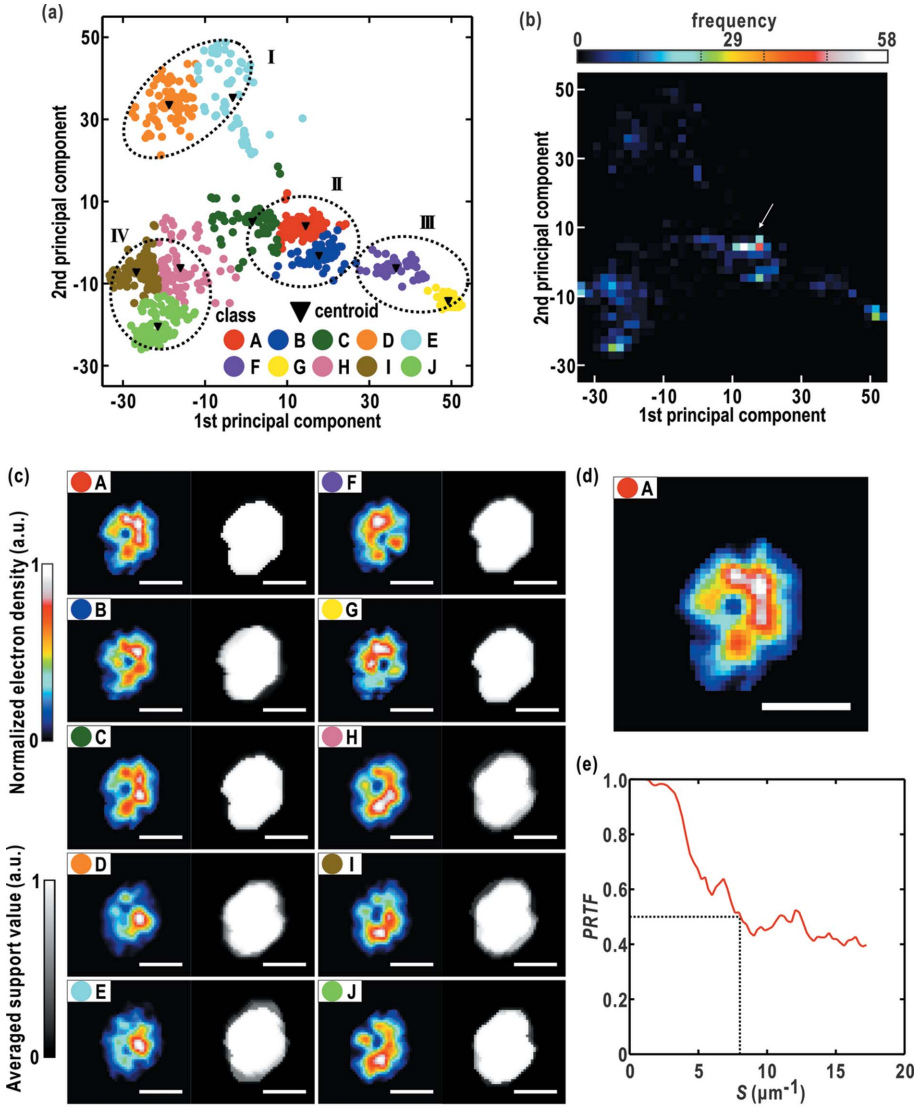


Figure 7
Result of the second stage of estimating the most probable electron density maps inside the support of Fig. 6(*d*). (*a*) A projection of 1000 electron density maps in the 2500-dimensional space onto the plane spanned by the first two PCs determined in PCA. The positions of maps are indicated by symbols colored according to the classes from the *k*-means clustering. (*b*) The distribution in panel (*a*) expressed as the frequency. (*c*) The averaged electron density map and averaged support shape of each class. The scale bars indicate 500 nm. The values of parameters for each class are compiled in Table 4. (*d*) A magnified view of the averaged electron density map of the selected class A. (*e*) A PRTF curve calculated from 227 electron density maps belonging to class A. The dotted lines are used to estimate the effective resolution of the averaged map.

algorithms or techniques can be easily implemented in each stage. For instance, more robust estimation of the support area may be possible by incorporating the dark-field PR techniques (Martin *et al.*, 2012; Kobayashi *et al.*, 2014) suitable for diffraction patterns missing data in the small-angle regions largely. However, it should be noted that the proposed scheme only contributes to selection of probable classes of maps and not to improvements of PR algorithms. Indeed, it is still difficult for the proposed scheme to retrieve reliable electron density maps from diffraction patterns with small oversampling ratio or low SNR.

The computational cost of the proposed scheme is somewhat heavy. To ensure statistical significance of structure analyses, the scheme requires more than 1000 PR trials for a diffraction pattern starting from different initial maps. Even with the parallelized software on the 576 CPU cores of the supercomputer we used, 1000 HIO-SW calculations and 1000 OSS calculations took about 15 and 25 min, respectively. The PCA and *k*-means clustering takes less than 1 min on a single CPU core. A larger number of PR calculations ensures the finer sampling of a larger area in the multi-dimensional space. Therefore, thoughtless reduction of the sampling points would cause a decline in the reliability of the analyses. Any idea to effectively reduce the number of PR calculations may be incorporated in the future development of the analysis scheme.

5.2. Characteristics in the distribution of maps on the plane spanned by a few PCs

The retrieved maps are non-uniformly and discretely distributed into a few clusters in planes spanned by the small number of PCs [Figs. 4(*a*), 5(*a*), 6(*a*) and 7(*a*)]. The discrete distributions are advantageous in selecting the most probable class of maps. Here we consider why discrete distributions occur in the multi-dimensional space.

The distributions of maps shown in Figs. 4–7 are similar to the population of protein structures in their energy landscapes (Moritsugu *et al.*, 2012). If retrieved maps are similar to the true map, maps probably distribute around the true map in a multi-dimensional space or in a space spanned by major PCs. Therefore, the landscape viewed using the similarity of a retrieved map to the true map has a basin of population around the position of the true map. In addition, a narrow distribution containing many maps with high consistency may be interpreted as the existence of a steep and narrow basin.

On the basis of this idea, we schematically illustrate the basins expected for clusters in Figs. 4(*a*) and 6(*a*). The maps in the clusters shown in Fig. 4(*a*) are thought to be distributed among three narrow basins (Fig. 8*a*). Comparing their population density, the basin inducing cluster I is the steepest among the three basins and is likely located closest to the true map. In the case of a chloroplast (Fig. 6*a*), three basins can be assumed (Fig. 8*b*). The width of cluster I may reflect the existence of a wide basin. The number of maps belonging to cluster I suggest that the true map is included in the basin.

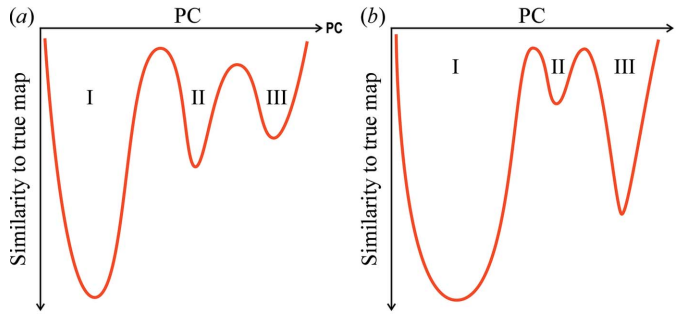


Figure 8
Schematic illustrations on the landscape regarding the similarity of maps to the true maps. The expected landscapes in (*a*) Fig. 4(*a*) and (*b*) Fig. 6(*a*).

The shape, size and depth in a landscape for retrieved maps in the multi-dimensional space probably depend on the SNR, oversampling ratio and the size of the small-angle area missing from the diffraction pattern. Non-uniform distribution of maps in the multi-dimensional space provides an opportunity to consider the landscape and the position of the true map. This idea may help us interpret the results from the multivariate analysis and provide ways to refine electron density maps.

5.3. Outlook

We are particularly interested in the visualization of internal structures inside cells and cellular organelles with complex and irregular shapes and low electron densities. In order to establish CXDI as a useful tool for structural analyses of biological specimens, the most probable electron density maps should be automatically and objectively proposed for given diffraction patterns without any prior information or reference images. Although the proposed scheme demonstrated the ability to present the most probable electron density map of a biological specimen, problems to be settled are the use of threshold δ for estimating the most probable support shape, the automated selection of the most probable class, and the number of classes assumed in the *k*-means clustering.

In the first stage, threshold δ yielding the most probable support shape depends on the type of specimens. From our experiences, a δ of 0.02–0.06 is suitable for metal particles with sharp edges and large electron densities. For biological specimens with low electron densities and irregular shapes, a δ of 0.01–0.05 tends to yield support shapes similar to those observed in other imaging techniques, with the size expected from the speckle. We empirically gave the optimal δ in the present study. However, it is better to treat δ as one of the variable parameters that is adjustable in every PR calculation, by inspecting the correlation between retrieved maps and δ. Thus, we are developing a PR calculation scheme incorporating this idea.

The second problem may be handled by introducing a comprehensive score. The score would comprise, for instance, parameterizations of each of the electron density maps, the sharpness of support shapes and the population in the

landscape discussed, in addition to the parameters listed in Tables 1–4. The weights of parameters are empirically tuned as is done in the *ab initio* determination of the molecular shape from small-angle scattering profiles of proteins (Svergun *et al.*, 2001), and the structure refinement under stereochemical restraints in protein crystallography (Hendrickson, 1985).

Several criteria for the determination of the number of classes in *k*-means clustering (Pham *et al.*, 2005) would help us solve the third problem. Furthermore, nonlinear dimensionality reduction, for instance, by using the diffusion map (Coifman *et al.*, 2005) may help us estimate the number of classes for retrieved maps in multi-dimensional space.

Acknowledgements

We selected representative diffraction data from cryogenic XFEL-CXDI experiments performed at SACLA (proposal Nos. 2013A8043, 2013B8049 and 2014A8033. The authors thank the members of the SACLA engineering team for their great help in the alignment and the operation of the focusing mirror optics, our diffractometer and the two detectors. The authors are also grateful to Professor Sachihiro Matsunaga and Dr Yayoi Inui for their preparation of the chloroplast samples. This study was supported by a grant for XFEL key technology and the X-ray Free Electron Laser Priority Strategy Program from the MEXT to MN, Grant-in-Aid for Scientific Research on Innovative Areas (Nos. 22244054, 25120725 to MN, and Nos. 24113723, 26104535 to TO), Grant-in-Aid for Young Scientists (B) (No. 26800227 to TO), Grant-in-Aid for Challenging Exploratory Research (No. 24654140 to MN) and Grant-in-Aid for JSPS Fellows (No. 15J01707 to YS) from the JSPS.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Chapman, H. N. *et al.* (2006*a*). *Nat. Phys.* **2**, 839–843.
Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006*b*). *J. Opt. Soc. Am. A*, **23**, 1179–1200.
Chapman, H. N., Caleman, C. & Timneanu, N. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130313.
Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F. & Zucker, S. W. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 7426–7431.
Drenth, J. (2007). *Principles of Protein X-ray Crystallography*. Berlin: Springer.
Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
Hantke, M. F. *et al.* (2014). *Nat. Photon.* **8**, 943–949.
Heel, M. van & Frank, J. (1981). *Ultramicroscopy*, **6**, 187–194.
Hendrickson, W. (1985). *Methods Enzymol.* **115**, 252–270.
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.
Joti, Y., Kameshima, T., Yamaga, M., Sugimoto, T., Okada, K., Abe, T., Furukawa, Y., Ohata, T., Tanaka, R., Hatsui, T. & Yabashi, M. (2015). *J. Synchrotron Rad.* **22**, 571–576.
Kameshima, T., Ono, S., Kudo, T., Ozaki, K., Kirihara, Y., Kobayashi, K., Inubushi, Y., Yabashi, M., Horigome, T., Holland, A., Holland, K., Burt, D., Murao, H. & Hatsui, T. (2014). *Rev. Sci. Instrum.* **85**, 033110.
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.
Kobayashi, A., Sekiguchi, Y., Takayama, Y., Oroguchi, T. & Nakasako, M. (2014). *Opt. Express*, **22**, 27892–27909.
Kodama, W. & Nakasako, M. (2011). *Phys. Rev. E*, **84**, 021902.
Loh, N. D. *et al.* (2012). *Nature (London)*, **486**, 513–517.
Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
MacQueen, J. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281–297.
Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., Weierstall, U. & Spence, J. C. H. (2003). *Phys. Rev. B*, **68**, 140101.
Martin, A. V. *et al.* (2012). *Opt. Express*, **20**, 13501–13512.
Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.
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.
Miao, J., Ishikawa, T., Anderson, E. H. & Hodgson, K. O. (2003). *Phys. Rev. B*, **67**, 174104.
Miao, J., Ishikawa, T., Shen, Q. & Earnest, T. (2008). *Annu. Rev. Phys. Chem.* **59**, 387–410.
Moritsugu, K., Terada, T. & Kidera, A. (2012). *J. Am. Chem. Soc.* **134**, 7094–7101.
Nakasako, M. *et al.* (2013). *Rev. Sci. Instrum.* **84**, 093705.
Nam, D., Park, J., Gallagher-Jones, M., Kim, S., Kim, S., Kohmura, Y., Naitow, H., Kunishima, N., Yoshida, T., Ishikawa, T. & Song, C. (2013). *Phys. Rev. Lett.* **110**, 098103.
Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.
Nishino, Y., Miao, J. & Ishikawa, T. (2003). *Phys. Rev. B*, **68**, 220101.
Nishino, Y., Takahashi, Y., Imamoto, N., Ishikawa, T. & Maeshima, K. (2009). *Phys. Rev. Lett.* **102**, 018101.
Oroguchi, T. & Nakasako, M. (2013). *Phys. Rev. E*, **87**, 022712.
Park, H. J. *et al.* (2013). *Opt. Express*, **21**, 28729–28742.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Pham, D. T., Dimov, S. S. & Nguyen, C. D. (2005). *Mech. Eng. Sci.* **219**, 103–119.
Rodriguez, J. A., Xu, R., Chen, C.-C., Zou, Y. & Miao, J. (2013). *J. Appl. Cryst.* **46**, 312–318.
Rosenthal, P. B. & Henderson, R. (2003). *J. Mol. Biol.* **333**, 721–745.
Schot, G. van der *et al.* (2015). *Nat. Commun.* **6**, 5704.
Seibert, M. M. *et al.* (2011). *Nature (London)*, **470**, 78–81.
Sekiguchi, Y., Oroguchi, T., Takayama, Y. & Nakasako, M. (2014*a*). *J. Synchrotron Rad.* **21**, 600–612.
Sekiguchi, Y., Yamamoto, M., Oroguchi, T., Takayama, Y., Suzuki, S. & Nakasako, M. (2014*b*). *J. Synchrotron Rad.* **21**, 1378–1383.
Shapiro, D., Thibault, P., Beetz, T., Elser, V., Howells, M., Jacobsen, C., Kirz, J., Lima, E., Miao, H., Neiman, A. M. & Sayre, D. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 15343–15346.
Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
Takayama, Y., Inui, Y., Sekiguchi, Y., Kobayashi, A., Oroguchi, T., Yamamoto, M., Matsunaga, S. & Nakasako, M. (2015). *Plant Cell Physiol.* **56**, 1272–1286.
Takayama, Y. & Nakasako, M. (2012). *Rev. Sci. Instrum.* **83**, 054301.
Tono, K., Togashi, T., Inubushi, Y., Sato, T., Katayama, T., Ogawa, K., Ohashi, H., Kimura, H., Takahashi, S., Takeshita, K., Tomizawa, H., Goto, S., Ishikawa, T. & Yabashi, M. (2013). *New J. Phys.* **15**, 083035.
Williams, G. J., Pfeifer, M. A., Vartanyants, I. A. & Robinson, I. K. (2003). *Phys. Rev. Lett.* **90**, 175501.
Wilson, A. J. C. (1950). *Acta Cryst.* **3**, 397–398.
Xu, R. *et al.* (2014). *Nat. Commun.* **5**, 4061.