

Similarity score for screening phase-retrieved maps in X-ray diffraction imaging – characterization in reciprocal space

Yuki Takayama^{a,b,c,d} and Masayoshi Nakasako^{b,e*}

Received 14 July 2023

Accepted 10 November 2023

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

Keywords: X-ray diffraction imaging; phase retrieval calculation; phase problem.

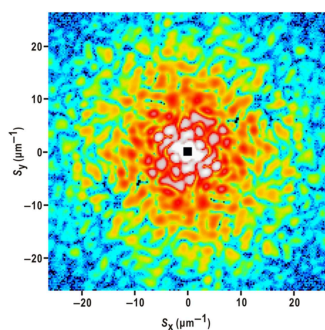
^aGraduate School of Science, University of Hyogo, 3-2-1 Kouto, Kamigori-cho, Ako-gun, Hyogo 678-1297, Japan, ^bRIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Sayo-gun, Hyogo 679-5148, Japan, ^cSynchrotron Radiation Research Center, Hyogo Science and Technology Association, 1-490-2 Kouto, Shingu, Tatsuno, Hyogo 679-5148, Japan, ^dInternational Center for Synchrotron Radiation Innovation Smart, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan, and ^eDepartment of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan. *Correspondence e-mail: nakasako@phys.keio.ac.jp

X-ray diffraction imaging (XDI) is utilized for visualizing the structures of non-crystalline particles in material sciences and biology. In the structural analysis, phase-retrieval (PR) algorithms are applied to the diffraction amplitude data alone to reconstruct the electron density map of a specimen particle projected along the direction of the incident X-rays. However, PR calculations may not lead to good convergence because of a lack of diffraction patterns in small-angle regions and Poisson noise in X-ray detection. Therefore, the PR calculation is still a bottleneck for the efficient application of XDI in the structural analyses of non-crystalline particles. For screening maps from hundreds of trial PR calculations, we have been using a score and measuring the similarity between a pair of retrieved maps. Empirically, probable maps approximating the particle structures gave a score smaller than a threshold value, but the reasons for the effectiveness of the score are still unclear. In this study, the score is characterized in terms of the phase differences between the structure factors of the retrieved maps, the usefulness of the score in screening the maps retrieved from experimental diffraction patterns is demonstrated, and the effective resolution of similarity-score-selected maps is discussed.

1. Introduction

X-ray diffraction imaging (XDI) is a technique for visualizing the structures of non-crystalline particles of size in the micro-to sub-micrometre range (Miao *et al.*, 1999, 2015; Nakasako, 2018; Nakasako *et al.*, 2020). In an XDI experiment, the particle to be imaged is irradiated by an X-ray beam with almost complete spatial coherence, and the Fraunhofer diffraction pattern of the particle is recorded at a high sampling frequency. When the sampling frequency satisfies the oversampling condition (Miao *et al.*, 2003a), the electron density map of the particle projected along the direction of the incident X-ray beam is, in principle, reconstructed from the oversampled diffraction amplitudes alone using phase-retrieval (PR) algorithms (Fienup, 1978, 1982).

Owing to the penetration power of short-wavelength X-rays, XDI is advantageous for visualizing whole structures of non-crystalline particles without sectioning and chemical labeling. Therefore, XDI has been applied to structural analyses of non-crystalline particles in material sciences and biology by using synchrotron radiation (SR) X-rays (Williams *et al.*, 2003; Shapiro *et al.*, 2005; Miao *et al.*, 2006; Nishino *et al.*, 2009; Jiang *et al.*, 2010; Takayama & Nakasako, 2012; Nam *et al.*, 2013; Takayama *et al.*, 2018; Kobayashi *et al.*, 2018a) and



X-ray free-electron laser (XFEL) pulses (Seibert *et al.*, 2011; Loh *et al.*, 2012; Nakasako *et al.*, 2013; Takahashi *et al.*, 2013; Gallagher-Jones *et al.*, 2014; Hantke *et al.*, 2014; Xu *et al.*, 2014; Kimura *et al.*, 2014; Oroguchi *et al.*, 2015; Takayama *et al.*, 2015a; van der Schot *et al.*, 2015; Ekeberg *et al.*, 2015; Kobayashi *et al.*, 2016a; Kameda *et al.*, 2017; Oroguchi *et al.*, 2018; Nakasako, 2018; Nakasako *et al.*, 2020; Ayer *et al.*, 2021; Cho *et al.*, 2021; Kobayashi *et al.*, 2021; Uezu *et al.*, 2023). The resolutions of the structure analyses were several tens of nanometres and sometimes reached several nanometres.

When the diffraction pattern of a particle is recorded without any pattern loss or noise, the projected electron density map of the particle can be, in principle, retrieved (Barakat & Newsam, 1984). However, PR calculations for experimental diffraction patterns frequently give non-realistic maps due to the loss of small-angle regions hidden by the beamstop and Poisson noise in X-ray detection (Huang *et al.*, 2010; Kobayashi *et al.*, 2014; Takayama *et al.*, 2015b; Sekiguchi *et al.*, 2016, 2017). In particular, in our experiences the loss of larger small-angle regions makes the convergence of PR calculations to realistic maps more difficult (Kobayashi *et al.*,

2014), because the small-angle regions contain structural information on the overall shape and the total electrons of the particle.

As typical examples, we show PR calculations for single-shot diffraction patterns from clusters of colloidal gold particles in Fig. 1 (Sekiguchi *et al.*, 2017). In the diffraction pattern in Fig. 1(a), the area of the lost small-angle region was more than three times the reciprocal of the cluster size, and the PR calculations yielded realistic maps with approximate probabilities of 50% [Fig. 1(b)]. The diffraction pattern of Fig. 1(c) loses the small-angle region as well as the pattern in Fig. 1(a), and signal-to-noise ratios beyond $10 \mu\text{m}^{-1}$ were smaller than those in Fig. 1(a). As a result, the probability yielding realistic maps was approximately 30% [Fig. 1(d)].

Therefore, to efficiently perform XDI structural analysis, protocols and/or metrics are necessary for screening the retrieved maps from a number of independently performed PR calculations (Chen *et al.*, 2007; Martin *et al.*, 2012; Park *et al.*, 2013; Rodriguez *et al.*, 2013; Kobayashi *et al.*, 2014; van der Schot *et al.*, 2015; Ekeberg *et al.*, 2015; Sekiguchi *et al.*, 2016; Favre-Nicolin *et al.*, 2020). For screening maps (Sekiguchi *et*

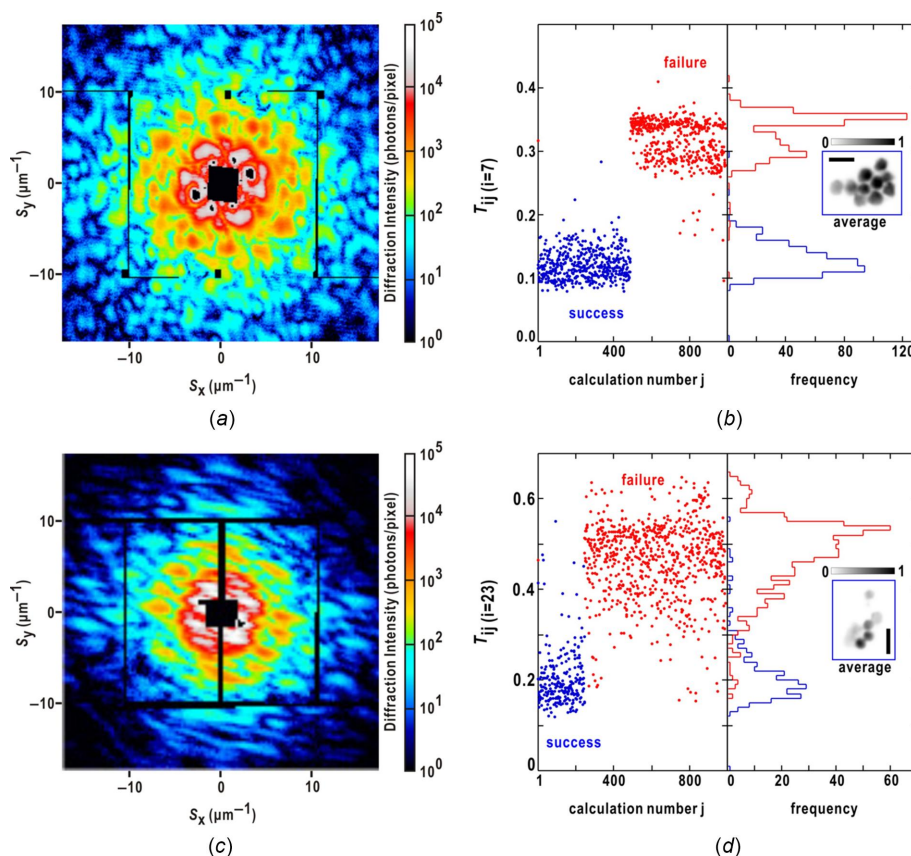


Figure 1

(a) Single-shot diffraction pattern from an aggregate of ten colloidal gold particles with 250 nm diameter (Sekiguchi *et al.*, 2016, 2017). The black square at the center is the shadow of the beamstop, and the stripes in the horizontal and vertical directions are the gaps between the CCD panels of the detectors. (b) The similarity scores of 1000 retrieved maps against the reference (left plot), and the frequency distribution of the scores of 999 maps (right). The reference map was obtained from the seventh trial ($i = 7$) of the 1000 PR calculations. The symbols and lines for correct and incorrect maps are colored in blue and red, respectively. The inset shows electron density maps averaged over the correct maps. (c) Single-shot diffraction pattern from an aggregate of colloidal gold particles. (d) Distribution of the similarity scores of 1000 maps retrieved from the diffraction pattern in panel (c). The reference map was obtained from the 23rd trial ($i = 23$) of the 1000 PR calculations. In panels (b) and (d), calculation trials were numbered after manual classifications judging whether the final maps were realistic. Panels (a) and (c) are reused from our previous paper (Sekiguchi *et al.*, 2017) after modification with permission from International Union of Crystallography.

al., 2017), we propose the use of the similarity score (Miao *et al.*, 2003b), defined as

$$T_{ij} = \frac{\sum_{x,y} |\rho_i(x, y) - \rho_j(x, y)|}{\sum_{x,y} |\rho_i(x, y) + \rho_j(x, y)|}, \quad (1)$$

where $\rho_i(x, y)$ denotes the projected electron density of the i th map. As XDI theory is formulated based on the first Born-approximation on a weak-phase object, $\rho_i(x, y)$ actually represents the projected electron density and not the complex transmission function formulated in ptychography (Pfeiffer, 2018).

Our previous structural analyses on the aggregates of colloidal metal particles (Sekiguchi *et al.*, 2017; Oroguchi *et al.*, 2018) showed that probable images were obtained with a high probability when each pair of the retrieved maps among a set of independently retrieved maps displayed similarity scores smaller than a threshold value (0.2) (Sekiguchi *et al.*, 2017) (Fig. 1). In contrast, when PR calculations failed, the resultant maps were inconsistent with each other and yielded similarity scores greater than the threshold value. The threshold value acts as a necessary condition to extract probable maps, although it is insufficient for the purpose because the score becomes smaller than the threshold value even for pairs of failed maps with similar structures. Owing to the success and usefulness in screening out non-realistic maps, we have been using the similarity score for XDI structural analyses (Kobayashi *et al.*, 2016a,b; Sekiguchi *et al.*, 2017; Oroguchi *et al.*, 2018; Nakasako *et al.*, 2020; Kobayashi *et al.*, 2021). However, it is unclear as to why the similarity score works as a good metric for screening retrieved maps.

In this study, we investigate how the similarity score works in the screening of retrieved maps in XDI. We describe equation (1) in reciprocal space to understand the correlation between the quality of the retrieved maps, and characterize the similarity score in terms of phase differences between the structure factors of the maps. Through screening the maps retrieved from diffraction patterns obtained in XFEL-XDI

and SR-XDI experiments, we assess the usefulness of the score and the resolution-dependent contributions of phase differences to the score.

2. Theoretical background

2.1. Interpretation of the similarity score in reciprocal space

2.1.1. Expression of the similarity score by Fourier transform. To characterize the similarity score in reciprocal space, we express the i th projection map, $\rho_i(x, y)$ in equation (1), as the inverse Fourier transform of the structure factor, $F_i(\mathbf{S})$, where \mathbf{S} is the scattering vector. Equation (1) in reciprocal space can be written as

$$T_{ij} = \frac{\sum_{\mathbf{r}} |\int \{F_i(\mathbf{S}) - F_j(\mathbf{S})\} \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S|}{\sum_{\mathbf{r}} |\int \{F_i(\mathbf{S}) + F_j(\mathbf{S})\} \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S|}. \quad (2)$$

According to the PR algorithm (Fienup, 1982), the final map is given by the inverse Fourier transform of the structure factor composed of the observed structure amplitude, $|F_{\text{obs}}(\mathbf{S})|$, and the phase in the final calculation cycle, $\alpha_i(\mathbf{S})$. Therefore, equation (2) is modified as

$$T_{ij} = \frac{\sum_{\mathbf{r}} |\int |F_{\text{obs}}(\mathbf{S})| \{\exp[i\alpha_i(\mathbf{S})] - \exp[i\alpha_j(\mathbf{S})]\} \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S|}{\sum_{\mathbf{r}} |\int |F_{\text{obs}}(\mathbf{S})| \{\exp[i\alpha_i(\mathbf{S})] + \exp[i\alpha_j(\mathbf{S})]\} \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S|}. \quad (3)$$

Equation (3) indicates that the similarity score depends on the phase differences between the structure factors of the two maps.

We designate the term $\exp[i\alpha_i(\mathbf{S})] - \exp[i\alpha_j(\mathbf{S})]$ as the ‘lack of identity’ (LoI) [Fig. 2(a)], analogous to the lack of closure of structure factors in the multiple isomorphous replacement method (Blow & Crick, 1959). Using the following relation on the difference and sum of the phase terms [Fig. 2(b)],

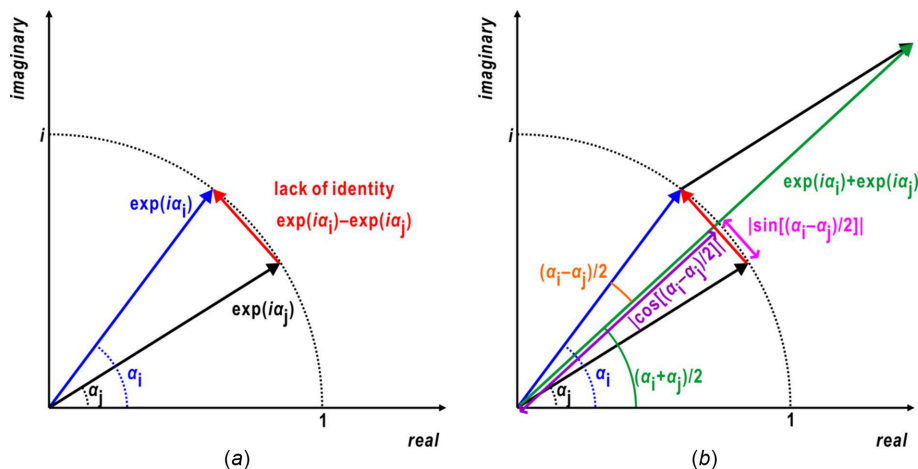


Figure 2

(a) Definition of the LoI in an Argand diagram. (b) Diagram showing the phase terms appearing in equation (4).

$$\begin{aligned} \exp[i\alpha_i(\mathbf{S})] - \exp[i\alpha_j(\mathbf{S})] &= \\ & 2i \sin\left[\frac{\alpha_i(\mathbf{S}) - \alpha_j(\mathbf{S})}{2}\right] \exp\left[i\frac{\alpha_i(\mathbf{S}) + \alpha_j(\mathbf{S})}{2}\right], \\ \exp[i\alpha_i(\mathbf{S})] + \exp[i\alpha_j(\mathbf{S})] &= \\ & 2 \cos\left[\frac{\alpha_i(\mathbf{S}) - \alpha_j(\mathbf{S})}{2}\right] \exp\left[i\frac{\alpha_i(\mathbf{S}) + \alpha_j(\mathbf{S})}{2}\right], \end{aligned} \quad (4)$$

then equation (3) is modified as

$$T_{ij} = \frac{\sum_{\mathbf{r}} \left| \int 2i \sin \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right|}{\sum_{\mathbf{r}} \left| \int 2 \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right|}, \quad (5)$$

$$\Delta\alpha_{ij}(\mathbf{S}) = \alpha_i(\mathbf{S}) - \alpha_j(\mathbf{S}) A(\mathbf{S}) = |F_{\text{obs}}(\mathbf{S})| \exp\left[i\frac{\alpha_i(\mathbf{S}) + \alpha_j(\mathbf{S})}{2}\right].$$

The trigonometric terms of $\Delta\alpha_{ij}(\mathbf{S})$ originating from the LoI act as weights on the structure factor $A(\mathbf{S})$, which is a product of $|F_{\text{obs}}(\mathbf{S})|$ and the phase averaged between the structure factors of the i th and j th maps. As the structure amplitudes at the lower resolution are greater than those at the higher resolution, the phase differences of the structure factors at the lower resolution dominate the score. With respect to two similar maps, the phase difference of the structure factors at low resolution will be as small as zero and reduce the numerator.

2.1.2. Resolution-dependent variation of the Fourier term.

As described above, the trigonometric terms of $\Delta\alpha_{ij}(\mathbf{S})$ and the structure amplitudes vary depending on the resolution. This implies that the similarity score can display a correlation with the resolution of the pair of maps, although the similarity score does not obviously depend on the resolution. Thus, it is meaningful if we modify each of the denominator and numerator in equation (5) into formulae suitable for monitoring the resolution-dependent variations of the trigonometric term. By using $|\int f(x) dx| \leq \int |f(x)| dx$, the numerator and denominator in equation (5) satisfy the following inequalities,

$$\begin{aligned} \left| \int 2i \sin \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right| &\leq \\ & \int \left| 2i \sin \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) \right| d^2 S, \\ \left| \int 2 \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right| &\leq \\ & \int \left| 2 \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) \right| d^2 S. \end{aligned}$$

Then, using the equation $|f(x)g(x)| = |f(x)||g(x)|$, the right-hand-side term in each of the above inequalities is expressed as the sum of the integrals on the N -divided resolution shells and each integrand is separated as

$$\begin{aligned} & \sum_{k=1}^N \int_{S_k \leq S < S_{k+1}} \left| 2i \sin \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) \right| d^2 S \\ &= \sum_{k=1}^N \int_{S_k \leq S < S_{k+1}} \left| 2i \sin \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} \right| |A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r})| d^2 S, \\ & \sum_{k=1}^N \int_{S_k \leq S < S_{k+1}} \left| 2 \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) \right| d^2 S \\ &= \sum_{k=1}^N \int_{S_k \leq S < S_{k+1}} \left| 2 \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} \right| |A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r})| d^2 S. \end{aligned} \quad (6)$$

Therefore, the separated trigonometric term may be helpful for numerically evaluating the influence of each resolution-shell to the similarity score.

2.1.3. Influences of phase differences in the resolution shell. Based on equation (6), to roughly estimate the influences of the phase differences in each resolution shell we use the following two quantities averaged over the maps in the narrow k th resolution shell,

$$\begin{aligned} & \int_{-\pi}^{+\pi} P(\Delta\alpha_k) \left| \cos \frac{\Delta\alpha_k}{2} \right| d(\Delta\alpha_k), \\ & \int_{-\pi}^{+\pi} P(\Delta\alpha_k) \left| \sin \frac{\Delta\alpha_k}{2} \right| d(\Delta\alpha_k), \\ & \int_{-\pi}^{+\pi} P(\Delta\alpha_k) d(\Delta\alpha_k) = 1, \end{aligned} \quad (7)$$

where $P(\Delta\alpha_k)$ is the normalized appearance frequency of $\Delta\alpha_k$. Hereafter, we designate the first and second equations as averaged cosine and sine terms in the k th shell, respectively.

Using equation (7), the influence of the trigonometric terms is qualitatively evaluated. For maps similar to each other, small phase differences between the structure factors in the low-resolution shell will yield values of the averaged cosine and sine term as large as 1 and 0, respectively. In contrast, for high-resolution shells, where the ambiguity of structure amplitudes due to Poisson noise causes large phase differences between the two structure factors, the averaged cosine term decreases, while the averaged sine term increases.

2.2. Reference values from special cases

As equation (6) is complicated, here we show values of the averaged cosine and sine terms in two special cases, which may be useful as references for roughly estimating the influences of the phase differences on the similarity score.

In the first special case, we assume that a particle has a circular shape with radius r_0 and uniform density. Then, the structure factor is written by the following equation (Born & Wolf, 2001),

$$F(S) = C(\pi r_0^2) \frac{2J_1(2\pi S r_0)}{2\pi S r_0},$$

where J_1 is the first-order cylindrical Bessel function and C is a constant. The phase value between any pair of adjacent zero-crossing points of J_1 is zero or π . When PR calculations for the diffraction amplitudes give two maps of circular shapes with different radius, the phase difference values are zero or π in any narrow resolution shell. Therefore, the phase differences between the two maps in the k th resolution shell are a constant.

In such a case, without the two inequalities introduced for the general case described above, the similarity score in the k th resolution shell is simplified as

$$T_{ij}(S_{k-1} < S < S_k) = \frac{\left| \sin \frac{\Delta\alpha_k}{2} \right| \sum_{\mathbf{r}} \left| \int_{S_k \leq S < S_{k+1}} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right|}{\left| \cos \frac{\Delta\alpha_k}{2} \right| \sum_{\mathbf{r}} \left| \int_{S_k \leq S < S_{k+1}} A(\mathbf{S}) \exp(-2\pi i \mathbf{S} \cdot \mathbf{r}) d^2 S \right|} = \frac{\left| \sin \frac{\Delta\alpha_k}{2} \right|}{\left| \cos \frac{\Delta\alpha_k}{2} \right|},$$

where $\Delta\alpha_k$ is a constant phase-difference value in the k th resolution shell. When $\Delta\alpha_k = 0$, the sine and cosine terms are 0 and 1, respectively, resulting in the similarity score in the resolution shell becoming 0. In contrast, when $\Delta\alpha_k = \pi$, the maximum phase difference, then the sine and cosine terms are 1 and 0, and the score diverges to infinity.

In this regard, we express an electron density distribution using the multipole expansion (Stuhrmann & Miller, 1978). Then, the electron density distribution is the sum of the first monopole term, the second dipole term, the third quadrupole and higher-order terms. If the weight for the first monopole term is approximated as uniform, the structure factor has the formula for a circular shape described above. In addition, when the monopole term approximating the size of the density distribution has a large electron density contrast (Ibel & Stuhrmann, 1975), the Fourier transform of the monopole term is dominant in the structure factor of the density distribution in the small-angle region. Therefore, this simple numerical calculation may be helpful to roughly evaluate the influences of the phase-difference terms in equation (6) on the similarity score.

In the next special case, we numerically calculate the averaged cosine and sine terms for the case that the frequency distribution of phase differences is random among the structure factors of retrieved maps in the k th resolution shell. Then, the values of the averaged cosine and sine terms approach 0.64 as

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \cos \left(\frac{\Delta\alpha_k}{2} \right) \right| d\Delta\alpha_k = \frac{2}{2\pi} \int_0^{\pi} \cos \left(\frac{\Delta\alpha_k}{2} \right) d\Delta\alpha_k = \frac{4}{2\pi}, \quad (8)$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sin \left(\frac{\Delta\alpha_k}{2} \right) \right| d\Delta\alpha_k = \frac{2}{2\pi} \int_0^{\pi} \sin \left(\frac{\Delta\alpha_k}{2} \right) d\Delta\alpha_k = \frac{4}{2\pi}.$$

Hereafter, this value is designated the random phase limit.

2.3. Figure of merit for estimated phase

In X-ray crystallography and transmission electron microscopy (TEM) of biological macromolecules, the reliability of a reconstructed map is evaluated using the figure of merit (FoM) (Blow & Crick, 1959), based on the probability distribution of phase values in the experimental estimation, as follows,

$$\text{FoM}(\mathbf{S}) = \left| \sum_k P[\alpha_k(\mathbf{S})] \exp[i\alpha_k(\mathbf{S})] \right|, \quad \sum_k P[\alpha_k(\mathbf{S})] = 1, \quad (9)$$

where $P[\alpha_k(\mathbf{S})]$ is the probability of the k th bin of the phase values at the scattering vector \mathbf{S} . The reconstructed maps displaying a FoM greater than the threshold of 0.5 are regarded as interpretable (Lunin & Woolfson, 1993; Perrakis *et al.*, 1997). The threshold indicates that the phase values are distributed within $\pm 60^\circ$ from the true value (Drenth, 1994). The FoM for the structure factors of two maps is in correlation with the LoI as

$$\text{FoM}(\mathbf{S}) = \cos \left(\frac{\Delta\alpha_{ij}(\mathbf{S})}{2} \right) \left| \exp \left[i \left(\frac{\alpha_i(\mathbf{S}) + \alpha_j(\mathbf{S})}{2} \right) \right] \right| = \cos \left[\frac{\Delta\alpha_{ij}(\mathbf{S})}{2} \right].$$

Thus, the similarity-score-selected maps were evaluated using FoM and LoI.

2.4. Phase-retrieval transfer function for estimated phase

In XDI structural analyses, the phase retrieval transfer function (PRTF) is frequently used to estimate the effective resolution for a set of retrieved maps (Chapman *et al.*, 2006). The PRTF for a set of retrieved maps is defined as

$$\text{PRTF}(S) = \left\langle \frac{1}{M} \left| \sum_{j=1}^M \frac{|F_{\text{obs}}(\mathbf{S})| \exp[i\alpha_j(\mathbf{S})]}{|F_{\text{obs}}(\mathbf{S})|} \right| \right\rangle_{\text{radial}} = \left\langle \frac{1}{M} \left| \sum_{j=1}^M \exp[i\alpha_j(\mathbf{S})] \right| \right\rangle_{\text{radial}} \quad (10)$$

where M is the number of retrieved maps incorporated into the calculation, and $\langle \dots \rangle_{\text{radial}}$ refers to circular averaging with respect to the scattering vector length. The effective resolution is determined from the reciprocal of the scattering vector length, where the PRTF curve decreased to a specified threshold value. The PRTF for the structure factors of the two retrieved maps is derived using equation (4) as

$$\text{PRTF}(S) = \left\langle \frac{1}{2} \left| 2 \cos \frac{\alpha_i(\mathbf{S}) - \alpha_j(\mathbf{S})}{2} \exp \left[i \frac{\alpha_i(\mathbf{S}) + \alpha_j(\mathbf{S})}{2} \right] \right| \right\rangle_{\text{radial}} = \left\langle \cos \frac{\Delta\alpha_{ij}(\mathbf{S})}{2} \right\rangle_{\text{radial}}$$

where we use the fact that the cosine term is positive for $-\pi \leq \Delta\alpha_{ij}(\mathbf{S}) \leq +\pi$. The effective resolution estimated using the PRTF will be discussed in conjunction with the similarity score in the *Discussion* section.

3. Experimental procedure and structure analysis

3.1. Specimen preparation

For XFEL-XDI experiments, colloidal gold particles with a mean diameter of 250 nm (BBI Solutions, UK) were dispersed on custom-made specimen disks, each of which had a 1 mm × 1 mm Si₃N₄ membrane window of 100 nm thickness (Norcada, Canada). Further details were described in our previous publications (Sekiguchi *et al.*, 2014a, 2016; Kobayashi *et al.*, 2016a,b; Nakasako, 2018; Nakasako *et al.*, 2020).

For the SR-XDI experiments, we used commercially available specimen disks with a 5 mm × 5 mm Si₃N₄ membrane window (Silson Ltd, UK). The windows were coated with ~15 nm-thick carbon layers using a JEE-420 vacuum evaporator (Jeol, Japan) and then covered with poly L-lysine layers (Sigma-Aldrich, USA) to assist the adhesion of the specimen particles to the Si₃N₄ membranes (Takayama & Yonekura, 2016; Kobayashi *et al.*, 2016b). We prepared specimen disks adsorbing colloidal gold particles with a mean diameter of 400 nm (Sigma-Aldrich, USA) or gold urchin particles synthesized according to the literature (Wang *et al.*, 2012).

3.2. XFEL-XDI experiments and data processing

XFEL-XDI experiments were performed using our custom-made apparatus KOTOBUKI-1 (Nakasako *et al.*, 2013) at beamline BL3 of SACLA (Tono *et al.*, 2013). Focused XFEL pulses with a photon energy of approximately 5.5 keV (corresponding to a wavelength of 0.225 nm) had intensities of 10¹⁰–10¹¹ photons pulse⁻¹ μm⁻² and almost complete spatial coherence (Kobayashi *et al.*, 2018a). The specimen mounted on the goniometer of KOTOBUKI-1 was scanned with a 50 μm step against the incident XFEL pulses, and single-shot diffraction patterns were recorded using multi-port CCD (MPCCD) octal and dual detectors (Kameshima *et al.*, 2014). After each scan, the recorded diffraction patterns were processed using the *G-SITENNO* program suite (Sekiguchi *et al.*, 2014a,b), which extracted diffraction patterns from a cluster of colloidal gold particles with a signal-to-noise ratio greater than 2 at a resolution of 15 μm⁻¹. Further details of the experimental setup, procedures and data processing were reported previously (Sekiguchi *et al.*, 2014a, 2016; Kobayashi *et al.*, 2016a,b; Nakasako, 2018; Nakasako *et al.*, 2020; Kobayashi *et al.*, 2021; Uezu *et al.*, 2023).

3.3. XDI experiments using synchrotron X-rays and data processing

The SR-XDI experiments were performed using an originally developed atmospheric XDI system (Takayama *et al.*, 2018, 2021) installed at the imaging station of BL24XU in SPring-8 (Takayama *et al.*, 2020). Specimen particles were irradiated with the nearly plane-wave region of the incident X-rays produced using a 30 μm pinhole, which was placed approximately 1.5 m upstream of the specimen position.

An aggregate of gold urchin particles was irradiated by X-rays with a photon energy of 8.000 keV (corresponding to a

wavelength of 0.1550 nm), and the diffraction pattern was recorded using a PILATUS 100k detector (Dectris Ltd, Switzerland) (Kraft *et al.*, 2009) placed approximately 4.1 m downstream of the specimen. An aggregate of 400 nm colloidal gold particles was irradiated by X-rays with a photon energy of 8.310 keV (corresponding to a wavelength of 0.1492 nm), and the diffraction pattern was recorded using an EIGER X 1M detector (Dectris Ltd, Switzerland) (Dinapoli *et al.*, 2011) placed approximately 3.2 m downstream of the specimen.

For each specimen, a background pattern was recorded from an area of the Si₃N₄ membrane yielding no speckles, and was subtracted from the diffraction pattern of the specimen. Centro-symmetric averaging was applied to the background-subtracted diffraction pattern to improve the signal-to-noise ratio owing to good Friedel symmetry.

In the SR-XDI experiments described above, as the coherent flux of SR X-rays from the undulator of BL24XU was around 0.1% due to the size of the light source (electron bunch) in the SPring-8 storage ring (approximately 300 μm along the horizontal direction), we extracted the coherent part of the incident X-rays using a pinhole with a diameter of 30 μm, and a specimen was irradiated by the plane-wave part at the peak area of the diffraction from the pinhole (Takayama *et al.*, 2018). In this case, the background scattering predominantly came from the Si₃N₄ membrane irradiated by the foot region of the pinhole-diffracted X-rays, where the spatial coherence was low. The electron density in the membrane was assumed to be uniform at the maximum resolution of the experiment. In addition, the background scattering from the membrane was assumed to be incoherent and the interference between the background scattering and the diffraction wave from the specimen was negligible. Therefore, the background subtraction procedure was routinely applied as reported for SR-XDI experiments at BL24XU (Takayama *et al.*, 2018) and BL29XUL (Nishino *et al.*, 2009; Nam *et al.*, 2013; Kobayashi *et al.*, 2018b) at SPring-8 and also for XDI experiments at BL9C at Pohang Light Source II (Ahn *et al.*, 2021).

3.4. PR calculation and multivariate analysis

We performed PR calculations for the diffraction patterns using a custom-made Python3 software, in which the hybrid-input-output (HIO) (Fienup, 1982), shrink-wrap (SW) (Marchesini *et al.*, 2003) and error reduction (ER) algorithms (Gerchberg & Saxton, 1972) were implemented. Each of 1000 independent PR calculations started with a random-valued map different among the 1000 calculations. An initial support was set as the area where the magnitude of the real-space autocorrelation function calculated from the diffraction pattern was greater than 4% of the maximum. Each projected electron density map was retrieved by 10000 iterations of the HIO cycles with SW modifications at every 100 HIO-cycles and additional 1000 iterations of the ER calculation. Details of HIO and SW were reported previously (Kodama & Nakasako, 2011; Oroguchi & Nakasako, 2013).

Based on the assumption that realistic maps were obtained at high frequency and similar to each other when a PR calculation is successful, we searched a reference map for a calculation of the similarity score among the set of retrieved maps as follows. The 1000 retrieved maps were first translated to coincide the center of gravity of the density with that of an arbitrarily selected reference map, and we calculated the similarity score for whole combinations of 1000 maps (499 500 pairs). Then, we selected the pair of maps yielding the smallest similarity score as the most probable pair, and one of the pair was used as a reference map for the subsequent calculation. Each map was translated to maximize phase-only correlation (Kuglin & Hines, 1975) against the reference map. In the PR calculation, several maps in the π -rotation relative to the reference map appeared owing to the centrosymmetry of the diffraction patterns. These maps were corrected with respect to π -rotation during translational alignment using the phase-only correlation. After the correction, the similarity scores for whole combinations of 1000 maps were again calculated against the reference. It should be noted that either of the map pair gave almost the same result of the calculated similarity scores.

Independently, the maps were classified into ten groups by K -means clustering (MacQueen, 1967) without the information from the similarity score. When assuming that the maps are classified into ten groups, we minimize the sum of squared distances between the maps and the centroids of the groups defined as

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

where \mathbf{y}_{im} is a vector indicating the position of the i th map belonging to the m th group and $\langle \mathbf{y}_m \rangle$ is the centroid of the m th group. For visualizing the distributions of the classified ten groups, the map positions were projected onto a plane spanned by the two principal component (PC) vectors, which had the largest and second largest eigenvalues in the principal component analysis (PCA) (Jolliffe & Cadima, 2016) applied to the 1000 retrieved maps. It should be noted that, on the plane spanned by the first and second PC vectors, clusters separated along the third-order or higher-order vectors will be overlaid because the K -means clustering was applied before the dimensionality reduction by PCA. Details of the PCA for a number of PR maps were reported previously (Sekiguchi *et al.*, 2016). The *scikit-learn* library (Pedregosa *et al.*, 2011) was used for the K -means clustering and PCA.

Finally, in each class, the phase distributions among the structure factors of the maps were evaluated using the FoM at every scattering vector [equation (9)] (Sekiguchi *et al.*, 2017), and the phase differences in the structure factors between any pair of maps were represented using the averaged cosine and sine terms of equation (7) in each resolution shell.

4. Results

Here, we performed the screening of maps phase-retrieved from experimental diffraction patterns using the similarity

score in real space and interpreted the score in terms of phase differences between the structure factors of the retrieved maps. In the analyses, we targeted diffraction patterns from aggregates of colloidal gold particles, because the known size and shape of the particles made identifying realistic maps easy.

4.1. Application to single-shot diffraction patterns recorded using XFEL pulses

Here, we assessed the similarity score for the maps retrieved from single-shot diffraction patterns recorded using XFEL pulses (Fig. 1). As a lot of single-shot diffraction patterns can be collected within a short time in our XFEL-XDI experiments (Kobayashi *et al.*, 2016a; Nakasako, 2018), efficient and automatic screening of probable and non-realistic maps is necessary in the structure analyses.

One thousand maps were retrieved from the diffraction pattern of an aggregate of ten colloidal gold particles in Fig. 1(a), and were divided into ten classes on the plane spanned by the first and second PC vectors [Fig. 3(a)]. The maps in classes 2, 5, 8 and 9 forming a cluster on the plane displayed similarity scores smaller than 0.19 against the reference map, which was one of the pair yielding the smallest similarity score among all the maps [Fig. 3(b)].

Here, we define a quantity in reciprocal space designated the Fourier error of a retrieved map as

$$\text{Fourier error} = \frac{\sum_{S_x, S_y} \left| |F_{\text{obs}}(S_x, S_y)| - |F_{\text{cal}}(S_x, S_y)| \right|^2}{\sum_{S_x, S_y} |F_{\text{obs}}(S_x, S_y)|^2},$$

where $F_{\text{cal}}(S_x, S_y)$ is the structure factor of a retrieved map masked by the support. Because this quantity is related to the target function of the phase retrieval to be reduced, we examined whether this quantity for each map was useful for extracting realistic maps. In this case, as the frequency distributions of the Fourier-error values of the ten classes were heavily overlapped [Fig. 3(c)], we abandoned the use of this quantity for extracting maps of classes 2, 5, 8 and 9 [see Fig. 3(d)].

Fig. 3(d) depicts the map in each of the ten classes, yielding the smallest similarity score against the reference map. In the maps of classes 2, 5, 8 and 9, the ten colloidal gold particles appeared with clear borders and were separated from each other, indicating that the maps in the four classes were successfully retrieved. Among the four correct classes, the maps of class 8 yielded the smallest similarity scores [Fig. 3(b)]. The PR calculations for the other six classes failed, as judged by the unclear and blurred electron densities of the particles. These results imply that the similarity score is a good measure for screening maps in real space. As indicated by the similarity scores, the probable maps were similar to each other, whereas non-realistic maps were different from each other [Fig. 3(b)].

The similarity of the maps in real space was evaluated using the FoM [Fig. 3(e)] and LoI [Fig. 3(f)] in reciprocal space. In the four successfully retrieved classes, particularly classes 2

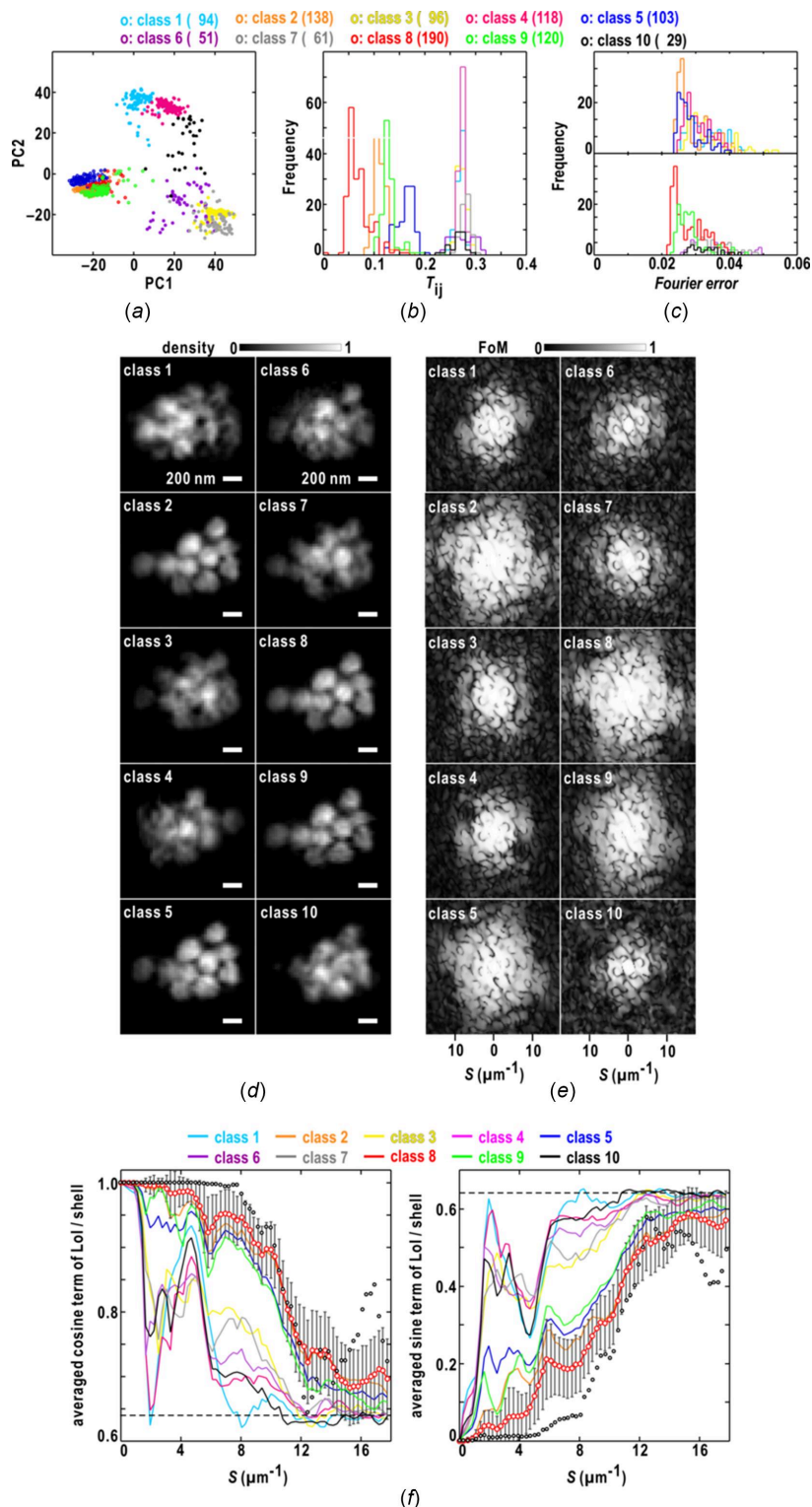


Figure 3

Characterization of the 1000 maps retrieved from the single-shot diffraction pattern in Fig. 1(a). (a) Distribution of the 1000 maps divided into ten classes on the plane spanned by the first and second PC vectors. The symbols indicate the positions of the maps on the plane and are colored according to the scheme at the top of the panel. The coloring scheme for the ten classes is used throughout the panels. The number of the map in each class is shown in parentheses. Frequency distributions of the similarity score (b) and the Fourier error (c) of the maps in each class. (d) Comparison of class-representative maps displaying the smallest similarity score against the reference in each class. The densities are illustrated according to the bar at the top of the panel. (e) FoMs of structure factors in the ten classes up to a resolution of $18 \mu\text{m}^{-1}$. The FoM values are shown according to the bar at the top of the panel. (f) Resolution-dependences of the averaged cosine (left panel) and sine (right panel) terms in equation (7) in each class. The curves of class 8 of the two terms are depicted using red open circles and error bars for the standard deviations of the two terms. The open black circles are the two terms for only the pair of maps yielding the smallest similarity scores among the 1000 maps. The dashed line indicates the values of the two terms for the random phase limit [equation (8)].

and 8, the reciprocal regions with FoM values greater than 0.5 extended to a resolution of $18 \mu\text{m}^{-1}$. In contrast, the FoM values of the non-realistic classes were smaller than 0.5, with a resolution lower than $10 \mu\text{m}^{-1}$.

Fig. 3(f) shows the resolution-dependent variations of the cosine and sine terms of equation (7). For the pair of maps yielding the smallest similarity score among the 1000 phase-retrieved maps, the phase differences gradually approached the random phase limit [equation (8)] beyond a resolution of $16 \mu\text{m}^{-1}$. In the realistic classes, the phase differences likely varied in the range $20\text{--}30^\circ$ among the maps up to a resolution of $8 \mu\text{m}^{-1}$. In the non-realistic classes, judging from the values close to the random phase limit of 0.64 [see equation (8)], the phase differences became random even at a resolution of approximately $2 \mu\text{m}^{-1}$; this might be a major cause of blurry densities and unclear edges of the averaged maps. It should be noted that there is a boundary of the data-missing area due to the beamstop around the resolution of approximately $2 \mu\text{m}^{-1}$ [Fig. 1(a)].

Next, we assessed the similarity score for the maps retrieved from another single-shot diffraction pattern shown in Fig. 1(c). Among the ten classes of the 1000 retrieved maps [Fig. 4(a)], only the maps in class 3 showed similarity scores smaller than 0.18 against the reference map [Fig. 4(b)]. In this case, the frequency distribution of the Fourier-error values for maps in class 3 had a peak at 0.026 [Fig. 4(c)], but, in contrast to the similarity score [Fig. 4(b)], the overlap of the frequency distributions among the ten classes made it difficult to extract most of the maps in class 3. Fig. 4(d) depicts the map in each of the ten classes displaying the smallest similarity score against the reference. In the class 3 map [Fig. 4(d)], the three major and three minor densities of colloidal particles were clear, and the minor density at the upper edge was interpreted as the particle being located in the foot part of the XFEL pulse. In the maps of classes 1, 7 and 9, which yielded similarity scores greater than 0.2, the edges of the three major particles were unclear and blurred, indicating failures of the PR calculations. Regarding the FoM of the structure factors of the maps in each class [Fig. 4(e)], classes 3 and 8 maintained FoM values greater than 0.5 beyond a resolution of $10 \mu\text{m}^{-1}$. In classes 2, 5, 6 and 10, the FoM values were direction-dependent, suggesting that the maps in the four classes had anisotropic similarity.

As shown in Fig. 4(f), the phase differences between the pair of maps yielding the best similarity score approached the random phase limit [equation (8)] beyond a resolution of $14 \mu\text{m}^{-1}$. In contrast, the maps in each of classes 2, 4, 5, 6, 8 and 10 displayed sudden changes in the cosine and sine terms in the resolution range $1\text{--}2 \mu\text{m}^{-1}$, around the boundary of the data-missing area due to the beamstop [Fig. 1(c)]. The two terms for the maps in each of classes 1, 7 and 9 indicated greater phase differences of structure factors up to a resolution of $17 \mu\text{m}^{-1}$.

Based on the results, we concluded that the similarity score is good for screening maps phase-retrieved from single-shot diffraction patterns. The similarity score in real space was more convenient for automatically screening maps than the FoM and the two trigonometric terms in reciprocal space

[equation (7)]. The trigonometric terms in reciprocal space are important to explain how the similarity score works.

4.2. Application to diffraction patterns recorded using synchrotron X-rays

In this section, we apply the similarity score to the maps retrieved from the diffraction patterns of clusters composed of several colloidal gold particles recorded in the SR-XDI experiments. In the SR-XDI experiment, the collection of diffraction patterns from specimens with sizes of several micrometres is possible owing to the wider spatially coherent irradiation area of approximately $10 \mu\text{m}$ diameter and the smaller beamstop covering a much smaller diffraction angle than those in XFEL-XDI (Takayama *et al.*, 2018; Kobayashi *et al.*, 2018b). In addition, good signal-to-noise ratios in high-diffraction-angle regions might be expected owing to both the relatively large scattering cross-section of specimens with sizes of several micrometres and the exposure time.

In Fig. 5, we summarize the application of the similarity score for the maps retrieved from the diffraction pattern of an aggregate composed of more than 20 colloidal gold particles of 400 nm diameter. Of the ten classes divided from the 1000 retrieved maps, only the maps of class 2 were localized in the plane spanned by the two PC vectors [Fig. 5(b)]. Fig. 5(b) depicts the map in each of the ten classes, yielding the smallest similarity score against the reference map. Only the maps of class 2 were realistic, as judged from the clear borders and void-like fine structures of the particles, and yielded a similarity score smaller than 0.08 against the reference [Fig. 5(c)]. As the Fourier-error values of the maps in class 2 displayed a single peak in the frequency distribution [Fig. 5(d)] and correlated with the similarity score values, the Fourier error may be used to extract the class 2 maps in this case.

Fig. 5(e) shows the resolution-dependent variations of the averaged cosine and sine terms of equation (7). For a pair of maps yielding the smallest similarity score, the phase differences in the structure factors of the maps gradually approached the random phase limit beyond a resolution of $30 \mu\text{m}^{-1}$ [equation (8)]. The two terms of class 2 also asymptotically reached this limit beyond $30 \mu\text{m}^{-1}$, while those of the other non-realistic classes reached the value within $5 \mu\text{m}^{-1}$.

As a second example, we assessed the similarity score for maps retrieved from the diffraction pattern of an aggregate of gold urchin particles with irregular shapes (Fig. 6). As the diffraction pattern included a part of the central speckle in the very small angle region and was recorded with good signal-to-noise ratios up to a resolution of $25 \mu\text{m}^{-1}$ [Fig. 6(a)], almost all of the maps were successfully retrieved [Fig. 6(b)]. Although the correct maps displayed similarity scores smaller than 0.19, the maps were divided into classes 2–10, probably because of the small differences in the local fine structures, such as convections and protrusions on the surfaces [Figs. 6(b) and 6(c)]. All of the class 2 maps yielded similarity scores smaller than 0.09 against the reference. The Fourier error values of all maps were approximately 0.0241. As the values of the non-realistic maps of class 1 overlapped with the frequency

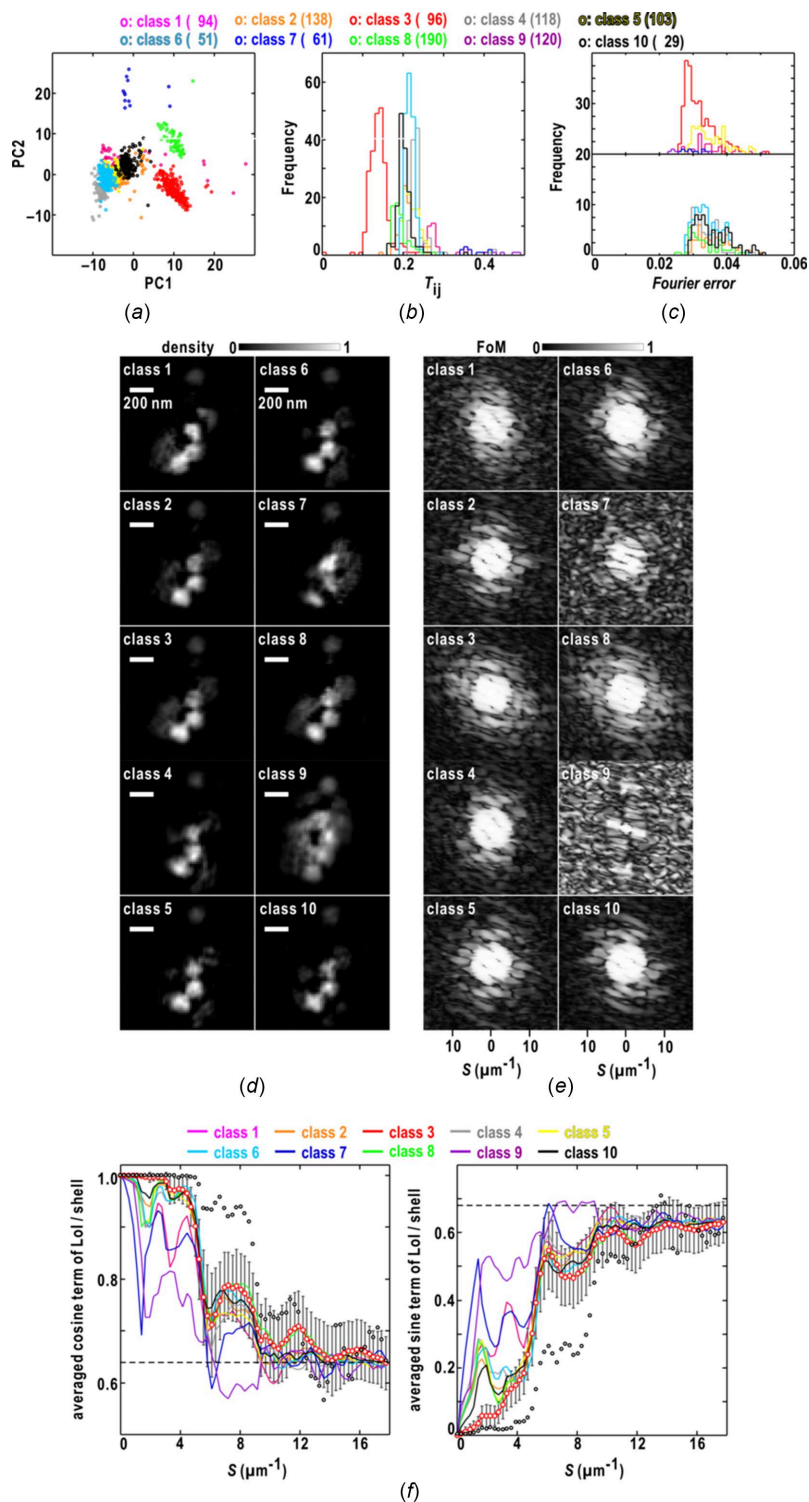


Figure 4 Characterization of the 1000 maps retrieved from the single-shot diffraction pattern in Fig. 1(c). Panels (a)–(f) are illustrated as in Fig. 3. In panel (f), showing the resolution-dependences of the averaged cosine (left panel) and sine (right) terms in equation (7), the curves of class 3 of the averaged terms are depicted using red open circles and error bars for the standard deviations of the two terms.

distributions of the other class, it was difficult to exclude the Fourier errors in the maps of class 1 [Fig. 6(d)].

In Fig. 6(e), for the two maps yielding the best similarity score, the phase differences of the structure factors were still non-random beyond a resolution of $25 \mu\text{m}^{-1}$. For only

the maps of class 2, the resolution-dependent variations of the cosine and sine terms of equation (7) were similar to those of the best pair and displayed no sudden changes up to a resolution of $5 \mu\text{m}^{-1}$. Beyond $23 \mu\text{m}^{-1}$, the two terms reached the random phase limit of 0.64 [equation (8)], probably

because the phase differences among the maps became random.

From the structural analyses of the two diffraction patterns in the SR-XDI experiments, the similarity score in real space

acts as a good measure in screening retrieved maps, and the trigonometric terms of equation (7) in reciprocal space were helpful for understanding the correlation between the similarity score and the validity of the maps.

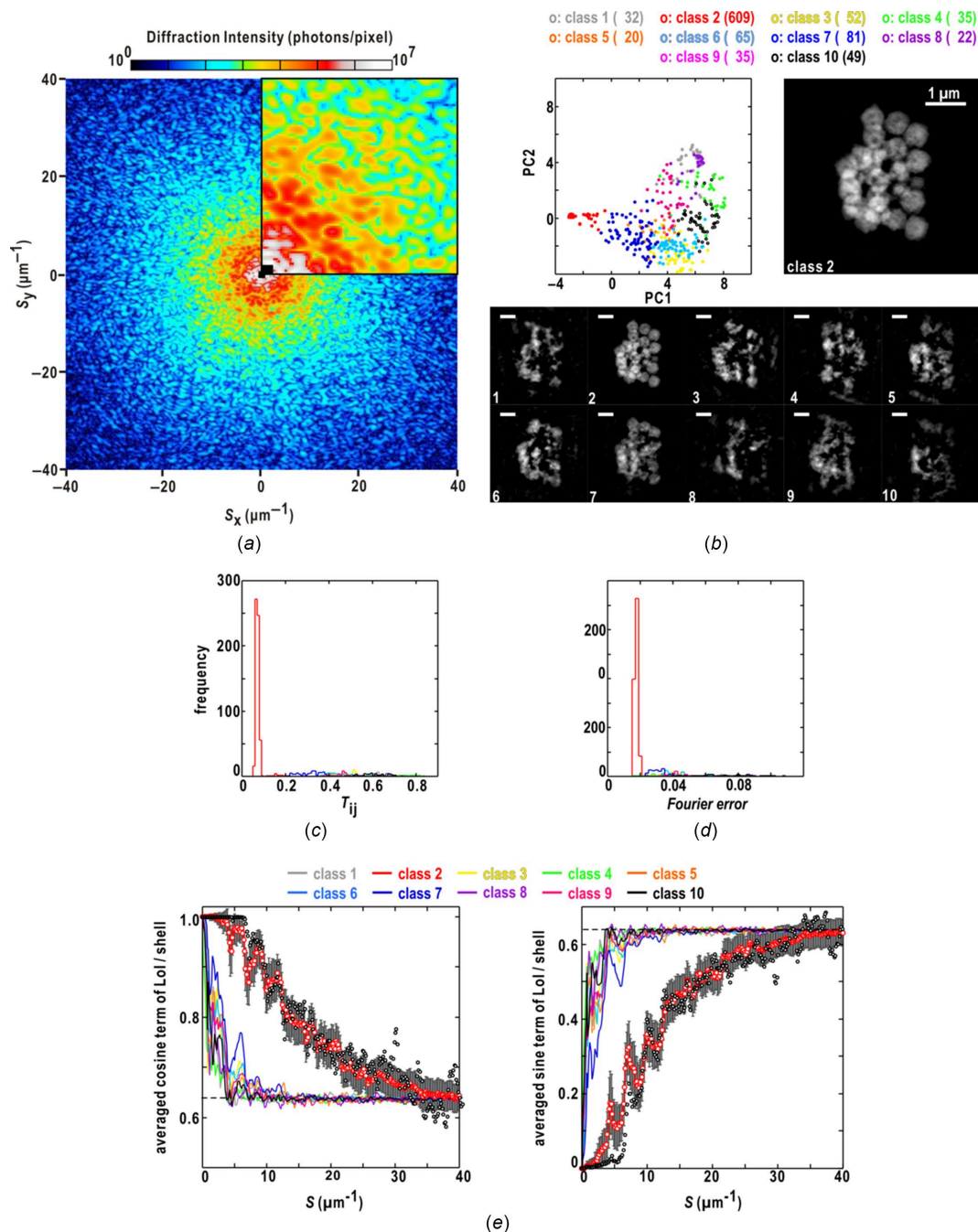


Figure 5

Characterization of the 1000 maps retrieved from the diffraction pattern of an aggregate of more than 20 colloidal gold particles of 400 nm diameter recorded in the SR-XDI experiment. (a) Diffraction pattern from the aggregate. The upper-right quadrant is a magnified view of the small-angle region up to $10 \mu\text{m}^{-1}$. (b) Distribution of the retrieved maps in the ten classes on the plane spanned by the first and second PC vectors. The symbols indicate the positions of the maps on the plane and are colored according to the scheme at the top of the panel. The number of the maps in each class are shown in the parentheses. The enlarged map at the top right is the reference map, which is one of a pair yielding the smallest similarity score among all the maps. The bottom panel compares the class-representative maps displaying the smallest similarity score against the reference in each class. Frequency distributions of the similarity score (c) and the Fourier error (d) of the maps in each class. (e) Resolution-dependences of the averaged cosine (left) and sine (right) terms of equation (7) in each class. The curves of class 2 of the two terms are depicted using red open circles and error bars for the standard deviations of the two terms. The open circles are the two terms for the pair of maps yielding the smallest similarity scores among the 1000 maps. The dashed line indicates the values of the two terms for the random phase limit [equation (8)].

5. Discussion

In this study, we characterized the similarity score in correlation with the phase differences in the structure factors of the retrieved maps. Through application to structural analyses for experimental diffraction patterns, we demonstrated the usefulness of the similarity score for screening phase-retrieved maps. Here, we discuss the weak point in the use of the similarity score, the estimation of the effective resolution of the retrieved maps using the score, and the utilization of the score for efficiently performing PR calculations.

5.1. Selection of maps using the similarity score

In this study, for selecting realistic maps, which approximate the shape, size and internal structure of a particle, we use the similarity score under the assumption that realistic maps yield the smallest similarity score among the retrieved maps. However, when pairs of non-realistic maps different from the structures of particles are quite similar, this assumption cannot be made. In our experience in structure analyses of aggregates of colloidal gold particles, the breakdown of the assumption occurs at a probability smaller than 0.1%. However, cross-

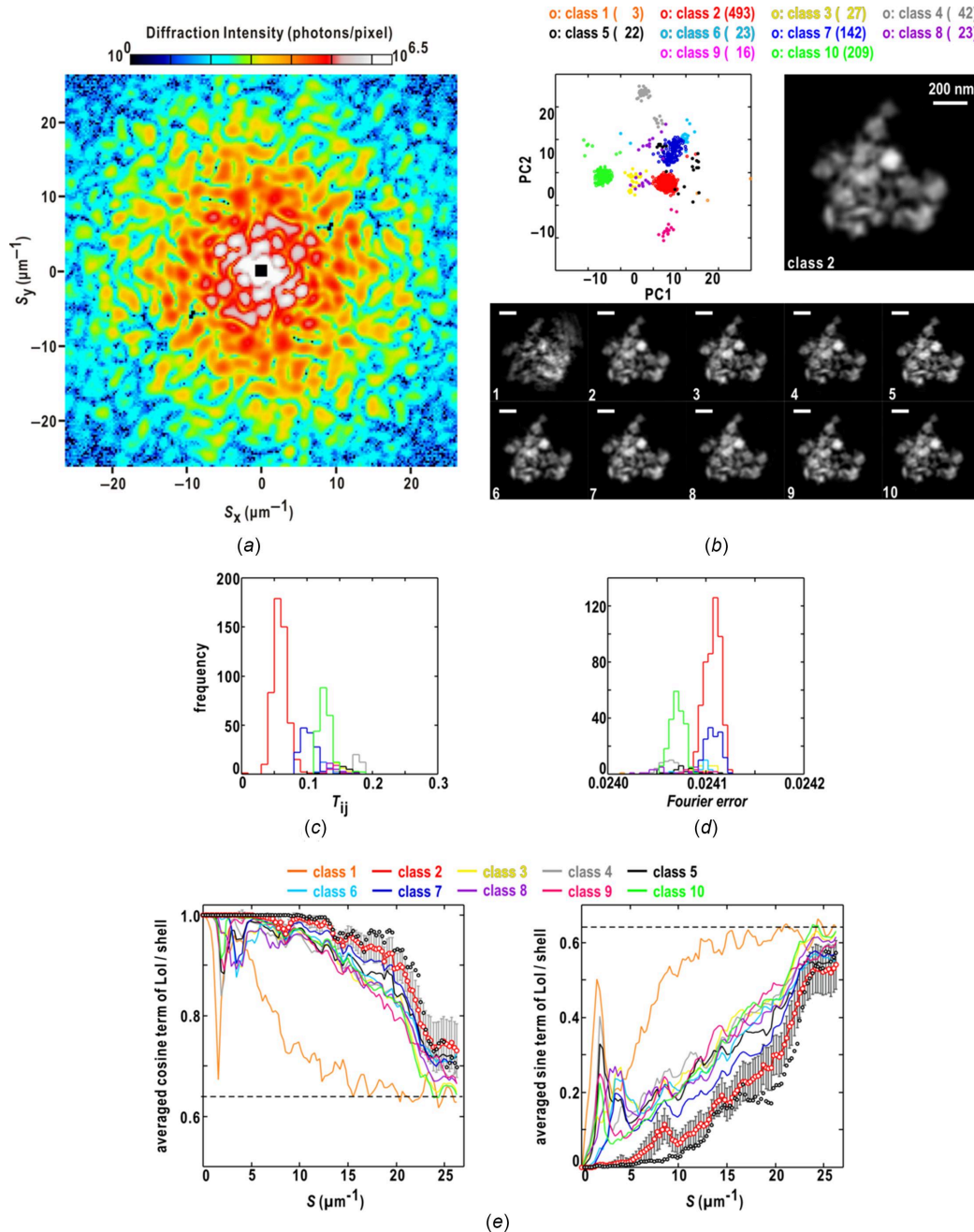


Figure 6 Characterization of the 1000 maps retrieved from the diffraction pattern of a cluster of gold urchin particles recorded in the SR-XDI experiment. All panels are illustrated as in Fig. 5. The retrieved map at the top right in panel (b) is enlarged to clearly show the spikes of the particles. In panel (e), showing the resolution-dependences of the averaged cosine (left) and sine (right) terms in equation (7), the curves of class 2 of the two terms are depicted using red open circles and error bars for the standard deviations of the two terms.

validation using known structural information, for instance from electron microscopy (EM) and fluorescence microscopy, is still necessary for the selected maps, as carried out in our previous works (Kobayashi *et al.*, 2014; Takayama *et al.*, 2015a; Oroguchi *et al.*, 2018; Kobayashi *et al.*, 2021).

In this regard, we describe two examples on cross-validation in the structure determination of protein molecules. Even in the estimation of phase set using the multiple heavy-atom replacement method in protein crystallography, we validated the heavy-atom positions in the unit cell of derivative crystals using the phase set in the centric zone of the Bragg reflections obtained from cryogenic TEM (Toyoshima *et al.*, 2000).

In small-angle X-ray scattering for proteins in solution, the molecular shapes of protein molecules are predicted using an *ab initio* algorithm (Svergun *et al.*, 2001). As the algorithm is applied to a one-dimensional scattering profile under restraints with respect to the arrangement of small spheres approximating the molecular shape, both realistic and non-realistic models appear in more than 500 trial calculations. Then, after classification of the models using multivariate analysis (Oide *et al.*, 2018), cross-validation using structural information from other imaging techniques is indispensable for finding realistic molecular structures (Oide *et al.*, 2020; Oide & Nakasako, 2021).

The similarity score may be a useful tool for screening realistic maps from a number of trial calculations. However, as we have no experimentally determined phase data in XDI, we keep in mind the possibility of the breakdown of the assumption. Therefore, even for maps yielding small similarity scores, validation for the shape, size and internal structure of the particle is always necessary.

5.2. Similarity score and differences in structure factors of maps at low resolution

In the four examples of structure analyses (Figs. 3–6), the frequency distributions of the similarity scores of phase-retrieved maps in each class are likely correlated with the resolution-dependent variations of the cosine and sine terms for the phase differences of the structure factors between the maps. When the similarity scores in a class are distributed in a range smaller than 0.2, the cosine term is close to 1 and the sine term is near 0 in the small-angle region [Figs. 3(*f*), 4(*f*), 5(*e*) and 6(*e*)]. Owing to the structure amplitudes of the small-angle region, which are much greater than those of the high-angle region [Figs. 1(*a*), 1(*c*), 5(*a*) and 6(*a*)], when the phase differences of the structure factors of maps are small in the small-angle region, the contribution of the structure factors to the numerator of equation (6) is reduced. Therefore, the small value of the similarity score is interpreted as the retrieved maps viewed at low resolution being similar to each other.

As the contribution of fine structures to the structure factors are dominant at high resolution, the screening of retrieved maps using the similarity score predominantly focuses on the low-resolution structures of the retrieved maps. At low resolution, the FoM takes values close to 1 [Figs. 3(*e*) and 4(*e*)], as expected from equation (9). Therefore, the

similarity score may be useful in the preliminary screening of realistic maps before applying, for instance, the unweighted pair group method with arithmetic mean analysis, which is the distance-matrix method and employs sequential clustering to yield a rooted phylogenetic tree of image data as used in XFEL-XDI structure analysis (Ekeberg *et al.*, 2015).

5.3. Effective resolution estimated using the similarity score

As the similarity score is dominated by the phase differences between structure factors of retrieved maps in the small-angle region, we discuss how the effective resolution for a set of retrieved maps is quantitatively estimated using the resolution-dependent variation of the cosine (or sine) term in equation (7). Then, we referred to the fine structures of the averaged maps, such as resolved narrow gaps between colloidal particles (arrows in the averaged map in Fig. 7 and Table 1). As the phase differences in a resolution shell between two maps are random, the cosine (or sine) term approaches the random phase limit (0.64) [equation (8), Figs. 3(*f*), 4(*f*), 5(*e*) and 6(*e*)], we adopted the reciprocal of the scattering vector length, where the cosine curve for sets of retrieved maps approaches the random-phase limit, as an effective resolution of the averaged map.

As a result, the effective resolution estimated in each of Figs. 7(*b*) and 7(*c*) was approximately twice the gap dimensions. From the point of view of the resolution in microscopy that is given by the Fourier component with the shortest period, half of the minimum period of the Fourier component is resolved. In Figs. 7(*a*) and 7(*d*), as the cosine terms were still greater than 0.64 at the highest resolution recorded, the effective resolution of the averaged map was assumed to be the highest resolution of the diffraction pattern used in the PR calculation (Table 1).

We also estimated the effective resolution using the PRTF curves [equation (10)]. The curves were calculated independently for ten and 50 maps from the best classes in each structure analysis and additionally for two maps yielding the smallest similarity score as a reference (Fig. 7). For maps displaying small variations in electron density, the greater number of maps used in the PRTF calculation likely correlated with the more rapid decrease of the curves [Figs. 7(*a*)–7(*c*)]. In contrast, for the maps that were very similar to each other, the PRTF curves were almost independent of the number of maps [Fig. 7(*d*)]. A comparison of the four cases indicated that the PRTF curves were likely advantageous for monitoring the degree of the convergence of the maps.

When the threshold for estimating the effective resolution of the maps from the PRTF curve was set to 0.5 or $1/e$ (~ 0.37) (Chapman *et al.*, 2006), the estimated effective resolution was lower than that to resolve the narrow gaps between particles. Instead, we tried to use the half-bit information threshold curve (van Heel & Schatz, 2005), which was proposed in conjunction with the Fourier shell correlation (FSC) curve (Rosenthal & Henderson, 2003) for estimating the effective resolution in single particle analysis using TEM. Although the definition of the FSC curve is different from that of the PRTF

curve, both curves essentially measure the distribution of phase values. The half-bit curve, which is correlated with the information content in voxels (pixels) in a resolution shell, is known to allow a reliable interpretation of the resolution level in TEM analysis (van Heel & Schatz, 2005). The effective resolution estimated from the scattering vector length, where the PRTF curve crossed the half-bit curve, was comparable

with the doubled gap dimension between colloidal particles (Fig. 7 and Table 1).

As a result, the effective resolution estimated using the cosine term and the random-phase limit of 0.64 was comparable with that using the PRTF and half-bit curves. The estimation of the effective resolution of the retrieved maps is still under debate – further experiences regarding the struc-

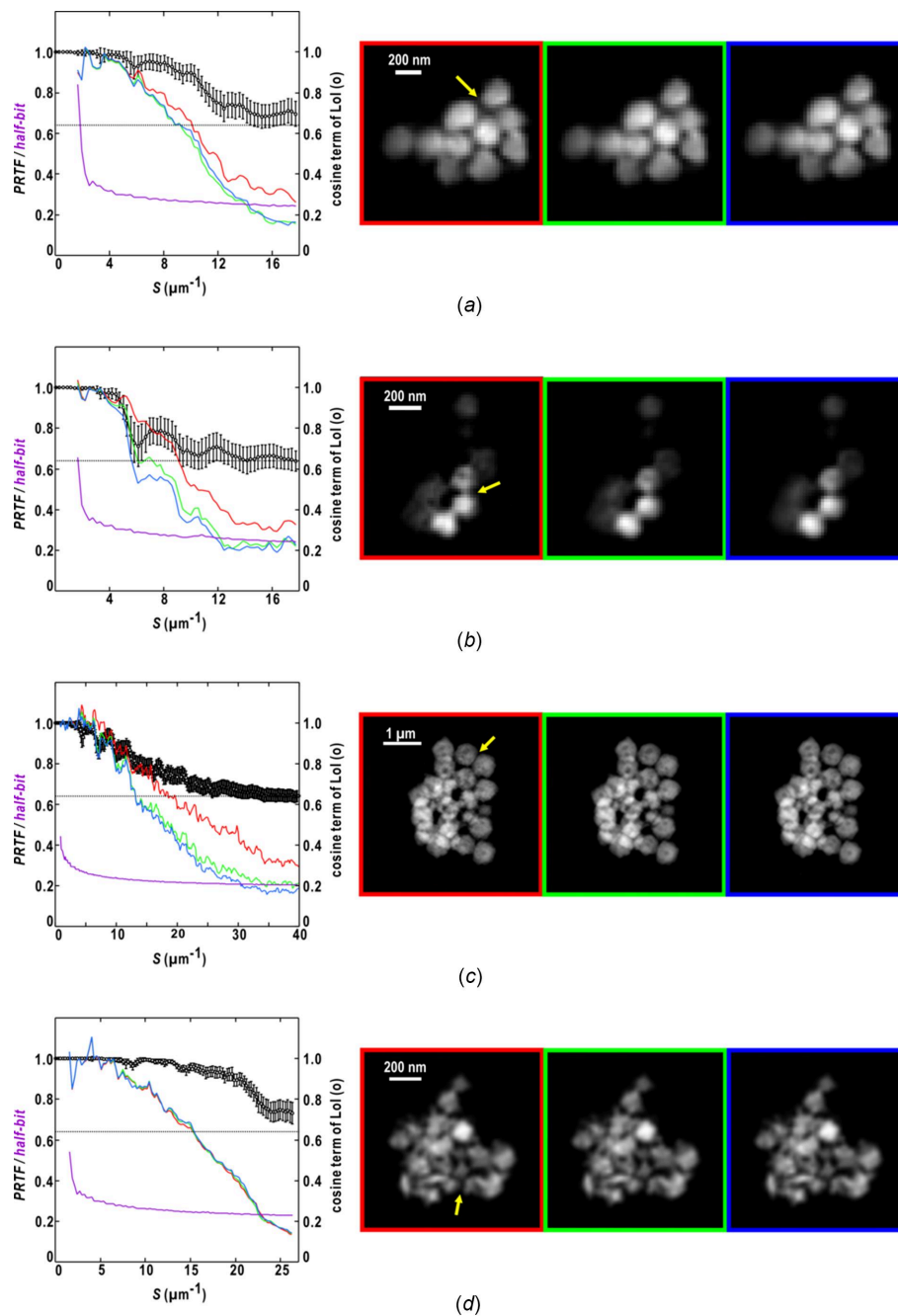


Figure 7

Resolution dependencies of the cosine term of equation (7) (open circles with standard deviations), the PRTF curves (red, green and blue) and the half-bit curves (purple). The cosine terms in panels (a)–(d) are taken from Figs. 3–6, respectively. The PRTF curves colored in red, green and blue were calculated for the 2, 10 and 49 smallest similarity scores including the reference, respectively. In the plots, the dotted line indicates the random phase limit [equation (8)]. The maps depicted in the right panel are averaged for the 2 (indicated by red box), 10 (green) and 50 maps (blue) used in the PRTF calculations. The arrows in the maps indicate the small gaps between particles in real space referenced to estimate the effective resolution using the cosine terms and the PRTF curves in reciprocal space.

Table 1

Effective resolution estimated using the cosine term of equation (7) and the PRTF curves of Fig. 7 with appropriate threshold values.

Gap between particles (nm)	Effective resolution estimated (μm^{-1} / nm) (number of maps used)			
	From cosine term using random phase limit	From PRTF curve and half-bit threshold curve		
Panel (a) / 27.8	18.0 / 55.6 (190)	18.0 / 55.6 (2)	16.0 / 62.5 (10)	15.6 / 64.3 (50)
Panel (b) / 27.8	14.2 / 70.4 (235)	18.0 / 55.6 (2)	18.0 / 55.6 (10)	18.0 / 55.6 (50)
Panel (c) / 12.5	32.6 / 30.6 (609)	40.0 / 25.0 (2)	40.0 / 25.0 (10)	34.5 / 29.0 (50)
Panel (d) / 18.5	26.0 / 38.5 (493)	24.6 / 40.6 (2)	24.6 / 40.6 (10)	24.6 / 40.6 (50)

ture analyses of various aggregates of metal particles are necessary.

5.4. Similarity score and landscape on the distribution of retrieved maps

In the previous study, we assumed that the phase-retrieved maps were distributed on a funnel-shaped landscape in multidimensional space spanned by the similarity score and pixel values of the maps, reflecting the structure, as schematically illustrated in Fig. 8(a) (Sekiguchi *et al.*, 2017). Based on this idea, the retrieved maps in Figs. 3–6 were plotted in the three-dimensional space spanned by the similarity score and the two PC vectors representing the major characteristics of the electron density distributions in all the retrieved maps [Figs. 8(b)–8(e)]. The maps were roughly distributed along funnel-shaped surfaces, and the distributions were different from each other probably due to the area of the missing small-angle regions, the signal-to-noise ratios and the oversampling ratios. The maps displaying small similarity scores were concentrated near the bottom part of the funnel. In contrast, the other maps with large scores were distributed around the upper edges of the funnel [Fig. 8(b)], or in local minima [Fig. 8(c)].

This representation is helpful for understanding characteristics of the funnel-shaped surface. As even maps displaying large scores are located in a local area on the two-dimensional planes, the landscape of the area is expected to be rugged as illustrated in Fig. 8(a). In the case of the retrieved maps of the aggregate of gold urchin particles, the nine classes discretely distributed near the bottom of the funnel [Fig. 8(e)] suggested local and shallow minima even at the bottom of the funnel, probably because of the small differences in the fine structures.

The visualized distributions encouraged us to steer the PR calculations toward the most probable solution for the observed diffraction pattern by referring to the similarity scores during the progress of the calculations. For instance, in parallelly conducted PR calculations, the modification of maps by those near the bottom of the funnel may increase the probability of successful calculations (Yoshida *et al.*, 2024).

Acknowledgements

We thank Dr Kensuke Tono and the SACLA engineering team of the Japan Synchrotron Radiation Research Center for their help in the XFEL-XDI experiments, and Professor

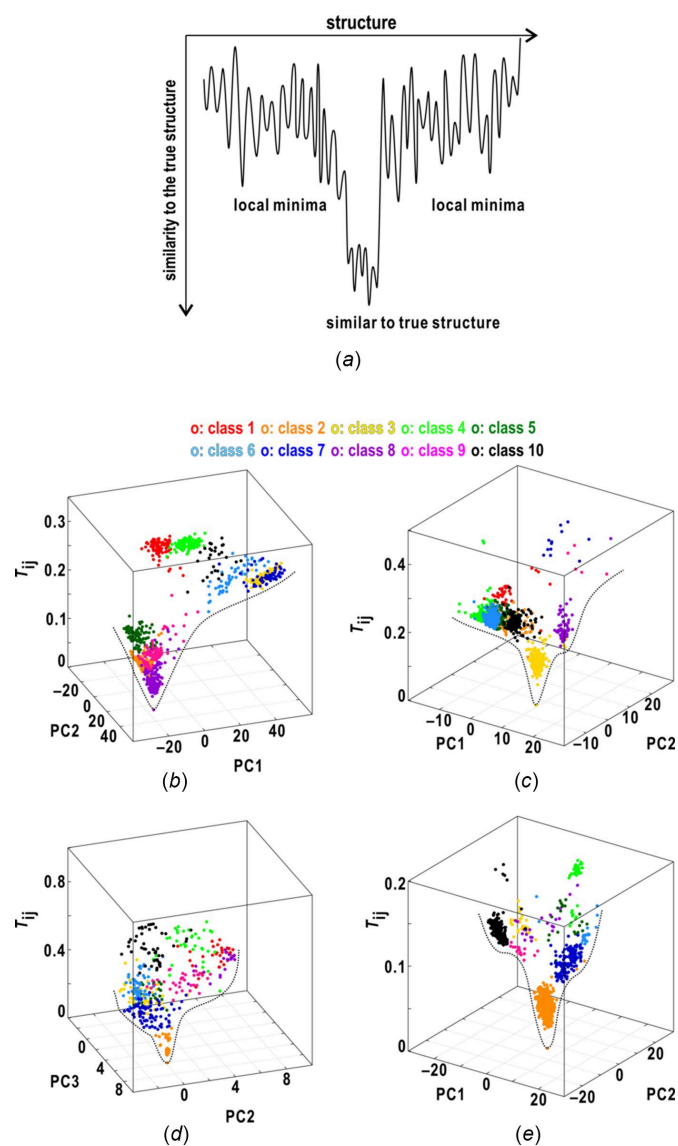


Figure 8

(a) Schematic illustration of the landscape in the multidimensional space spanned by the similarity score and structure axes. Panels (b)–(e) show the distributions of the retrieved maps taken from Figs. 3–6, respectively, in the three-dimensional space spanned by the similarity score and two PC vectors representing the variations of electron densities. The dots indicate the positions of the maps in the space, and are colored according to the scheme at the top of panels (b) and (c). In each panel, the dotted lines are the rough borders of the distributions of the maps.

Yasushi Kagoshima and the members of his laboratory for their help in the SR-XDI experiments. SR-XDI diffraction data were collected at BL24XU of SPring-8 (Proposal No.

2016B3201, 2018A3201). Data processing for XFEL-XDI was performed on an HPC computer at the SACLA facility.

Funding information

This study was supported by grants for XFEL key technology and the X-ray Free Electron Laser Priority Strategy Program to MN from The Ministry of Education, Culture, Sports, Science and Technology (MEXT). In addition, this study was supported by grants from The Japan Society for The Promotion of Science to MN (No. jp17654084, jp23120525, jp24654140, jp25120725, jp16H02218 and jp21H01050) and YT (No. jp16K21621 and jp19K14678), grants from MEXT to MN (No. jp15076210, and jp20050030) and by The CREST grant from the Japan Science and Technology Agency (No. JPMJCR2233 to YT). XFEL-XDI diffraction data were collected at SACLA (proposals 2013B8049 and 2014A8033)

References

Ahn, K., Cho, I. H., Kim, J., Lee, S. Y., Sung, D., Jung, C., Song, C., Kang, H. C. & Noh, D. Y. (2021). *J. Synchrotron Rad.* **28**, 505–511.

Ayyer, K., Xavier, P. L., Bielecki, J., Shen, Z., Daurer, B. J., Samanta, A. K., Awel, S., Bean, R., Barty, A., Bergemann, M., Ekeberg, T., Estillore, A. D., Fangohr, H., Giewekemeyer, K., Hunter, M. S., Karneviy, M., Kirian, R. A., Kirkwood, H., Kim, Y., Koliyadu, J., Lange, H., Letrun, R., Lübke, J., Michelat, T., Morgan, A. J., Roth, N., Sato, T., Sikorski, M., Schulz, F., Spence, J. C. H., Vagovic, P., Wollweber, T., Worbs, L., Yefanov, O., Zhuang, Y., Maia, F. R. N. C., Horke, D. A., Küpper, J., Loh, N. D., Mancuso, A. P. & Chapman, H. N. (2021). *Optica*, **8**, 15–23.

Barakat, R. & Newsam, G. (1984). *J. Math. Phys.* **25**, 3190–3193.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Born, M. & Wolf, E. (2001). *Principles of Optics*, 7th (expanded) ed. Cambridge University Press.

Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006). *J. Opt. Soc. Am. A*, **23**, 1179–1200.

Chen, C.-C., Miao, J., Wang, C. W. & Lee, T. K. (2007). *Phys. Rev. B*, **76**, 064113.

Cho, D. H., Shen, Z., Ihm, Y., Wi, D. H., Jung, C., Nam, D., Kim, S., Park, S.-Y., Kim, K. S., Sung, D., Lee, H., Shin, J.-Y., Hwang, J., Lee, S. Y., Lee, S. Y., Han, S. W., Noh, D. Y., Loh, N. D. & Song, C. (2021). *ACS Nano*, **15**, 4066–4076.

Dinapoli, R., Bergamaschi, A., Henrich, B., Horisberger, R., Johnson, I., Mozzanica, A., Schmid, E., Schmitt, B., Schreiber, A., Shi, X. & Theidel, G. (2011). *Nucl. Instrum. Methods Phys. Res. A*, **650**, 79–83.

Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. Berlin, Heidelberg: Springer-Verlag.

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

Favre-Nicolin, V., Leake, S. & Chushkin, Y. (2020). *Sci. Rep.* **10**, 2664.

Fienup, J. R. (1978). *Opt. Lett.* **3**, 27–29.

Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.

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

Gerchberg, R. W. & Saxton, W. O. (1972). *Optik*, **35**, 237–246.

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.

Heel, M. van & Schatz, M. (2005). *J. Struct. Biol.* **151**, 250–262.

Huang, X., Nelson, J., Steinbrener, J., Kirz, J., Turner, J. J. & Jacobsen, C. (2010). *Opt. Express*, **18**, 26441–26449.

Ibel, K. & Stuhrmann, H. B. (1975). *J. Mol. Biol.* **93**, 255–265.

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.* **107**, 11234–11239.

Jolliffe, I. T. & Cadima, J. (2016). *Philos. Trans. R. Soc. A* **374**, 20150202.

Kameda, H., Usugi, S., Kobayashi, M., Fukui, N., Lee, S., Hongo, K., Mizobata, T., Sekiguchi, Y., Masaki, Y., Kobayashi, A., Oroguchi, T., Nakasako, M., Takayama, Y., Yamamoto, M. & Kawata, Y. (2017). *J. Biochem.* **161**, 55–65.

Kameshima, T., Ono, S., Kudo, T., Ozaki, K., Kirihara, Y., Kobayashi, K., Inubushi, Y., Yabashi, M., Horigome, T., Holland, A., Holland, K., Burt, D., Muraio, 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., Oroguchi, T., Yamamoto, M. & Nakasako, M. (2018a). *Sci. Rep.* **8**, 831.

Kobayashi, A., Sekiguchi, Y., Takayama, Y., Oroguchi, T. & Nakasako, M. (2014). *Opt. Express*, **22**, 27892–27909.

Kobayashi, A., Sekiguchi, Y., Takayama, Y., Oroguchi, T., Shirahama, K., Torizuka, Y., Manoda, M., Nakasako, M. & Yamamoto, M. (2016a). *Rev. Sci. Instrum.* **87**, 053109.

Kobayashi, A., Takayama, Y., Hirakawa, T., Okajima, K., Oide, M., Oroguchi, T., Inui, Y., Yamamoto, M., Matsunaga, S. & Nakasako, M. (2021). *Sci. Rep.* **11**, 3877.

Kobayashi, A., Takayama, Y., Okajima, K., Oide, M., Yamamoto, T., Sekiguchi, Y., Oroguchi, T., Nakasako, M., Kohmura, Y., Yamamoto, M., Hoshi, T. & Torizuka, Y. (2018b). *J. Synchrotron Rad.* **25**, 1803–1818.

Kodama, W. & Nakasako, M. (2011). *Phys. Rev. E*, **84**, 021902.

Kraft, P., Bergamaschi, A., Broennimann, Ch., Dinapoli, R., Eikenberry, E. F., Henrich, B., Johnson, I., Mozzanica, A., Schlepütz, C. M., Willmott, P. R. & Schmitt, B. (2009). *J. Synchrotron Rad.* **16**, 368–375.

Kuglin, C. D. & Hines, D. C. (1975). *Proceedings of the IEEE International Conference on Cybernetics and Society*, New York, NY, USA, pp. 163–165.

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., Gumprecht, L., Coppola, N., Wunderer, C., Graafsma, H., Maia, F. R. N. C., 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., Soltan, 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*, **486**, 513–517.
- Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* **D49**, 530–533.
- MacQueen, J. (1967). *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California, USA, Vol. 1, pp. 281–297. University of California Press.
- 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., Loh, N. D., Hampton, C. Y., Sierra, R. G., Wang, F., Aquila, A., Bajt, S., Barthelmess, M., Bostedt, C., Bozek, J. D., Coppola, N., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Frank, M., Graafsma, H., Gumprecht, L., Hartmann, A., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Kassemeyer, S., Kimmel, N., Liang, M., Lomb, L., Maia, F. R. N. C., Marchesini, S., Nass, K., Pedersoli, E., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Schulz, J., Shoeman, R. L., Soltau, H., Starodub, D., Steinbrener, J., Stellato, F., Strüder, L., Ullrich, J., Weidenspointner, G., White, T. A., Wunderer, C. B., Barty, A., Schlichting, I., Bogan, M. J. & Chapman, H. N. (2012). *Opt. Express*, **20**, 13501–13512.
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature*, **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. USA*, **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. (2018). *X-ray Diffraction Imaging of Biological Cells*, Vol. 210 of *Springer Series in Optical Sciences*. Tokyo: Springer Nature.
- Nakasako, M., Kobayashi, A., Takayama, Y., Asakura, K., Oide, M., Okajima, K., Oroguchi, T. & Yamamoto, M. (2020). *Biophys Rev*, **12**, 541–567.
- 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.
- Oide, M., Hikima, T., Oroguchi, T., Kato, T., Yamaguchi, Y., Yoshihara, S., Yamamoto, M., Nakasako, M. & Okajima, K. (2020). *FEBS J.* **287**, 1612–1625.
- Oide, M. & Nakasako, M. (2021). *Sci. Rep.* **11**, 2827.
- Oide, M., Sekiguchi, Y., Fukuda, A., Okajima, K., Oroguchi, T. & Nakasako, M. (2018). *J. Synchrotron Rad.* **25**, 1379–1388.
- Oroguchi, T. & Nakasako, M. (2013). *Phys. Rev.* **E87**, 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 At. Mol. Opt. Phys.* **48**, 184003.
- Oroguchi, T., Yoshidome, T., Yamamoto, T. & Nakasako, M. (2018). *Nano Lett.* **18**, 5192–5197.
- 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.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.
- Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* **D53**, 448–455.
- Pfeiffer, F. (2018). *Nat. Photon.* **12**, 9–17.
- 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., Odić, D., Hasse, D., Carlsson, G. H., Larsson, D. S. D., Barty, A., Martin, A. V., Schorb, S., Bostedt, C., Bozek, J. D., Rolles, D., Rudenko, A., Epp, S., Foucar, L., Rudek, B., Hartmann, R., Kimmel, N., Holl, P., Englert, L., Duane Loh, N., Chapman, H. N., Andersson, I., Hajdu, J. & Ekeberg, T. (2015). *Nat. Commun.* **6**, 5704.
- Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., Hantke, M., DePonte, D. P., Barty, A., Schulz, J., Gumprecht, L., Coppola, N., Aquila, A., Liang, M., White, T. A., Martin, A., Caleman, C., Stern, S., Abergel, C., Seltzer, V., Claverie, J.-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*, **470**, 78–81.
- Sekiguchi, Y., Hashimoto, S., Kobayashi, A., Oroguchi, T. & Nakasako, M. (2017). *J. Synchrotron Rad.* **24**, 1024–1038.
- 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. USA*, **102**, 15343–15346.
- Stuhrmann, H. B. & Miller, A. (1978). *J. Appl. Cryst.* **11**, 325–345.
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). *Biophys. J.* **80**, 2946–2953.
- Takahashi, Y., Suzuki, A., Zetsu, N., Oroguchi, T., Takayama, Y., Sekiguchi, Y., Kobayashi, A., Yamamoto, M. & Nakasako, M. (2013). *Nano Lett.* **13**, 6028–6032.
- Takayama, Y., Tsusaka, Y., Kagoshima, Y., Kuwamoto, S., Urushihara, Y., Li, L., Nose, S., Sudo, T., Yoshimura, M., Yokoyama, K. & Matsui, J. (2020). *SPRING-8/SACLA Annual Report FY2018*, pp. 132–135. SPRING-8, Sayo-gun, Hyogo, Japan.
- Takayama, Y., Fukuda, K., Kawashima, M., Aoi, Y., Shigematsu, D., Akada, T., Ikeda, T. & Kagoshima, Y. (2021). *Commun. Phys.* **4**, 48.

- Takayama, Y., Inui, Y., Sekiguchi, Y., Kobayashi, A., Oroguchi, T., Yamamoto, M., Matsunaga, S. & Nakasako, M. (2015a). *Plant Cell Physiol.* **56**, 1272–1286.
- Takayama, Y., Maki-Yonekura, S., Oroguchi, T., Nakasako, M. & Yonekura, K. (2015b). *Sci. Rep.* **5**, 8074.
- Takayama, Y. & Nakasako, M. (2012). *Rev. Sci. Instrum.* **83**, 054301.
- Takayama, Y., Takami, Y., Fukuda, K., Miyagawa, T. & Kagoshima, Y. (2018). *J. Synchrotron Rad.* **25**, 1229–1237.
- Takayama, Y. & Yonekura, K. (2016). *Acta Cryst.* **A72**, 179–189.
- 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.
- Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000). *Nature*, **405**, 647–655.
- Uezu, S., Yamamoto, T., Oide, M., Takayama, Y., Okajima, K., Kobayashi, A., Yamamoto, M. & Nakasako, M. (2023). *Sci. Rep.* **13**, 10802.
- Wang, X., Yang, D.-P., Huang, P., Li, M., Li, C., Chen, D. & Cui, D. (2012). *Nanoscale*, **4**, 7766–7772.
- 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.
- Yoshida, S., Harada, K., Uezu, S., Takayama, Y. & Nakasako, M. (2024). *J. Synchrotron Rad.* **31**. Accepted. [YN5104].