



A protocol for searching the most probable phase-retrieved maps in coherent X-ray diffraction imaging by exploiting the relationship between convergence of the retrieved phase and success of calculation

Yuki Sekiguchi,^{a,b} Saki Hashimoto,^{a,b} Amane Kobayashi,^{a,b} Tomotaka Oroguchi,^{a,b} and Masayoshi Nakasako^{a,b*}

Received 25 January 2017

Accepted 6 June 2017

Edited by Y. Amemiya, University of Tokyo, Japan

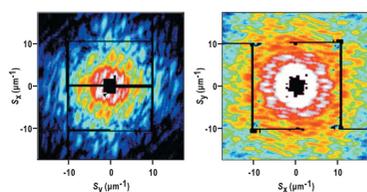
Keywords: coherent X-ray diffraction imaging; X-ray free-electron laser; structure analysis of non-crystalline particles; phase retrieval calculation.

^aDepartment of Physics, Faculty of Science and Technology, Keio University, Hiyoshi 3-14-1, Yokohama, Kohoku, Kanagawa 223-8522, Japan, and ^bRIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Sayo-gun, Hyogo 679-5148, Japan.
*Correspondence e-mail: nakasako@phys.keio.ac.jp

Coherent X-ray diffraction imaging (CXDI) is a technique for visualizing the structures of non-crystalline particles with size in the submicrometer to micrometer range in material sciences and biology. In the structural analysis of CXDI, the electron density map of a specimen particle projected along the direction of the incident X-rays can be reconstructed only from the diffraction pattern by using phase-retrieval (PR) algorithms. However, in practice, the reconstruction, relying entirely on the computational procedure, sometimes fails because diffraction patterns miss the data in small-angle regions owing to the beam stop and saturation of the detector pixels, and are modified by Poisson noise in X-ray detection. To date, X-ray free-electron lasers have allowed us to collect a large number of diffraction patterns within a short period of time. Therefore, the reconstruction of correct electron density maps is the bottleneck for efficiently conducting structure analyses of non-crystalline particles. To automatically address the correctness of retrieved electron density maps, a data analysis protocol to extract the most probable electron density maps from a set of maps retrieved from 1000 different random seeds for a single diffraction pattern is proposed. Through monitoring the variations of the phase values during PR calculations, the tendency for the PR calculations to succeed when the retrieved phase sets converged on a certain value was found. On the other hand, if the phase set was in persistent variation, the PR calculation tended to fail to yield the correct electron density map. To quantify this tendency, here a figure of merit for the variation of the phase values during PR calculation is introduced. In addition, a PR protocol to evaluate the similarity between a map of the highest figure of merit and other independently reconstructed maps is proposed. The protocol is implemented and practically examined in the structure analyses for diffraction patterns from aggregates of gold colloidal particles. Furthermore, the feasibility of the protocol in the structure analysis of organelles from biological cells is examined.

1. Introduction

Coherent X-ray diffraction imaging (CXDI) is a lens-less imaging technique for visualizing the structures of non-crystalline particles with dimensions in the submicrometer to micrometer range at resolutions of several tens of nanometers (Miao *et al.*, 2015). In CXDI experiments, a spatially isolated non-crystalline specimen particle is illuminated by a coherent X-ray beam, and the Fraunhofer diffraction pattern is recorded with a sufficiently high sampling frequency to satisfy the oversampling condition (Miao *et al.*, 2003a). The electron



density map of the specimen particle projected along the direction of the incident X-ray beam is reconstructed by applying the phase-retrieval (PR) algorithms (Fienup, 1982) to the oversampled diffraction 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 material sciences 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).

Recently, CXDI experiments utilizing X-ray free-electron lasers (XFELs) have been used to perform structure analyses of non-crystalline particles (Seibert *et al.*, 2011; Loh *et al.*, 2012; Nakasako *et al.*, 2013; Hantke *et al.*, 2014; Xu *et al.*, 2014; Kimura *et al.*, 2014; van der Schot *et al.*, 2015; Ekeberg *et al.*, 2015). Diffraction patterns are collected at the repetition rate of the XFEL pulses with fresh specimens being delivered into the irradiation area pulse by pulse. For instance, our diffraction apparatus TAKASAGO-6, which can move the frozen-hydrated or dry specimens on thin film at a speed of 25 μm per 33 ms, provides more than 35000 diffraction patterns within 1 h at SACLA (Kobayashi *et al.*, 2016a), where XFEL pulses are supplied at a repetition rate of 30 Hz.

Diffraction patterns lose the phase information necessary to reconstruct the electron density map of the specimen by the inverse Fourier transform. In X-ray protein crystallography, for instance, the phase of a diffracted wave is experimentally estimated by measuring the changes in diffraction intensities caused by the heavy-atom labeling of protein molecules (Blow & Crick, 1959). In contrast, the phase values of a diffraction pattern in CXDI are estimated by an entirely computational procedure executed by a large number of PR calculation cycles (for instance, 10000 cycles) under the real-space and reciprocal-space constraints (Fienup, 1982).

In experimental diffraction patterns, small-angle regions, where structural information on the overall shape and total electrons of specimen particles are contained, are missed due to the beam stop and saturation of the detector pixels. In addition, Poisson noise in X-ray detection modifies the diffraction patterns, particularly in high-angle regions. These factors often make it difficult to efficiently obtain correct maps as demonstrated in our previous simulation studies (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013; Kobayashi *et al.*, 2014; Takayama *et al.*, 2015a; Yoshidome *et al.*, 2015) and structure analyses (Takayama *et al.*, 2015b; Oroguchi *et al.*, 2015; Sekiguchi *et al.*, 2016; Kobayashi *et al.*, 2016b). As a typical example, we show the results of PR calculations for a diffraction pattern from an aggregate of ten 250 nm gold colloidal particles [Figs. 1(a) and 1(b)]. Successful PR calculations provide maps displaying clear and well separated images of ten gold colloidal particles, while ill-defined particle images appear in maps from failed calculations. Among 1000 PR runs starting from different random electron density maps,

the number of successful and failed runs were 487 and 513, respectively.

In structure determination, phases of diffraction waves are more important than amplitudes to describe electron density maps (Taylor, 2003). However, in CXDI, little attention is paid to the variation of phases in the PR calculations. We speculated that phase values in successful runs would change differently from those in failed runs from our experiences in simulations and structure analyses (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013; Kobayashi *et al.*, 2014, 2016b; Takayama *et al.*, 2015a,b; Oroguchi *et al.*, 2015; Yoshidome *et al.*, 2015; Sekiguchi *et al.*, 2016). In fact, as demonstrated in Fig. 1(c), the phase values in the small-angle regions of $S < 10 \mu\text{m}^{-1}$ converged and remained around certain values in the early stage of the PR cycles, while phase values persistently varied in failed runs. This point is visualized by the frequency distribution of phase values in the PR cycles [Fig. 1(d)]. These tendencies regarding the variation of phase values suggest a clue to identify successful runs in PR calculations.

In this study, we propose a protocol for identifying successful runs by monitoring the phase values. The key features of the protocol are (i) parameterization regarding the variations of the phase values in PR cycles, and (ii) use of scores measuring the similarity of maps produced by a number of PR runs. Here we describe the details of the proposed protocol and its practical application to experimental diffraction patterns from aggregates of gold colloidal particles and organelles of biological cells.

2. Calculation methods

In this section, we describe the calculation methods used in the proposed protocol. To concretely explain the details of the calculation methods, we illustrate the process of structure analysis for the diffraction pattern shown in Fig. 1(a).

2.1. PR calculations

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. The algorithms were implemented in a program suite, *ZOCHO*, in our previous simulation studies (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013). We perform 1000 PR runs with different random initial maps for each single diffraction pattern. The PR runs in which the support shapes did not converge are not used in the subsequent analysis.

2.2. Quantification of variation of phase values during PR calculation

During a PR run, the frequency distribution of the phase values is determined for each pixel in the diffraction pattern by recording maps in every PR cycle. Because retrieved maps lose information on absolute translational positions, maps are

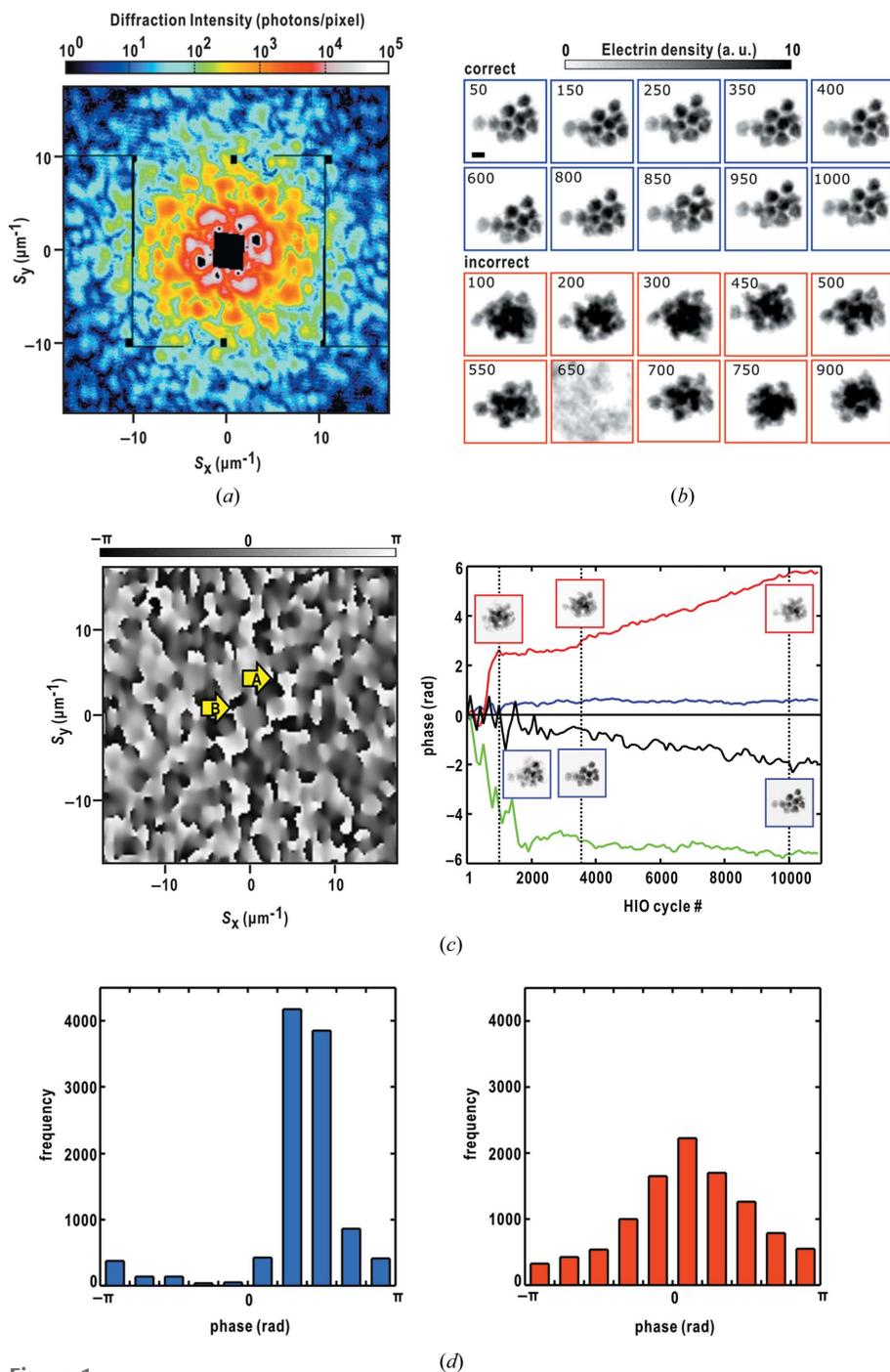


Figure 1
 (a) Diffraction pattern from an aggregate of ten 250 nm gold colloidal particles recorded using a single-shot XFEL pulse. The magnitude of scattering vector \mathbf{S} is defined as $|\mathbf{S}| = 2 \sin \theta / \lambda$, where 2θ and λ are the scattering angle and the wavelength of the X-rays, respectively. Then, the resolution at the edge is $17.4 \mu\text{m}^{-1}$ (corresponding to a resolution of 57.4 nm in the real space). Starting from electron density maps of different random noise, 1000 projection maps are retrieved by the HIO-SW algorithm. (b) Because the shape and size of each gold colloidal particle is known, we can determine which PR runs succeeded or failed. The blue-colored boxes indicate the maps of successful runs, while maps of failed runs are indicated by red boxes. The scale bar indicates 200 nm. (c) Variation of phase values in the course of PR calculations for the diffraction pattern shown in (a). The left-hand panel shows the phase values in the diffraction pattern at the 3000th PR cycle of a successful run. The right-hand panel shows the variation of phase values at pixels A (blue line for the successful run and red line for the failed run) and B (green line for the successful run and black line for the failed run). The maps appearing at the 1000th, 3500th and 10000th cycles in the runs are shown for the successful (blue box) and failed runs (red box). (d) Frequency distributions of phase values at pixel A during the 11000 cycles of the PR calculation are shown for the successful (left panel) and failed (right panel) runs.

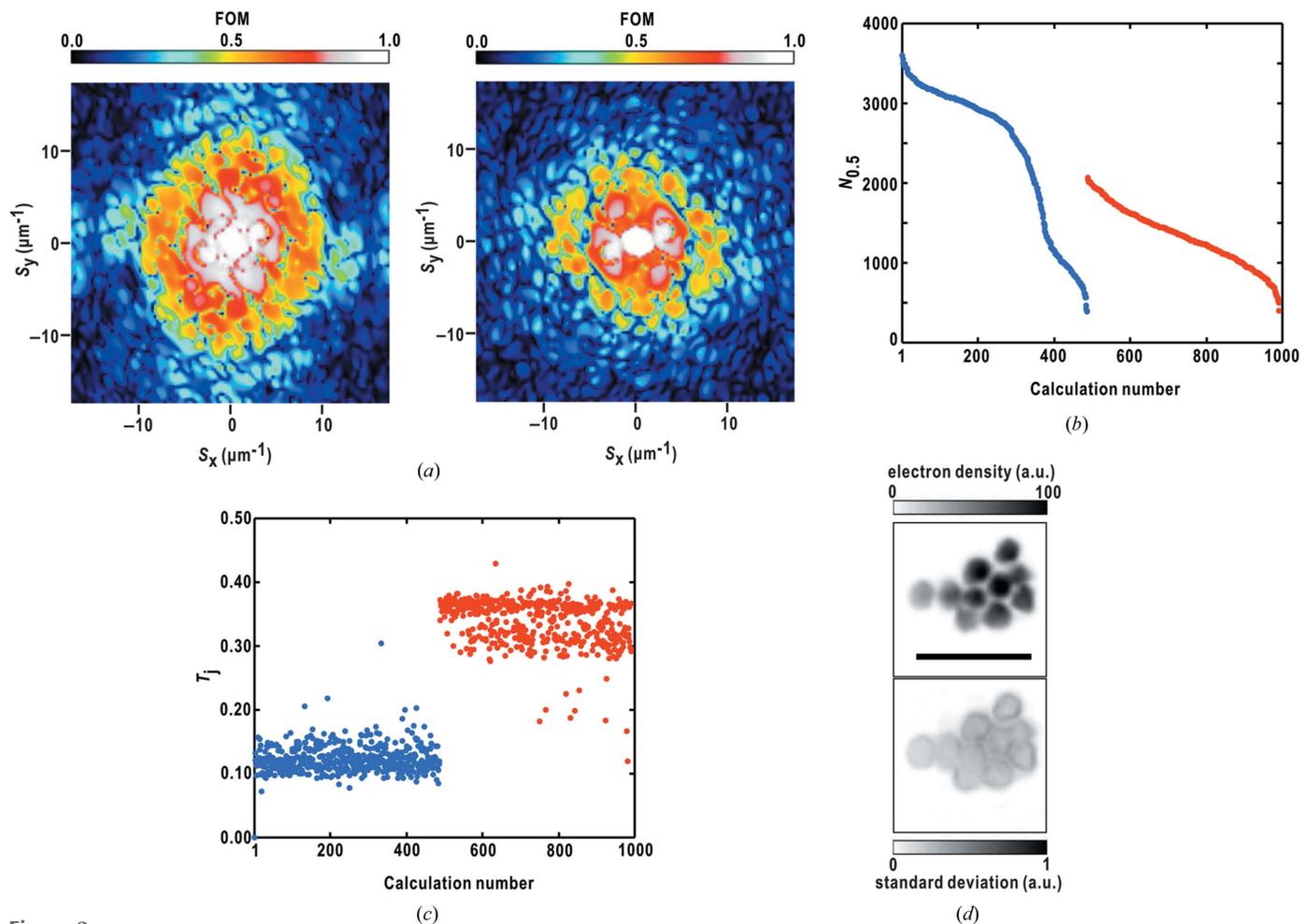
superimposed with regard to their centroids prior to calculating the phase values.

To quantitatively parameterize the frequency distribution, we introduce a figure of merit (FOM) for the frequency distribution of the phase values in a pixel at scattering vector \mathbf{S} ($|\mathbf{S}| = 2 \sin \theta / \lambda$, where 2θ and λ are the scattering angle and wavelength of incident X-rays, respectively) as

$$\text{FOM}(k, \mathbf{S}) = \frac{\left| \sum_k P(k, \mathbf{S}) \exp(i\alpha_k) \right|}{\sum_k P(k, \mathbf{S})}, \quad (1)$$

where $P(k, \mathbf{S})$ is the frequency of phase values in the k th bin at \mathbf{S} . We used a bin width of 0.2π rad suitable for describing frequency distributions of phase angles among a set of 1000 trial calculations [Fig. 1(d)]. α_k is the phase value at the center of the k th bin. The FOM values tend to be high in small-angle regions, and gradually decrease at high diffraction angles. As an example, the variations of the FOM values between the successful and failed runs are compared as shown in Fig. 2(a). In the successful run, the number of pixels in the diffraction pattern displaying high FOM values tended to be larger than those in the failed run.

In this study, we tentatively set a threshold for the FOM of 0.5. The number of pixels with FOM values larger than the threshold value, designated as $N_{0.5}$, is then counted for the diffraction pattern in each PR run [Fig. 2(b)]. For the diffraction pattern shown in Fig. 1(a), the successful runs gave an $N_{0.5}$ value of 300–3500. In most of the failed runs, a real-space constraint in the PR calculation sometimes works as an operator to convert incorrect maps into different incorrect maps. As a result, the phase values in PR cycles of failed runs would distribute over a wide range and then the broad frequency distributions give small $N_{0.5}$. When the phase values drop near the correct values in the final stage of the PR cycles, the frequency distributions of the phase values become broad and give a small FOM. The $N_{0.5}$ values of failed runs are less than 2000.


Figure 2

(a) FOM values of pixels in the diffraction pattern shown in Fig. 1(a) for successful (left panel) and failed (right) runs. (b) $N_{0.5}$ values for 487 successful runs (blue dots) and 503 failed runs (red). The calculation numbers are given in the order of the $N_{0.5}$ values within correct or incorrect maps. (c) T_j values of PR maps against the best map. The values for maps from successful and failed runs are indicated by blue and red dots, respectively. (d) The upper panel is a map averaged from those with T_j values smaller than 0.2, and the lower panel is the standard deviation from the average.

2.3. Similarity score

The PR run displaying the largest $N_{0.5}$ score is assumed to succeed in producing a map (the best map) that would be closely similar to the true one. Subsequently, runs that give maps similar to the best one can also be assigned as successful runs. Several CXDI studies postulated that frequently appearing maps in a large number of independent PR runs are usually more probable (Park *et al.*, 2013; Kimura *et al.*, 2014; van der Schot *et al.*, 2015; Sekiguchi *et al.*, 2016). Therefore, to find maps of successful runs and validate whether the best map is correct, we use a parameter to measure the similarity of a map to the best map defined as (Miao *et al.*, 2003b)

$$T_j = \frac{\sum_{x,y} |\rho_1(x,y) - \rho_j(x,y)|}{\sum_{x,y} |\rho_1(x,y) + \rho_j(x,y)|}, \quad (2)$$

where $\rho_1(x,y)$ is the electron density value of the best map and $\rho_j(x,y)$ is the electron density value of the other map. When the two maps become more similar, the T_j score tends to be close to 0.

Fig. 2(c) shows the T_j scores of 999 maps against the best map. The T_j values of the maps that resemble the best map are smaller than 0.2. In most cases, the maps showing T_j values larger than 0.25 have different shapes from the best map. Therefore, we set the threshold value of T_j to 0.2 for extracting the correct maps.

Finally, we calculate the average of the electron density maps with the T_j values smaller than 0.2 [Fig. 2(d)]. The averaged map is composed of the electron densities of ten gold colloidal particles with a standard deviation of less than 0.5% from the average. This result suggests that $N_{0.5}$ and T_j would be used as indicators for extracting correct electron density maps from a set of PR maps.

2.4. Protocol to extract correct electron density maps

Based on the analysis shown in Fig. 2, we propose a protocol for extracting maps from successful runs (Fig. 3). Firstly, 1000 PR calculations for a diffraction pattern are carried out starting from different random density maps. In each PR calculation run, the phase values calculated in each PR cycle

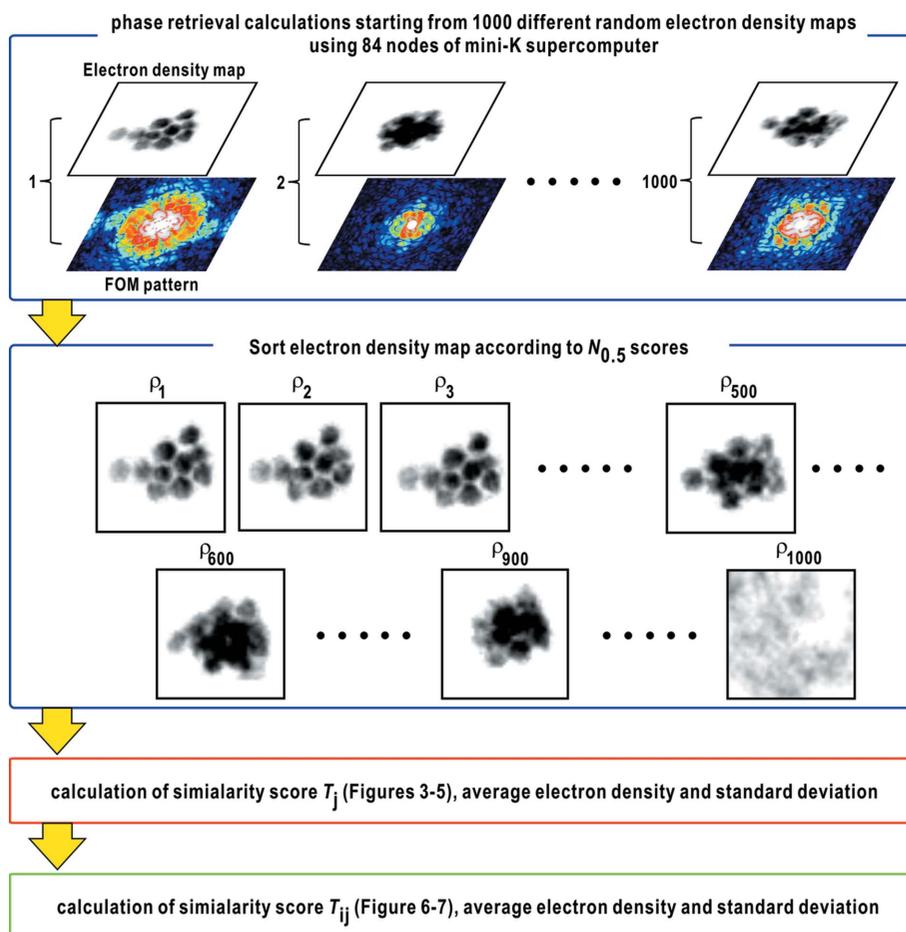


Figure 3
Flowchart used in this study to extract correct electron density maps from 1000 PR calculation runs.

are stored to construct the frequency distribution of the phase values for pixels in the region of interest ($S < 20 \mu\text{m}^{-1}$) in the diffraction pattern. After calculating the FOM values in the pixels of the diffraction pattern, the best map with the largest $N_{0.5}$ is used as the reference in calculating the T_j scores. Subsequently, maps with T_j values smaller than 0.2 are extracted as correct maps. The threshold value for the T_j score is set through the application of this protocol to several diffraction patterns in this study (see the *Results* section). Finally, the average electron density map and the standard deviation from the average are calculated.

3. Experimental procedure

3.1. XFEL-CXDI experiment and data processing

We performed CXDI experiments using our custom-made diffractometer KOTOBUKI-1 (Nakasako *et al.*, 2013) or TAKASAGO-6 (Kobayashi *et al.*, 2016a) at BL3 (Tono *et al.*, 2013) of the XFEL facility SACLA. Either diffraction apparatus was placed so that the specimen position of the apparatus was in the focus spot of the XFEL pulses focused by Kirkpatrick–Baez mirror system (Yumoto *et al.*, 2013). The intensity and duration of the X-ray pulses with an energy

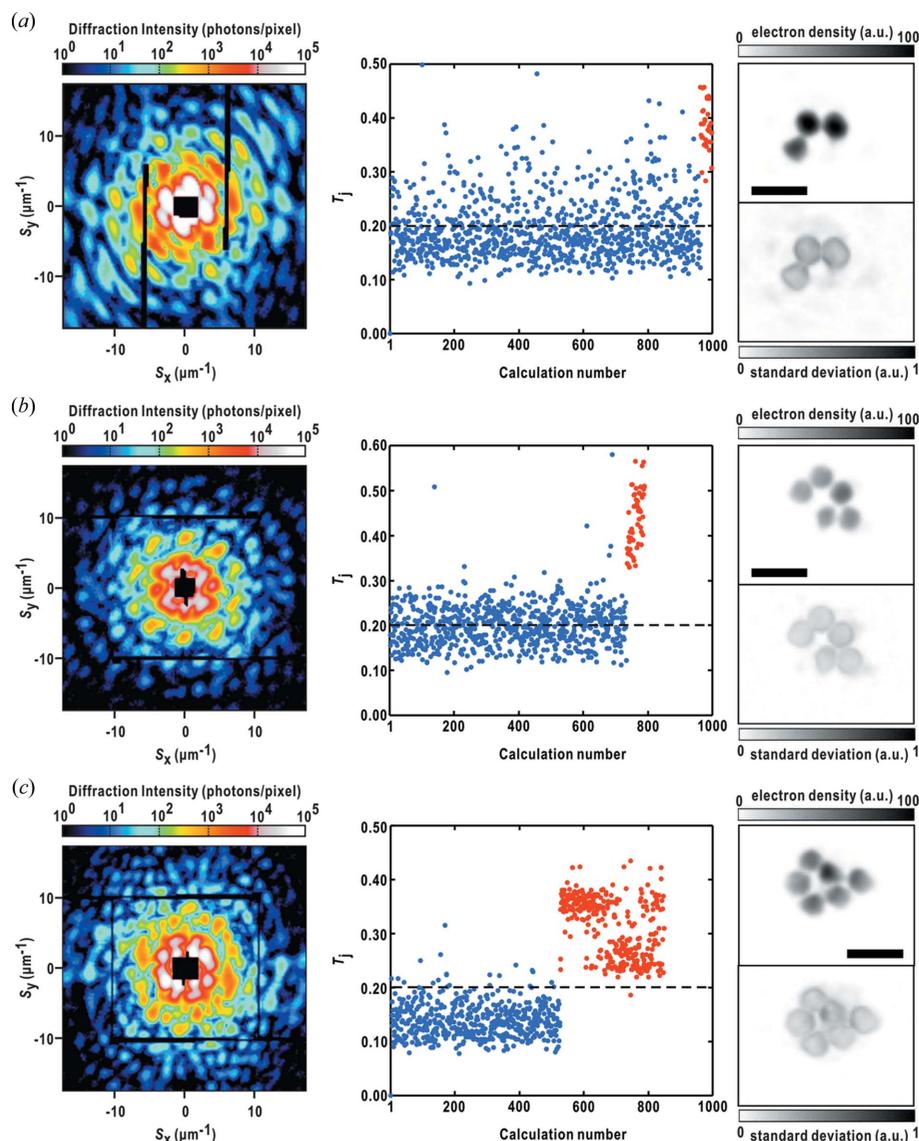
of 5.5 keV were approximately 10^{10} – 10^{11} photons μm^{-2} pulse $^{-1}$ and 10 fs, respectively. A specimen holder fixing a silicon nitride membrane was scanned against incident X-ray pulses. In the diffraction data collection using the KOTOBUKI-1 diffraction apparatus, X-ray pulses were extracted at a repetition rate of 1 Hz by the pulse-selector device installed on the beamline. In contrast, the TAKASAGO-6 apparatus allows us to collect diffraction patterns at a repetition rate of 30 Hz (Kobayashi *et al.*, 2016a).

The method for the preparation of specimens has been reported previously (Kobayashi *et al.*, 2016b). Diffraction patterns were recorded by using the multi-port CCD (MPCCD) Octal and the Dual detectors (Kameshima *et al.*, 2014) in a tandem arrangement. The camera distances of the MPCCD Octal and Dual detectors were approximately 1.6 m and 3.2 m downstream from the specimen position, respectively. The central aperture of the MPCCD Octal detector was changed depending on the diffraction intensity. Aluminium foils were placed in front of the Dual detector to attenuate strong diffraction patterns in the small-angle region.

The *G-SITENNO* program suite (Sekiguchi *et al.*, 2014a,b) was used for data processing. The suite first subtracts the background noise of the detectors and merges the diffraction patterns recorded by the two MPCCD detectors. Each diffraction pattern was binned by summing 2×2 pixel arrays into one pixel. By inspecting the montage, a graphical summary of the diffraction patterns worth analyzing, the diffraction patterns used for this study were selected. All data processing by the *G-SITENNO* suite was performed on a high-performance supercomputer, composed of 960 cores of Intel Xeon CPU X5690 (3.47 GHz per core), at SACLA (Joti *et al.*, 2015). The PR calculations were carried out on the mini-K supercomputer (Joti *et al.*, 2015).

4. Results

Here we examined the practical feasibility of the proposed protocol for experimental diffraction patterns with a variety of shapes and sizes of specimens, diffraction intensity, Poisson noise, oversampling (OS) ratios and the sizes of missing small-angle regions. We conducted 1000 PR runs for each diffraction pattern from an aggregate of gold colloidal particles, and determined which runs were successful or failed by inspecting the size, shape and edges of particles in PR maps (Figs. 4 and 5). As demonstrated in Fig. 1(b), maps from successful PR


Figure 4

Application of the T_j analysis to diffraction patterns from various aggregates of compact arranged gold colloidal particles. Each panel shows the diffraction pattern (left), the T_j values of 1000 PR maps (middle) and a set of the averaged map and the standard deviation map (right). T_j values from successful and failed runs are indicated by blue and red dots, respectively. The average map is calculated from maps with T_j values smaller than 0.2. The scale bar indicates 600 nm. The statistics of diffraction patterns and phase-retrieved electron density maps in panels (a)–(f) are summarized in Table 1.

runs displayed clear and well defined particle images, and were distinguished from those of failed runs with ill-defined particle images. By inspecting the variation of similarity scores of PR maps, we defined a threshold value of the similarity score for discriminating safely between successful and failed runs. Then, the protocol and the threshold value were examined further by the application to the PR calculation for other diffraction patterns from specimens with complicated structures (Figs. 6 and 7).

4.1. Feasibility of the proposed protocol

We examined the practical feasibility of the proposed protocol through the application to two types of diffraction

patterns from aggregates of gold colloidal particles (Figs. 4, 5 and Table 1). The first type is a set of diffraction patterns from compactly packed aggregates (Fig. 4). The second type is a set of diffraction patterns from aggregates with larger dimensions, *i.e.* smaller OS ratios, than the first type (Fig. 5).

Figs. 4(a), 4(b) and 4(c) show examples of PR calculations for the diffraction patterns from aggregates of three, five and six gold colloidal particles, respectively. The diffraction patterns were recorded without the saturation of detector pixels around the beam stop. Failed PR runs gave maps composed of densities of particles with unclear edges or lacking particle images. Successful PR runs gave maps composed of particles separated clearly. The T_j scores were different between the successful and failed runs. The T_j scores were predominantly in the range 0.1–0.3 for the maps from the successful runs, while those of the failed runs were 0.2–0.5. For automatically extracting maps of successful runs, the threshold of T_j scores is set at 0.2 to discriminate safely between successful and failed runs. When using the threshold value, the averaged electron density maps displayed clear densities of gold colloidal particles, and then the standard deviations were less than 0.5% of the maximum density value.

Figs. 4(d), 4(e) and 4(f) show PR calculations for diffraction patterns from aggregates of three, seven and six gold colloidal particles, respectively. In contrast to the diffraction patterns in Figs. 4(a)–4(c), the patterns missed small-angle regions due to the saturation of detector pixels. The T_j scores of maps from successful runs were in the range 0.1–0.3, while those from failed runs were predominantly 0.2–0.6. When the threshold of T_j scores is set at 0.2, maps from successful runs could be distinguished from failed runs. Then, the averaged density maps clearly present the images of aggregates with standard deviations from the average of less than 0.5%. Although a small number of maps from failed runs with the T_j scores less than 0.2 are included in the averaging, the influence on the averaged maps and the standard deviation was very small.

To further examine the practical feasibility, the protocol was applied to diffraction patterns from specimens composed of gold colloidal particles distributed in large areas (Fig. 5 and Table 1). The particles in large areas gave fine interference

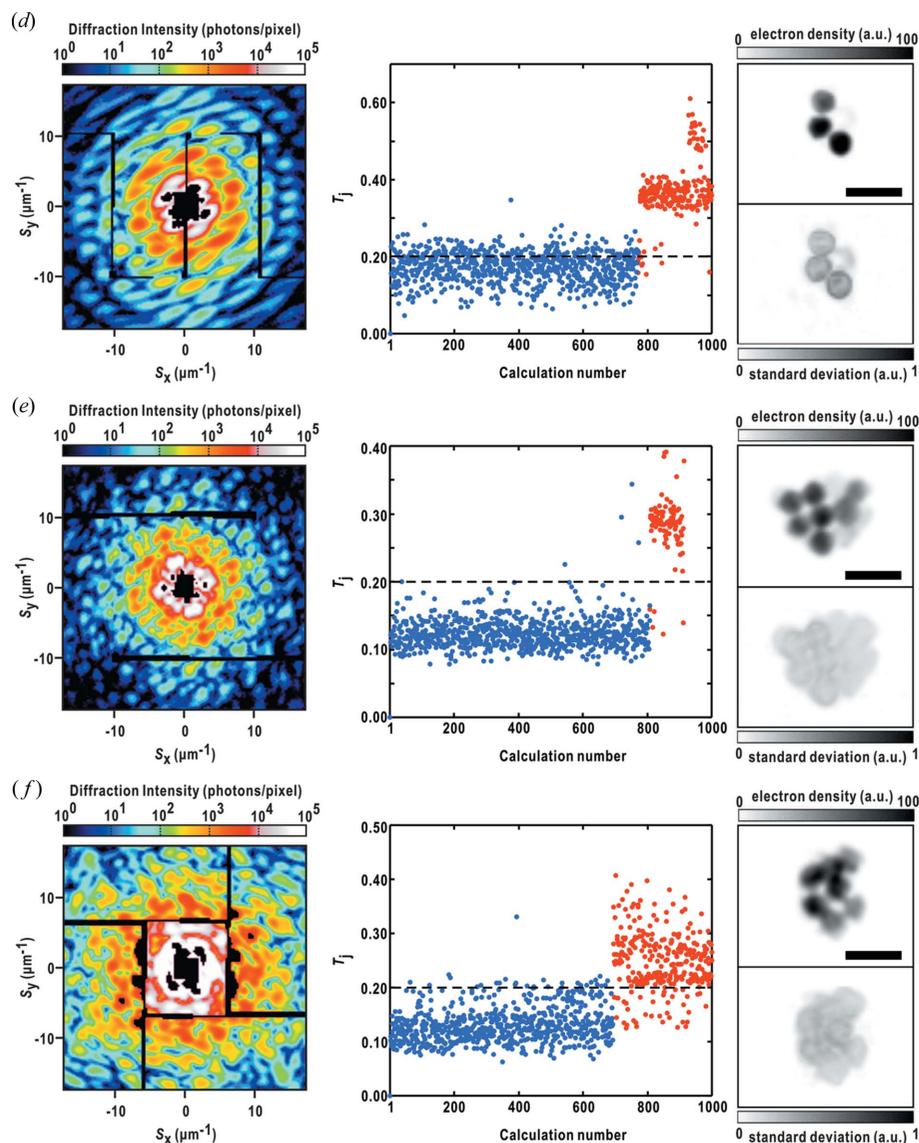


Figure 4 (continued)

Application of the T_j analysis to diffraction patterns from various aggregates of compact arranged gold colloidal particles. Each panel shows the diffraction pattern (left), the T_j values of 1000 PR maps (middle) and a set of the averaged map and the standard deviation map (right). T_j values from successful and failed runs are indicated by blue and red dots, respectively. The average map is calculated from maps with T_j values smaller than 0.2. The scale bar indicates 600 nm. The statistics of diffraction patterns and phase-retrieved electron density maps in panels (a)–(f) are summarized in Table 1.

patterns with small OS ratios. Thus, the diffraction patterns would give the opportunity for more severe examination of the protocol than those in Fig. 4. The T_j scores of maps from successful runs were predominantly in the range 0.1–0.4, while those from failed runs were 0.2–0.8 except a small number of maps with scores less than 0.2. The success of PR calculations likely depended on the size of the area to be retrieved. The numbers of failed runs for diffraction patterns from particles distributed within 1.3 μm were 200–400 [Figs. 5(a)–5(d)] and Table 1], while the numbers exceeded 600 for diffraction patterns from particles separated by more than 1.6 μm [Figs. 5(e), 5(f) and Table 1].

The tendencies in the distribution of T_j scores suggested that the threshold level of 0.2 for T_j scores is likely suitable to discriminate between most of the successful and failed runs. Although a small number of maps from failed runs were extracted in Figs. 5(c)–5(f) under the discrimination level, their influence on the averaged maps was negligible. The averaged electron density maps were composed of particles with standard deviations of less than 0.8% of the maximum density.

From the structure analyses for the 12 experimental diffraction patterns with a variety of intensity, Poisson noise, OS ratios and the arrangements of gold colloidal particles (Figs. 4 and 5), T_j of 0.2 was likely suitable as a threshold to discriminate between successful and failed runs. A small number of maps from failed runs having T_j scores of less than 0.2 are also extracted. However, the maps contribute little to the average maps and the standard deviations. In the following sections, the protocol and the discrimination level for the similarity scores were examined further by applying to diffraction patterns from complicated specimens.

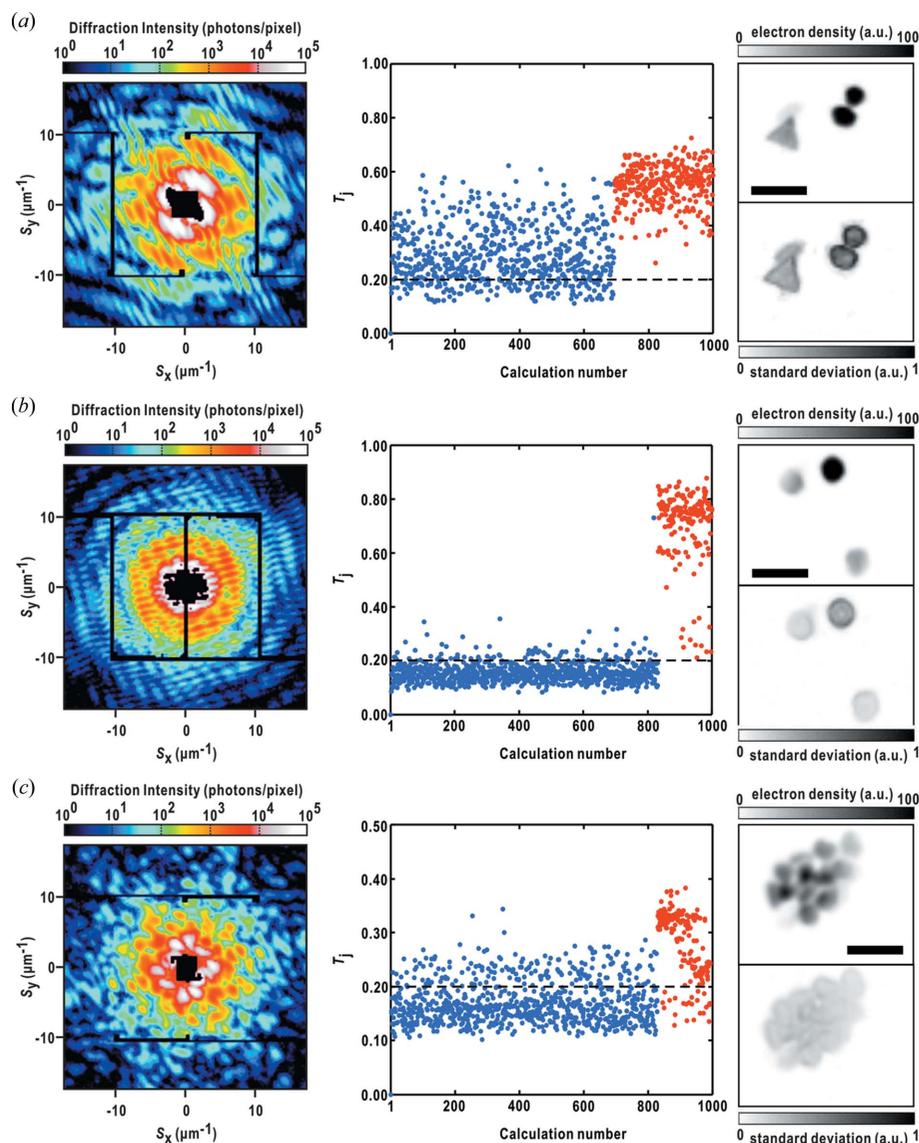
4.2. Similarity score of the second type

The threshold for the T_j score is useful for extracting correct maps as demonstrated in Fig. 4. However, it is difficult to deny the possibility that a map from a failed run rarely displayed the best $N_{0.5}$ value as shown in Fig. 5. When such a map is selected as the reference, T_j scores are distributed around 0.5, and scores smaller than 0.2 are rare (Table 1). Consequently, we introduced another score evaluating the similarity by exchanging the reference maps sequentially starting from the best

map to maps with lower $N_{0.5}$ values as

$$T_{ij} = \sum_{x,y} \left| \rho_i(x,y) - \rho_j(x,y) \right| / \sum_{x,y} \left| \rho_i(x,y) + \rho_j(x,y) \right|, \quad (3)$$

where $\rho_i(x,y)$ is the density value of the reference map, and $\rho_j(x,y)$ is that of the other map. We searched for the reference map that gave T_{ij} scores smaller than 0.2 most frequently. Through the application of the protocol to diffraction patterns, where the best maps are accidentally extracted from failed runs, we found that the maps from successful runs were included in the maps with the 100 highest $N_{0.5}$ scores.


Figure 5

Application of the T_j analysis to diffraction patterns from gold colloidal particles sparsely distributed or forming a large aggregate [panels (a)–(f)]. Each panel shows the diffraction pattern (left), the T_j values of 1000 PR maps (middle) and a set of the averaged map and the standard deviation map (right). T_j values from successful and failed runs are indicated by blue and red dots, respectively. The average map is calculated from maps with T_j values smaller than 0.2. The scale bar indicates 600 nm. The statistics of diffraction patterns and phase-retrieved electron density maps in panels (a)–(f) are summarized in Table 1.

In the case of the diffraction pattern shown in Fig. 6(a), which came from an aggregate composed of colloidal particles, the T_j values for the best map are greater than 0.5. The reference map of the 33rd highest $N_{0.5}$ score gave T_{ij} scores smaller than 0.2 for approximately 200 maps, with T_{ij} scores for the maps from failed runs being larger than 0.3 (Table 1). The average map gave a clear image of five gold colloidal particles with small standard deviation from the average.

Fig. 6(b) demonstrates another example, in which the T_j score is used to find maps from successful runs. The T_j scores using the best map as a reference were in the range 0.4–0.6, indicating that the best map came from a failed run. The map with the 15th highest $N_{0.5}$ score gave the largest number of

maps (approximately 200) with T_{ij} scores smaller than 0.2 (Table 1). The average map was composed of a compactly packed aggregate of seven colloidal particles accompanying an additional separate particle. These two examples indicate the potential of the T_{ij} scores for finding maps from successful runs, even when the best maps were selected from failed runs.

4.3. Application to biological non-crystalline particles

Because biological specimens, which are composed of light atoms, have total scattering cross sections for X-rays smaller than those of aggregates for gold colloidal particles, the diffraction patterns of biological specimens are characterized by a weak intensity. As reported in our previous simulation studies, the PR calculation using the HIO and SW algorithms are affected by Poisson noise (Kodama & Nakasako, 2011) and the electron density contrast (Oroguchi & Nakasako, 2013). Therefore, the number of successful PR runs would be smaller than those for metal particles, and then the T_{ij} score would be more effective in searching maps from successful runs.

For the diffraction pattern from an isolated chloroplast of *Cyanidioschyzon merolae* (*C. merolae*) [Fig. 7(a)] (Takayama *et al.*, 2015b), PR maps were divided roughly into two groups: approximately 70% of the maps with T_{ij} scores in the range 0.2–0.3, and the other maps with scores in the range 0.4–0.7. An average map calculated from those with T_{ij} scores smaller than 0.2 appears as an annular shape with four prominent peaks. This map is similar to

the most probable map, which is selected by the previously reported multivariate analysis against 1000 PR maps (Sekiguchi *et al.*, 2016).

Another example is an isolated nucleus from budding yeast at the G2/M phase in the cell cycle [Fig. 7(b)]. Similar to the chloroplast case, the PR maps were divided into two groups with respect to the T_{ij} scores. Approximately 25% of maps have T_{ij} scores smaller than 0.2. The average map calculated from those with the T_{ij} scores smaller than 0.2 is superimposable on the map selected as the most probable support in the previous study (Oroguchi *et al.*, 2015).

These results suggest the possibility that PR maps displaying T_{ij} scores smaller than 0.2 are candidates for being

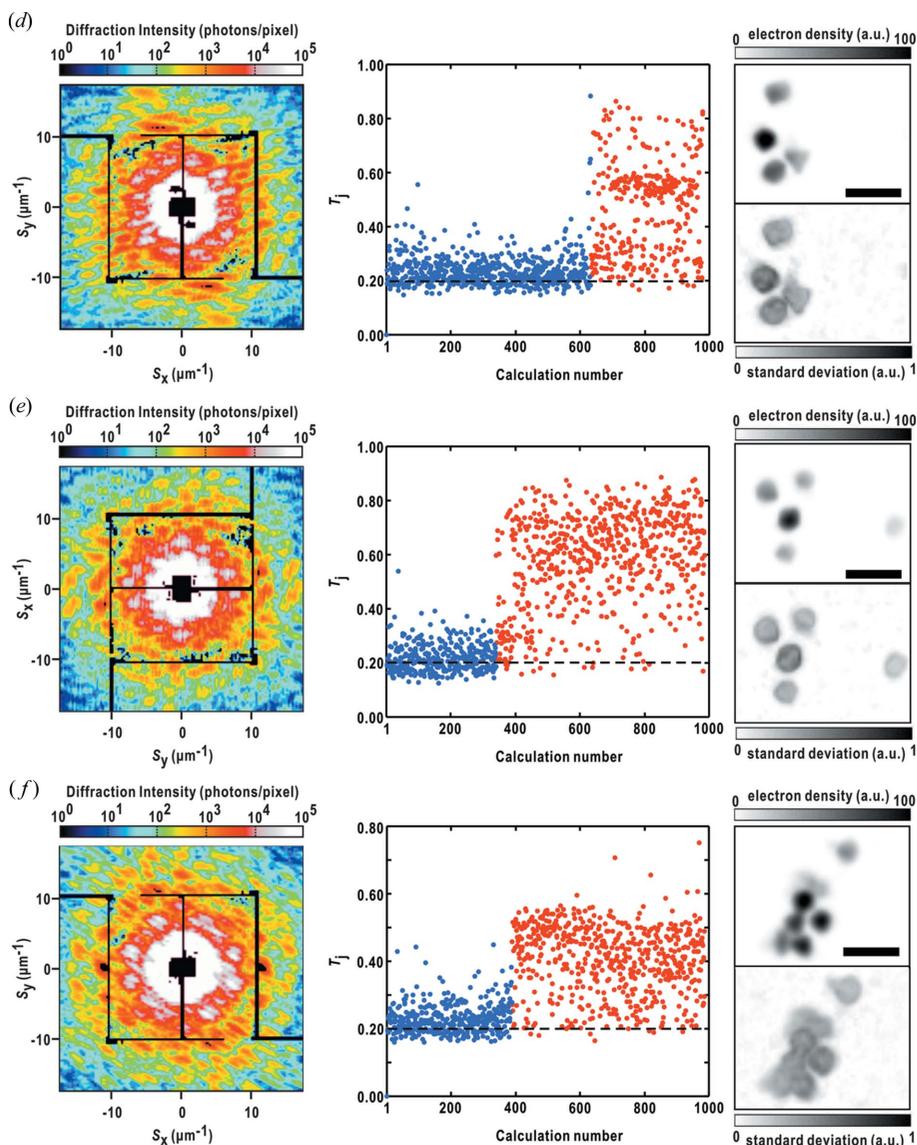


Figure 5 (continued)
 Application of the T_j analysis to diffraction patterns from gold colloidal particles sparsely distributed or forming a large aggregate [panels (a)–(f)]. Each panel shows the diffraction pattern (left), the T_j values of 1000 PR maps (middle) and a set of the averaged map and the standard deviation map (right). T_j values from successful and failed runs are indicated by blue and red dots, respectively. The average map is calculated from maps with T_j values smaller than 0.2. The scale bar indicates 600 nm. The statistics of diffraction patterns and phase-retrieved electron density maps in panels (a)–(f) are summarized in Table 1.

the most probable maps. This point will be discussed later by comparing the results of the present protocol with those of the previously reported protocol.

5. Discussion

In this study, we propose a protocol for extracting maps from successful runs in PR calculations of diffraction patterns in CXDI. In the protocol, we introduced a FOM to parameterize the variation of phase values in PR calculations and similarity scores as indicators to efficiently identify maps from successful runs. Here, we compare the results with those obtained

using the previously proposed protocol incorporating the multivariate analysis.

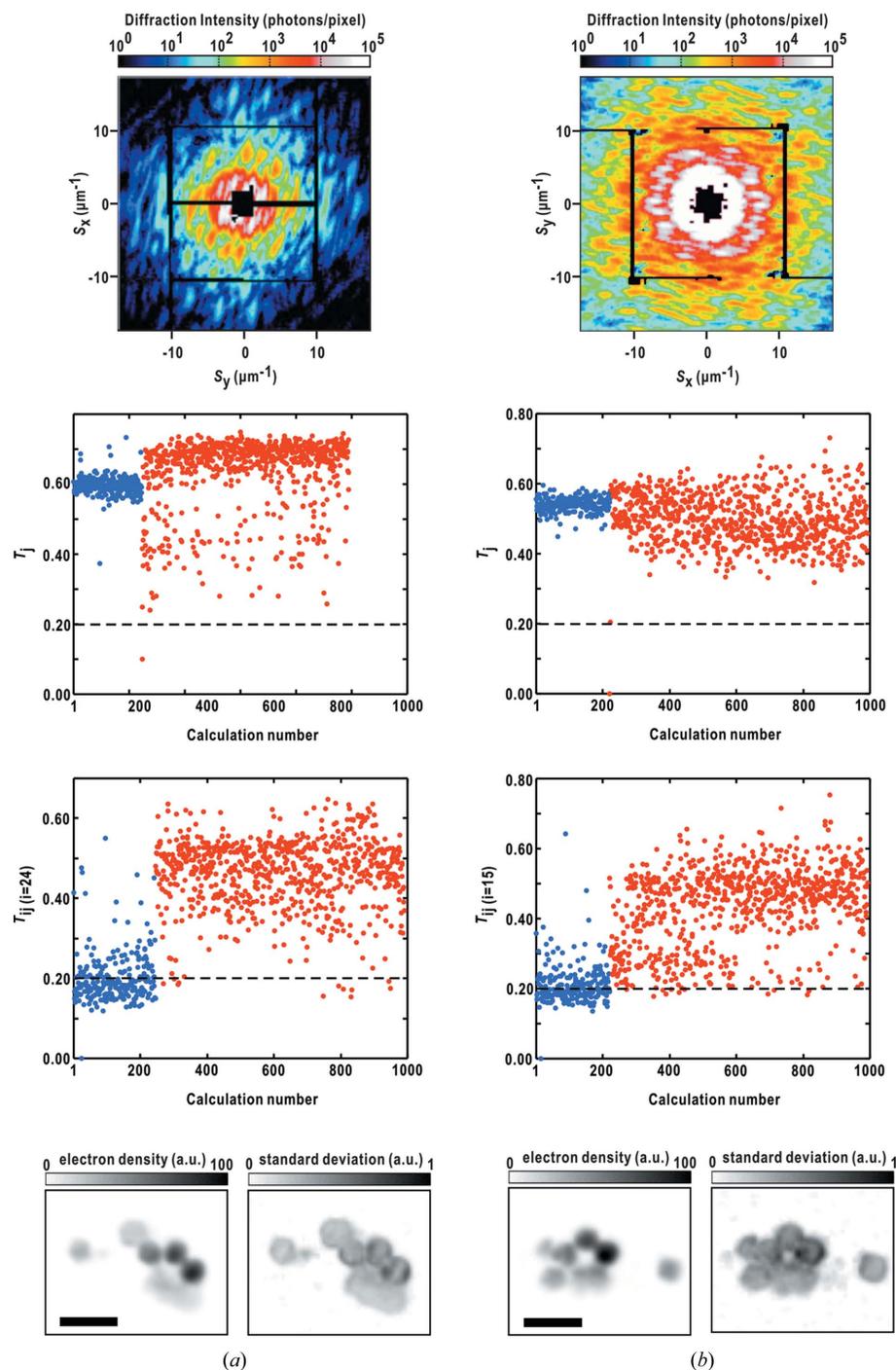
5.1. Feasibility study of the proposed protocol by using experimental data

We reported several simulation studies on PR procedures in CXDI (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013; Kobayashi *et al.*, 2014; Takayama *et al.*, 2015a; Yoshidome *et al.*, 2015). In those simulation studies, calculation conditions are limited to the variation of incident intensity and a small number of structural models. In simulation studies, PR maps can be classified into several groups by monitoring the degree of similarity to a structure model as done in our previous study (Kobayashi *et al.*, 2014). However, CXDI experiments are requested to visualize the structures of specimen particles, and therefore we propose the protocol to extract only correct maps from successful runs.

For this purpose, diffraction patterns from gold colloidal particles are advantageous for identifying correct maps from successful runs and are suitable for defining the threshold level of similarity scores for the practical application of the protocol. In addition, diffraction patterns from gold colloidal particles are varied with respect to structure, intensities, missing small-angle regions, Poisson noise and OS ratios. To examine the practical feasibility of the proposed protocol, diffraction patterns from dispersed gold colloidal particles can provide a variety of specimens rather than those from simulation models under limited conditions.

5.2. Benefit of the protocol

To date, various types of PR algorithms have been proposed to obtain the most probable maps from the PR calculations (Fienup, 1982; Elser, 2003; Luke, 2005; Chen *et al.*, 2007; Rodriguez *et al.*, 2013). Under current standard methods, PR maps that have similar shapes to images obtained by transmission electron microscopy and/or light microscopy are extracted and averaged as the most probable maps. In our recent XFEL-CXDI experiments, a large number of diffraction patterns were collected in a short period of time. Subsequently, maps from successful calculations were found to be


Figure 6

Application of the T_{ij} analysis for diffraction patterns from aggregates of gold colloidal particles. Each panel, from top to bottom, shows the diffraction pattern, the T_j values of PR maps, the T_{ij} values for a reference map giving the largest number of maps with T_{ij} values smaller than 0.2, and the average map with the standard deviation. The T_j and T_{ij} scores from successful and failed runs are indicated by blue and red dots, respectively. The statistics of diffraction patterns and phase-retrieved electron density maps in panels (a) and (b) are summarized in Table 1.

automatically extracted more reliably without information from other microscopic observations.

In the previous study, we proposed a protocol to provide opportunities for more objective assessment of PR maps by using the multivariate analysis (Sekiguchi *et al.*, 2016). Although the protocol is useful for suggesting PR maps from

successful PR calculations, it requires manual inspections of the results from multivariate analyses. In contrast, the protocol proposed in this study can suggest maps from successful PR calculations without manual inspection. Therefore, this protocol is suitable for automatically and efficiently extracting maps from successful calculations. Because the diffraction apparatus allows us to collect a large number of diffraction patterns within a short period of time, the automatic extraction without time-consuming inspections provides benefits in structure analyses in CXDI.

5.3. Tendencies in PR calculations and PR maps

In the structure analysis for the diffraction patterns of gold colloidal particles, the $N_{0.5}$ parameter and the similarity scores are useful for extracting correct maps (Figs. 2, 4, 5 and 6), even when more than 70% of the PR calculations fail (Fig. 6). In the structure analyses of biological specimens, the protocol allows us to pick up candidates for maps from successful runs (Fig. 7). These results for the extraction of maps from successful runs suggest the following tendencies in the variation of phase values in PR calculations and the similarities among PR maps.

As speculated in the *Introduction*, we firstly confirmed the tendency for phase values in successful PR calculations to converge around certain values in the early stages of the PR cycles, which are almost retained until the end of the cycles [Figs. 1(b) and 2(a)]. In contrast, in failed calculations, phase values vary cycle-by-cycle until the end of the PR cycles, probably because incorrect maps are likely modified by the real-space constraints to give different values of the phase set from those before.

Second, the maps from successful PR calculations were found to resemble each other, as indicated by the similarity

scores, but were different from almost all maps from failed calculations (Figs. 2–6). In addition, maps from failed calculations are mutually different as characterized by the similarity scores (Fig. 6). This tendency of the similarity scores is important to distinguish between maps from successful and failed calculations (Figs. 2–7). Third, when a map that has a

Table 1
Statistics of diffraction patterns and phase-retrieved electron density maps.

Specimen	Diffraction pattern	C_{sym}^\dagger	S_{max} (nm) ‡	Approximate dimension (μm)	No. of successful runs	$\langle T_{ij} \rangle$ for successful runs	No. of failed runs	$\langle T_{ij} \rangle$ for failed runs	No. of runs with support out of convergence
Gold colloid	Fig. 1(a)	0.90	21.5	1.1	487	0.12	503	0.32	10
	Fig. 4(a)	0.96	18.8	0.7	963	0.19	34	0.38	3
	Fig. 4(b)	0.85	29.5	0.8	735	0.20	57	0.43	208
	Fig. 4(c)	0.91	26.9	1.0	526	0.14	323	0.31	151
	Fig. 4(d)	0.92	19.2	0.8	770	0.17	330	0.37	0
	Fig. 4(e)	0.92	24.1	1.0	808	0.12	108	0.30	84
	Fig. 4(f)	0.85	11.6	1.0	693	0.13	307	0.25	0
	Fig. 5(a)	0.88	20.1	1.1	692	0.29	308	0.55	0
	Fig. 5(b)	0.86	24.2	1.3	831	0.16	169	0.70	0
	Fig. 5(c)	0.86	20.4	1.2	829	0.17	164	0.28	7
	Fig. 5(d)	0.90	6.5	1.1	633	0.24	347	0.47	20
	Fig. 5(e)	0.92	10.3	1.7	340	0.21	646	0.58	14
	Fig. 5(f)	0.90	7.9	1.6	388	0.23	599	0.41	13
	Fig. 6(a)	0.86	27.8	1.5	245	$\langle T_{ij} \rangle = 0.20$	745	$\langle T_{ij} \rangle = 0.46$	10
Fig. 6(b)	0.89	11.7	1.5	220	$\langle T_{ij} \rangle = 0.21$	773	$\langle T_{ij} \rangle = 0.44$	7	
Chloroplast	Fig. 7(a)	0.82	29.0	1.2			$\langle T_{ij} \rangle = 0.42$		
Nucleus	Fig. 7(b)	0.83	18.3	1.2			$\langle T_{ij} \rangle = 0.33$		

† The Friedel symmetry of a diffraction pattern is evaluated using the following correlation: $C_{\text{sym}} = (E - O)/(E + O)$, $E = \sum_{x,y} [I_0(x,y) + I_{\text{sym}}(-x,-y)]^2$, $O = \sum_{x,y} [I_0(x,y) - I_{\text{sym}}(-x,-y)]^2$ where $I_0(x,y)$ is the diffraction intensity in the region of interest with 100×100 pixels and $I_{\text{sym}}(-x,-y)$ is the diffraction intensity of the Friedel mate. For a diffraction pattern with ideal Friedel symmetry, the C_{sym} value is 1. ‡ Maximum resolution is defined as the highest-resolution shell including at least two detector pixels with more than four photons.

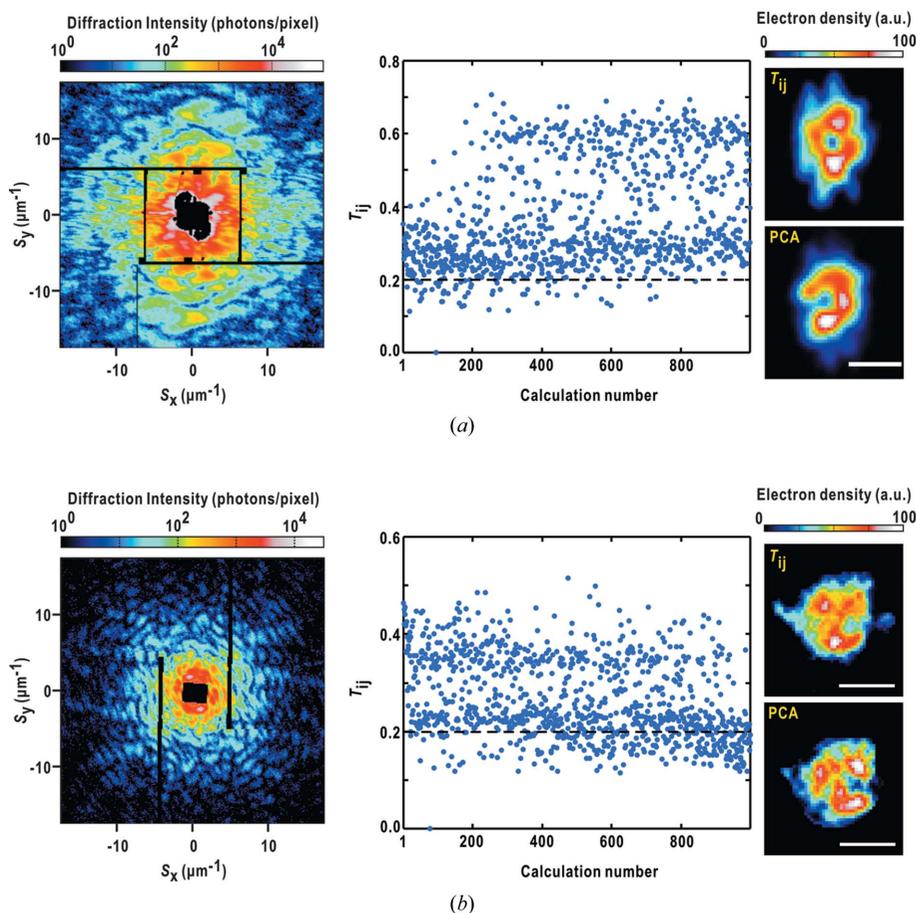


Figure 7
Application of the T_{ij} analyses for the diffraction patterns of (a) a chloroplast of *C. merolae* and (b) a nucleus of budding yeast. Each panel shows the diffraction pattern (left), the T_{ij} values for a reference map giving the largest number of maps with T_{ij} values smaller than 0.2 (middle), and the average map compared with the map estimated by the multivariate analysis (Oroguchi *et al.*, 2015; Sekiguchi *et al.*, 2016) (right). The scale bar indicates 500 nm.

density distribution close to the true value is used as a reference in the analysis of similarity scores, the number of maps with scores less than 0.2 becomes largest in the T_{ij} analysis conducted by exchanging the reference map (Figs. 2–7).

The number of correct or probable maps appearing in 1000 PR calculations depends on the intensities, OS ratios and the areas of detector saturation in the diffraction patterns. When a specimen particle with a large scattering cross section gives a diffraction pattern with a large OS ratio, a number of successful runs appear in the PR calculation. Subsequently, the T_j scores are useful in extracting correct maps (Figs. 4 and 5). In contrast, for diffraction patterns with small OS ratios (Fig. 6) and biological specimens with small scattering cross sections (Fig. 7), the T_{ij} analysis is better for searching for maps from successful calculations.

5.4. Relationship between similarity analysis and multivariate analysis

In the previous study, we proposed a protocol to suggest the most probable maps among 1000 PR maps by using the multivariate analysis (Sekiguchi *et al.*, 2016). The distribution of PR maps in the multidimensional image space is visualized in the plane spanned by the

two lowest principal components (PCs). In this regard, it is interesting to inspect where the maps extracted by the present protocol are distributed in the plane (Fig. 8).

The maps retrieved from Fig. 1(a) are classified into three clusters on the PC plane. The correct maps with T_{ij} scores smaller than 0.2 are distributed on cluster I, which was the most probable in the previous study [Fig. 8(a)]. Regarding a chloroplast of *C. merolae*, the PR maps are divided roughly into three clusters on the PC plane. The PR maps composing dense cluster I have shapes and sizes smaller than those known in optical microscopy, and then the central part of

cluster II is selected as the most probable in the previous study. The maps displaying T_{ij} scores smaller than 0.2 are located in the center of cluster II. The PR maps of a nucleus isolated from budding yeast are distributed in clusters I–IV in PC planes [Fig. 8(c)], and then the most probable maps are extracted from cluster I. The maps that display T_{ij} scores smaller than 0.2 are distributed within the cluster. These comparisons suggest that maps with T_{ij} scores smaller than 0.2 can be treated as the most probable maps.

The PR calculation searches sets of maps with the most adequate values of all the pixels to explain a diffraction

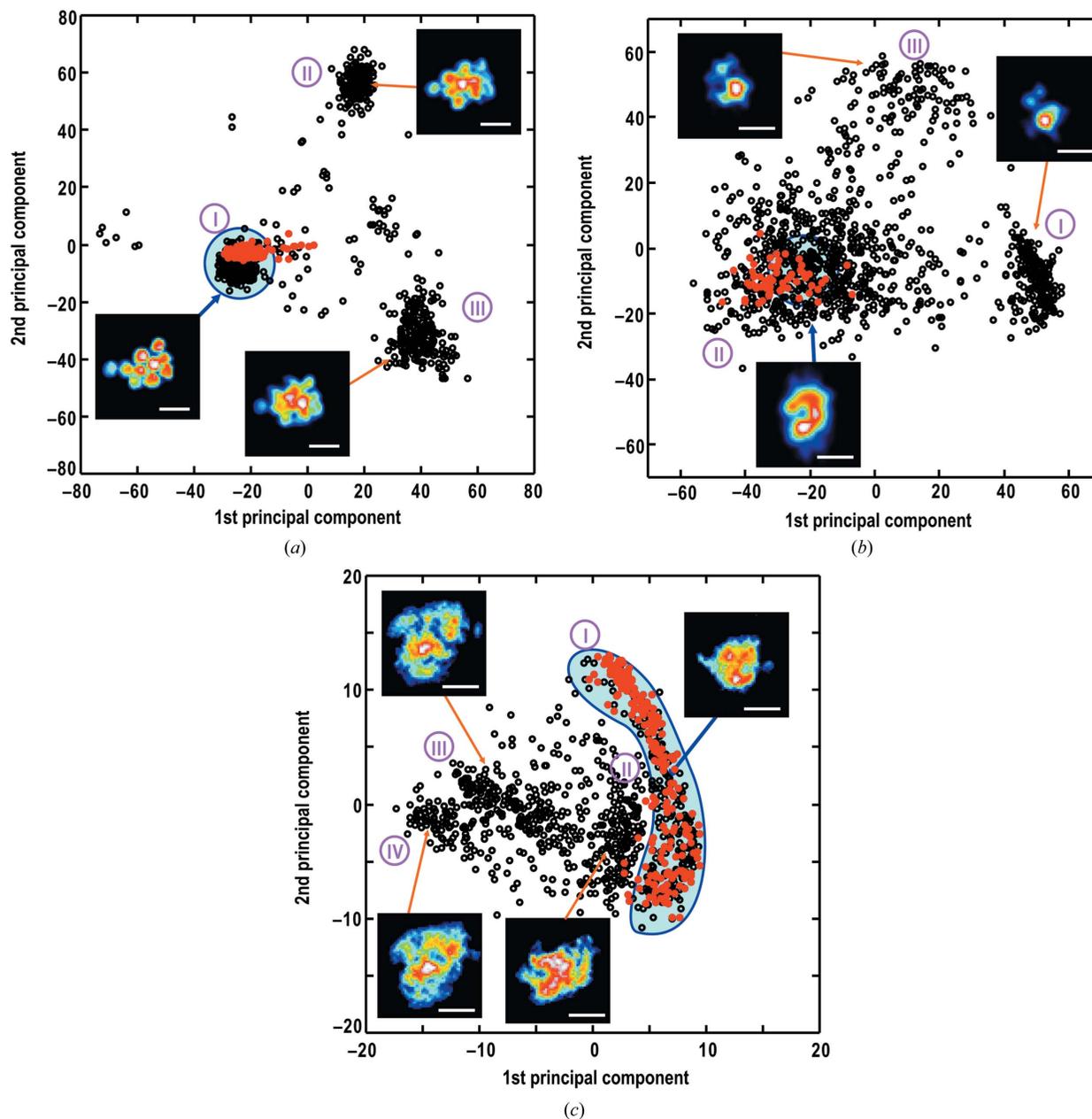


Figure 8

Two-dimensional plots illustrating where the maps with similarity scores smaller than 0.2 are distributed in the plane spanned by the first and second PCs, predominantly explaining the variance in 1000 PR maps, regarding the PR calculations for the diffraction patterns from (a) gold colloidal particles [Fig. 1(a)], (b) a chloroplast of *C. merolae* [Fig. 7(a)] and (c) a nucleus of budding yeast [Fig. 7(b)]. In each plot, the positions of PR maps are indicated by black dots. The red dots indicate the positions of maps with similarity scores smaller than 0.2 projected on the plane. The representative maps of the clusters appearing by the PCA are shown. The cyan background indicates the cluster selected as the most probable maps in the previous studies (Oroguchi *et al.*, 2015; Sekiguchi *et al.*, 2016). The scale bar indicates 500 nm.

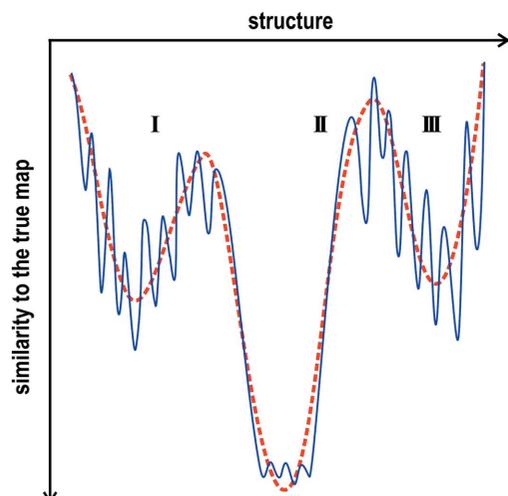


Figure 9
Schematic illustration of the landscape regarding the similarity of maps to the true maps. The red curve shows the smooth changes of the landscape viewed from PCs. The blue curve indicates the rugged landscape viewed by the similarity score.

pattern. The principal component analysis (PCA) used in the previous study determines a small number of principal components that describe the major variance among maps with a minimal loss of information. However, the PCs with large eigenvalues are insensitive to the variation of electron densities in individual pixels. In contrast, the similarity score, which is the normalized version of the Manhattan distance (Faith *et al.*, 1987), is sensitive to pixel-by-pixel variation.

The landscape regarding the distribution of PR maps in the multidimensional image space is described by the similarity of PR maps to the true map (Fig. 9). Taking the results from multivariate analyses (Fig. 8), the landscape illustrated by PCs is composed of smooth basins. Consequently, differences between PR maps in the same basin are difficult to distinguish. In contrast, when the landscape is illustrated by using the similarity score, PR maps from failed calculations differ from each other as indicated by their large T_j values (Fig. 6), suggesting that the landscape described by the similarity scores are significantly rugged. In addition, the small T_j values for maps from successful calculations suggest their similarity and localization in the multidimensional space. Although the shapes and sizes of basins in a rugged landscape observed using the similarity scores probably depend on the signal-to-noise ratio, oversampling ratio and the size of the small-angle area missing the diffraction pattern (Table 1), the similarity score is still useful for finding maps from successful PR calculations.

5.5. Threshold levels of FOM and T_j scores

In the present study, we counted pixels in diffraction patterns with a FOM larger than 0.5 in order to find the most probable or correct maps (Figs. 2 and 3). This threshold for the FOM was tentatively defined through the application of the proposed protocol for a number of diffraction patterns. In the experimental determination of phases by single-particle cryo-

electron microscopy (Rosenthal & Henderson, 2003) and X-ray protein crystallography (Lunin & Woolfson, 1993; Perrakis *et al.*, 1997), the averaged FOM of phase sets less than 0.5 is used as a major index to examine whether the obtained maps are interpretable. However, in contrast to the FOM in these techniques, large FOM values in the entirely computational PR calculations only suggest that the PR calculations converge into global or local minima. Therefore, as demonstrated in Fig. 6, maps in failed runs were accidentally extracted as the best map.

For the similarity scores, we applied a threshold of 0.2 to extract the correct electron density maps through the application of the protocol. This is most likely due to the characteristics of the rugged landscape of PR maps in the multidimensional space; similarity scores play a role in finding incorrect extractions of the best map by the FOM (Fig. 6). Since the minimum values and variation of the scores depend on the intensity, speckle size and the missed small-angle region, the threshold value would be better to accommodate the characteristics of the diffraction patterns.

Regardless, the threshold values of the FOM and similarity scores may be refined through the application of the protocol to various types of diffraction patterns in the future.

5.6. Future prospects

In this study, the FOM value calculated from all the phase sets that appeared in a single PR run was used to quantify phase variations. As seen in Fig. 1(c), the phase sets in successful runs often converged in the early stage of the iterative PR calculations. Subsequently, introducing on-the-fly monitoring criteria for the phase variations to confirm the convergence during PR runs would enable us to terminate calculations immediately after confirming success in retrieving correct electron density maps. This on-the-fly analysis might dramatically reduce computational costs. Moreover, monitoring criteria for the converging speed of retrieved phases would be a useful tool for evaluating the performance of various PR algorithms.

Since the introduced T_j score can sensitively evaluate the similarity between PR maps, T_j can contribute to the selection of the most probable maps among those with the same overall shapes resulting from high-resolution PR calculations with input of a fixed support area (Sekiguchi *et al.*, 2016). If high-resolution PR calculations are also automated, throughput of XFEL-CXDI structure analyses would be greatly improved and the analyses could come into use for nonprofessional users.

Acknowledgements

We selected representative diffraction data from cryogenic XFEL-CXDI experiments performed at SACLA (proposal Nos. 2014A8033, 2014A8033, 2014B8052, 2015A8051, 2016A8048 and 2016B8064). The authors would like to thank the members of the SACLA engineering team for their great help in the alignment and operation of the focusing mirror

optics, our diffractometer and the two detectors. This study was supported by a grant for XFEL Key Technology and the X-ray Free Electron Laser Priority Strategy Program from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to MN, Grant-in-Aid for Scientific Research on Innovative Areas to MN and to TO, Grant-in-Aid for Young Scientists (B) to TO, Grant-in-Aid for Challenging Exploratory Research to MN and Grant-in-Aid for JSPS Fellows to YS and to AK, and from the Japan Society for the Promotion of Science to MN, as detailed below. The PR calculations and multivariate analyses were performed using the mini-K supercomputer system at the SACLA facility.

Funding information

The following funding is acknowledged: Ministry of Education, Culture, Sports, Science and Technology (award No. X-ray Free Electron Laser Priority Strategy Program); Japan Society for the Promotion of Science (award No. jp15076210; award No. jp23120525; award No. jp251202725; award No. 15H01647; award No. jp24113723; award No. jp26104535; award No. jp26800227; award No. jp17654084; award No. jp24654140; award No. jp15J1707; award No. jp15J01831; award No. jp1920402; award No. jp16H02218).

References

- Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
- Chen, C.-C., Miao, J., Wang, C. W. & Lee, T. K. (2007). *Phys. Rev. B*, **76**, 064113.
- Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R. N. C., Seltzer, V., Claverie, J., Hantke, M., Jonsson, 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.
- Elser, V. (2003). *J. Opt. Soc. Am. A*, **20**, 40–55.
- Faith, D. P., Minchin, P. R. & Belbin, L. (1987). *Vegetatio*, **69**, 57–68.
- Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
- Hantke, M. F., Hasse, D., Maia, F. R. N. C., Ekeberg, T., John, K., Svenda, M., Loh, N. D., Martin, A. V., Timneanu, N., Larsson, D. S. D., van der Schot, G., Carlsson, G. H., Ingelman, M., Andreasson, J., Westphal, D., Liang, M., Stellato, F., DePonte, D. P., Hartmann, R., Kimmel, N., Kirian, R. A., Seibert, M. M., Mühlig, K., Schorb, S., Ferguson, K., Bostedt, C., Carron, S., Bozek, J. D., Rolles, D., Rudenko, A., Epp, S., Chapman, H. N., Barty, A., Hajdu, J. & Andersson, I. (2014). *Nat. Photon.* **8**, 943–949.
- 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., Oroguchi, T., Okajima, K., Fukuda, A., Oide, M., Yamamoto, M. & Nakasako, M. (2016b). *J. Synchrotron Rad.* **23**, 975–989.
- Kobayashi, A., Sekiguchi, Y., Takayama, Y., Oroguchi, T. & Nakasako, N. (2014). *Opt. Express*, **22**, 27892.
- Kobayashi, A., Sekiguchi, Y., Takayama, Y., Oroguchi, T., Shirahama, K., Torizuka, Y., Manoda, M., Nakasako, M. & Yamamoto, M. (2016a). *Rev. Sci. Instrum.* **87**, 053109.
- Kodama, W. & Nakasako, M. (2011). *Phys. Rev. E*, **84**, 021902.
- Loh, N. D., Hampton, C. Y., Martin, A. V., Starodub, D., Sierra, R. G., Barty, A., Aquila, A., Schulz, J., Lomb, L., Steinbrener, J., Shoeman, R. L., Kassemeyer, S., Bostedt, C., Bozek, J., Epp, S. W., Erk, B., Hartmann, R., Rolles, D., Rudenko, A., Rudek, B., Foucar, L., Kimmel, N., Weidenspointner, G., Hauser, G., Holl, P., Pedersoli, E., Liang, M., Hunter, M. M., Hunter, M. M., Gumprecht, L., Coppola, N., Wunderer, C., Graafsma, H., Maia, F. R., Ekeberg, T., Hantke, M., Fleckenstein, H., Hirsemann, H., Nass, K., White, T. A., Tobias, H. J., Farquar, G. R., Benner, W. H., Hau-Riege, S. P., Reich, C., Hartmann, A., Soltau, H., Marchesini, S., Bajt, S., Barthelmess, M., Bucksbaum, P., Hodgson, K. O., Strüder, L., Ullrich, J., Frank, M., Schlichting, I., Chapman, H. N. & Bogan, M. J. (2012). *Nature (London)*, **486**, 513–517.
- Luke, D. R. (2005). *Inverse Probl.* **21**, 37–50.
- Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* **D49**, 530–533.
- Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., Weierstall, U. & Spence, J. C. H. (2003). *Phys. Rev. B*, **68**, 140101.
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.
- Miao, J., 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., Hodgson, K. O., Ishikawa, T., Larabell, C. A., LeGros, M. A. & Nishino, Y. (2003b). *Proc. Natl Acad. Sci.* **100**, 110–112.
- Miao, J., Ishikawa, T., Anderson, E. H. & Hodgson, K. O. (2003a). *Phys. Rev. B*, **67**, 174104.
- Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. (2015). *Science*, **348**, 530–535.
- Nakasako, M., Takayama, Y., Oroguchi, T., Sekiguchi, Y., Kobayashi, A., Shirahama, K., Yamamoto, M., Hikima, T., Yonekura, K., Maki-Yonekura, S., Kohmura, Y., Inubushi, Y., Takahashi, Y., Suzuki, A., Matsunaga, S., Inui, Y., Tono, K., Kameshima, T., Joti, Y. & Hoshi, T. (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.
- 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.
- Oroguchi, T., Sekiguchi, Y., Kobayashi, A., Masaki, Y., Fukuda, A., Hashimoto, S., Nakasako, M., Ichikawa, Y., Kurumizaka, H., Shimizu, M., Inui, Y., Matsunaga, S., Kato, T., Namba, K., Yamaguchi, K., Kuwata, K., Kameda, H., Fukui, N., Kawata, Y., Kameshima, T., Takayama, Y., Yonekura, K. & Yamamoto, M. (2015). *J. Phys. B*, **48**, 184003.
- Park, H. J., Loh, N. D., Sierra, R. G., Hampton, C. Y., Starodub, D., Martin, A. V., Barty, A., Aquila, A., Schulz, J., Steinbrener, J., Shoeman, R. L., Lomb, L., Kassemeyer, S., Bostedt, C., Bozek, J., Epp, S. W., Erk, B., Hartmann, R., Rolles, D., Rudenko, A., Rudek, B., Foucar, L., Kimmel, N., Weidenspointner, G., Hauser, G., Holl, P., Pedersoli, E., Liang, M., Hunter, M. S., Gumprecht, L., Coppola, N., Wunderer, C., Graafsma, H., Maia, F. R. N. C., Ekeberg, T., Hantke, M., Fleckenstein, H., Hirsemann, H., Nass, K., Tobias, H. J., Farquar, G. R., Benner, W. H., Hau-Riege, S., Reich, C., Hartmann, A., Soltau, H., Marchesini, S., Bajt, S., Barthelmess, M., Strueder, L., Ullrich, J., Bucksbaum, P., Frank, M., Schlichting, I., Chapman, H. N., Bogan, M. J. & Elser, V. (2013). *Opt. Express*, **21**, 28729–28742.
- Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* **D53**, 448–455.
- 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, 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., Rucker, 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.-M., 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.-U., Andritschke, R., Schröter, C.-D., 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.
- Sekiguchi, Y., Oroguchi, T., Takayama, Y. & Nakasako, M. (2014a). *J. Synchrotron Rad.* **21**, 600–612.
- Sekiguchi, Y., Yamamoto, M., Oroguchi, T., Takayama, Y., Suzuki, S. & Nakasako, M. (2014b). *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.* **102**, 15343–15346.
- Takayama, Y., Inui, Y., Sekiguchi, Y., Kobayashi, A., Oroguchi, T., Yamamoto, M., Matsunaga, S. & Nakasako, M. (2015b). *Plant Cell Physiol.* **56**, 1272–1286.
- Takayama, Y., Maki-Yonekura, S., Oroguchi, T., Nakasako, M. & Yonekura, K. (2015a). *Sci. Rep.* **5**, 8074.
- Takayama, Y. & Nakasako, M. (2012). *Rev. Sci. Instrum.* **83**, 054301.
- Taylor, G. (2003). *Acta Cryst.* **D59**, 1881–1890.
- 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.
- 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.
- Yoshidome, T., Oroguchi, T., Nakasako, M. & Ikeguchi, M. (2015). *Phys. Rev. E*, **92**, 032710.
- Yumoto, H., Mimura, H., Koyama, T., Matsuyama, S., Tono, K., Togashi, T., Inubushi, Y., Sato, T., Tanaka, T., Kimura, T., Yokoyama, H., Kim, J., Sano, Y., Hachisu, Y., Yabashi, M., Ohashi, H., Ohmori, H., Ishikawa, T. & Yamauchi, K. (2013). *Nat. Photon.* **7**, 43–47.