Biophysics and Physicobiology

Regular Article

Parameter optimization for 3D-reconstruction from XFEL diffraction patterns based on Fourier slice matching

Miki Nakano[1], Osamu Miyashita[1] and Florence Tama[1,2,3]

[1]*RIKEN Center for Computational Science, Kobe, Hyogo 650-0047, Japan*
[2]*Department of Physics, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan*
[3]*Institute of Transformative Bio-Molecules, Nagoya University, Nagoya, Aich 464-8602, Japan*

Single-particle analysis (SPA) by X-ray free electron laser (XFEL) is a novel method that can observe biomolecules and living tissue that are difficult to crystallize in a state close to nature. To reconstruct three-dimensional (3D) molecular structure from two-dimensional (2D) XFEL diffraction patterns, we have to estimate the incident beam angle to the molecule for each pattern to assemble the 3D-diffraction intensity distribution using interpolation, and retrieve the phase information. In this study, we investigated the optimal parameter sets to assemble the 3D-diffraction intensity distribution from simulated 2D-diffraction patterns of ribosome. In particular, we examined how the parameters need to be adjusted for diffraction patterns with different binning sizes and beam intensities to obtain the highest resolution of molecular structure phase retrieved from the 3D-diffraction intensity. We found that resolution of restored molecular structure is sensitive to the interpolation parameters. Using the optimal parameter set, a linear oversampling ratio of around four is found to be sufficient for correct angle estimation and phase retrieval from the diffraction patterns of SPA by XFEL.

Corresponding author: Florence Tama, RIKEN Center for Computational Science, 6-7-1, Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
e-mail: florence.tama@riken.jp

Key words: single-particle analysis, coherent X-ray diffraction imaging, phase retrieval

Single-particle analysis (SPA) using femtosecond X-ray pulses from free electron laser (XFEL) is a new technique to observe the structure of biological samples in a state close to nature [1–8]. The bright and highly coherent beam of XFEL allows us to obtain diffraction data without crystallization. The short femtosecond pulses of the beam enable us to record the data without radiation damage, referred to as "diffraction-before-destruction" [1,4,5,9,10]. In addition, using XFEL beam, we can observe the inner structure of samples thicker than 500 nm without multiple scattering events, which is an unavoidable problem in cryo-electron microscopy (cryo-EM).

Currently, five XFEL facilities are under operation in the world [11], and the volume of XFEL experimental data is increasing [7,12–16]. Gallagher-Jones, M., *et al.* observed the nanostructure formation of RNA interference microsponges using a combination of XFEL and synchrotron X-rays [12]. Kimura, T., *et al.* demonstrated two-dimensional imaging of live cells using XFEL diffraction data at 28 nm full-period resolution [13]. Ekeberg, T., *et al.* presented the three-dimensional molecular structure of the giant mimivirus particle, reconstructed at 125 nm resolution from

◀ *Significance* ▶

In this study, we examined the effect of data binning on the resolutions of restored 3D-molecular structures from the 2D-diffraction images obtained by X-ray Free Electron Laser (XFEL) single particle analysis. We found that the advantage of pixel binning on reducing noise exceeds the disadvantage of losing the detailed pixel information, and that the interpolation parameters should be optimized in relation to binning size. By using an optimal parameter set, a lower linear oversampling ratio of around four is sufficient to restore the 3D-molecular structure by single particle analysis using XFEL experiments.
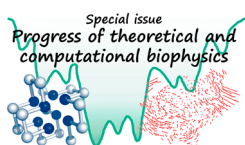
diffraction patterns obtained by XFEL experiments [14]. Recently, Rose, M., *et al.* reported three-dimensional reconstructions of PR772 virus structure without symmetry constraints with better than 10 nm resolution [7]. In the future, higher resolution should be expected as experiments are still being developed and theoretical work has demonstrated that given a large number of diffraction patterns and strong beam intensity, 0.1 nm resolution could be achieved even for a small protein such as the lysozyme [17].

In single particle XFEL experiments, the incident beam orientation of each diffraction pattern is unknown, therefore angles between 2D-diffraction patterns have to be estimated to assemble them into 3D-diffraction intensity distribution. Several algorithms have been developed for such purpose. One is based on "maximum correlation coefficients" [18–20], and the other is the "expand, maximize and compress (EMC)" algorithm [21]. In these approaches, 2D-diffraction patterns are assembled into a 3D-diffraction intensity distribution through iterative procedures. In the "maximum correlation coefficient" approach, a single orientation is assigned to each diffraction pattern based on its consistency against the tentative 3D-diffraction intensity at each iteration. This algorithm is simple and it has better scalability when it is applied to a large number of diffraction patterns [18,22]. In the EMC algorithm, a number of angular assignments are considered for each target diffraction pattern with relative weights based on the similarities between the target and reference diffraction patterns from the tentative 3D-diffraction intensity. This strategy significantly improves the convergence, but the resolution of the resulting structure may be overestimated [23]. In our previous study, we performed the reconstruction of the structure of a large biological molecule, ribosome, from the simulated diffraction data using maximum correlation coefficient approach [24]. We examined how the quantity and quality of the diffraction patterns affect the resulting 3D-molecular structure and discussed the minimum experimental conditions to reconstruct the molecular structure at certain resolutions.

Usually, experimental diffraction patterns are binned before analysis to increase signal and noise ratio, SNR, and to reduce image size to make data handling easier. Even using the bright pulse of XFEL, diffraction intensities from single particle are quite low especially on the high wavenumber area, which are located on the outer region of the diffraction pattern. Binning improves the quality of experimental diffraction pattern by reducing the number of zero and low-count pixels by averaging multiple pixels. However, by binning, there is possible loss of detailed information within the binned pixels. Therefore, we need to investigate the optimal binning size to keep the quality of restored structures.

Practically, the binning protocol also affects the optimal parameters to estimate the incident beam angle and assemble 2D-diffraction patterns into a 3D-diffraction intensity. In addition, the binning step size can affect the quality of the

phase recovery process [25]. Miao, J., *et al.* showed that missing phase information may be recovered if the total number of pixels in the diffraction patterns is sufficiently larger, *i.e. oversampled*, than the number of *Shannon* pixels, which is defined as the size of 1/(particle size) [26]. It has been also shown that a larger oversampling ratio improves the quality of phase retrieval [27]. Thus, excessive binning may affect the accuracy of phase recovery.

In this study, we examined how different oversampling ratio affects the resolution of a biomolecular structure (ribosome) reconstructed from simulated diffraction patterns. Achievable beam intensities by current and near future XFEL facilities were used for simulations. We investigated the optimal parameter set with different oversampling ratio, and how the ratio affects the resolutions of restored structures.
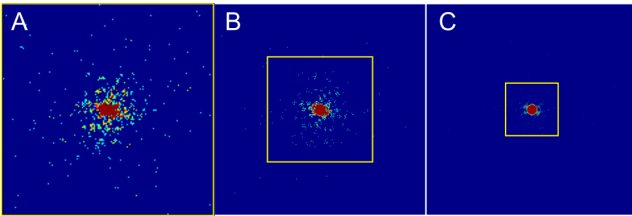
Materials and Methods

Simulated diffraction pattern dataset

We performed 3D-molecular structure reconstruction of the ribosome from simulated 2D-coherent diffraction patterns to demonstrate the application of the single particle analysis using XFEL data for large biological molecules. We chose the crystal structure of Thermus thermophilus 70S ribosome bound with release factor RF2 from Protein Data Bank (PDB ID: 4V67, [28] molecular size ~32 nm) as the target structure.
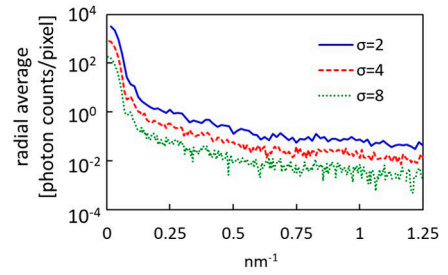
We converted the atomic structure of the X-ray crystallography structure in the PDB format to an electron density map, with 0.4^3 nm³/voxel resolution in real space by using Xmipp [29]. The electron density map was converted to 3D-diffraction intensity distributions by Fourier transform with padding factor 2, 4, or 8 to investigate the dependence of the resolution of restored structures on the linear oversampling ratio for each dimension, σ (described later in detail). By taking slices through cubic spline interpolation from these simulated 3D-diffraction intensities, we created 10242 2D-diffraction intensity distribution patterns. Ewald sphere was approximated as a plane in the slicing process (and in the assembly, discussed later). The effect of the curvature is not large, particularly for the regions used for pattern matching, and the effect of binning is negligible. Therefore, this approximation should not affect the results in this study. Poisson noise was applied onto the simulated diffraction patterns, corresponding to strong and medium beam intensities, 5.5×10^{13} and 5.5×10^{12} (photons/μm²), respectively. The medium beam intensity we assumed here could be achieved by current XFEL facilities. A stronger beam intensity was also considered since ongoing XFEL experiments developments would make such intensity achievable in the near future [30,31]. In our previous study, we also tested weak beam intensity, 5.5×10^{11} (photons/μm²), but the resolutions of restored molecular structures were quite low [24]. Therefore, we tested strong and medium

beam intensity in this study. Beam intensities were estimated by comparing against outputs from spsim [32] with wavelength 0.1 nm and quantum efficiency of the detector at 0.8. These slices were used as in place of experimental XFEL diffraction patterns. It should be noted that, in actual measurements, photon counts are integrated over the detector pixel and deconvolution procedure can be used to obtain "exact" measurement at the center of each detector pixel [33]; however, such integration is not considered in this simulation.

Three sets of diffraction patterns using different binning sizes were examined. In all the sets, the wavenumber at the detector edge was set to 1.25 nm^{-1}. The three diffraction pattern sizes considered were 640×640 pixels, 320×320 pixels, and 160×160 pixels, each corresponding to the pixel size, Δk, of 0.00391 nm^{-1}, 0.00781 nm^{-1}, and 0.0156 nm^{-1}, respectively. Since the size of the ribosome is $L\sim32$ nm, linear oversampling ratio, $\sigma=(1/L)/\Delta k$, of each data set corresponds to $\sigma=8$, and $\sigma=4$, and $\sigma=2$. The diffraction pattern sizes and resolutions with $\sigma=2$ or 4 correspond to those of the binned patterns with step 4 or 2 from the pattern with $\sigma=8$. To reduce computational time for the slice matching iteration, the patterns with $\sigma=4$ or 8 were cropped to 160×160 pixels. We note here that the diffraction patterns are cropped only to increase computation efficiency. Pattern comparisons during slice matching are performed using only the annular region defined within 160 pixels corresponding to the same wavenumber among the diffraction patterns with different σ values (described later). Then, the final 3D-diffraction intensity distribution was assembled from the original full-size diffraction patterns. Diffraction pattern samples for each σ, the photon count distributions on 2D-detectors, created with strong beam intensities for the different σ are shown in Figure 1. Radial averages of the photon count on a sample diffraction pattern for each σ are shown in Figure 2. The average photon counts per pixel differ significantly, which affects the S/N ratio of each simulated diffraction patterns.
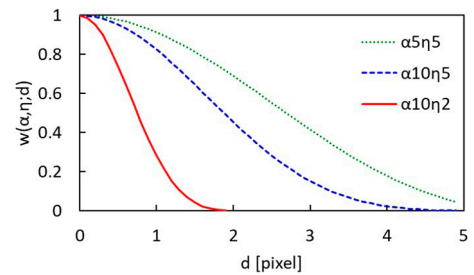

Figure 2 Radial average of diffraction intensity on a sample diffraction pattern for each σ with strong beam intensity.

Assembling 2D-diffraction patterns into 3D-diffraction intensity

To assemble 2D-diffraction patterns into a 3D-diffraction intensity, we used a weight function based on the Kaiser-Bessel window, $w(\alpha, \eta; d)$ defined by Eq. 1. The value of $w(\alpha, \eta; d)$ depends on the distance d which is between the center of the voxel where the diffraction intensity is being calculated and the exact mapped position from the 2D-diffraction pattern to the assembled 3D-diffraction intensity. η determines the cut-off interpolation length, and α determines the decreasing rate of the weight function. With large α, weight for interpolation would decrease quickly for larger d as shown in Figure 3. To investigate the most optimal interpolation parameter set for each slice set from different conditions, we assembled a 3D-diffraction intensity, V_{true}, using the correct (*true*) angles, *i.e.*, the angles used for the creation of the slice set. The interpolation parameter list used for the assembly of V_{true} for both strong and medium beam intensities are shown in Table 1.

$$w(d) = \frac{\xi(\alpha,\eta)}{I_0(\pi\alpha)} I_0\left\{\pi\alpha\sqrt{1-\left(\frac{2d-\eta}{\eta-1}-1\right)^2}\right\}, \quad 0\leq d\leq\eta$$

(1)

To determine the orientation angles of the samples against the incident beam captured in each diffraction pattern, we performed slice matching [34]. We used annular regions on the diffraction patterns defined by the inner and outer radii, R_i and R_o, for slice matching, and set these parameters corresponding to the same wavenumber among the diffraction

Figure 1 Diffraction pattern samples with a strong beam intensity for (A) $\sigma=2$, (B) $\sigma=4$, and (C) $\sigma=8$. The full sizes of the diffraction patterns are 160×160 pixels for $\sigma=2$, 320×320 pixels for $\sigma=4$, and 640×640 pixels for $\sigma=8$, respectively. The wavenumber at the edge corresponds to 1.25 nm^{-1} for all slices. Yellow squares in each figure show the cropped areas with 160×160 pixels, which were used for slice matching.

Figure 3 Profiles of interpolation weight function, $w(\alpha, \eta; d)$.

Table 1 Interpolation parameter sets used to reconstruction of V_{true}

σ	(α, η)						
2	<u>α15η2</u>	<u>α15η3</u>	α18η2	α20η1.5	<u>α20η2</u>	<u>α20η3</u>	α25η2
4	α05η3	<u>α10η2</u>	α10η5	<u>α10η3</u>	α15η2	<u>α15η3</u>	α15η5
8	<u>α05η2</u>	<u>α05η5</u>	α05η6	<u>α10η2</u>	<u>α10η5</u>	α15η5	

* All combinations of parameters were examined for both strong and medium beam intensity. FSC curves underlined combinations are shown in Figure 4.

patterns with different σ. All parameters, including matching region and interpolation parameters, are adjusted during the iterations. We started the slice matching using small α and η, which makes the resolution of the assembled 3D-diffraction intensity distribution lower. In the subsequent iterations, these parameters were updated to make the 3D-diffraction intensity have higher resolutions. These parameters were selected empirically so that the angles would be evenly assigned among the experimental diffraction patterns. The detailed parameter sets used for this study are shown in Supplementary Table S1. Using the angles estimated by slice matching and the most optimal interpolation parameters, we finally assembled the 3D-diffraction intensity, V_{match}, with full size of diffraction patterns having the same highest wavenumber.

Quantification of the accuracy of angle estimation and the restored 3D-molecular structures

To assess the accuracy of the angle estimation, we introduced the angle distance error average, Δangle, defined as follows:

$$\Delta\text{angle} = \langle |\beta_{ij} - \beta'_{ij}| \rangle \tag{2}.$$

Here, β_{ij} and β'_{ij} are the relative angles between i-th and j-th planes within V_{true} and V_{match}, respectively. We calculated the absolute differences between these angles as angle errors for all combination of the slices and their average as Δangle.

Using the final 3D-diffraction intensities, V_{true} and V_{match}, we performed phase retrieval by using hybrid input-output (HIO) approach [35]. The support region was fixed to be a sphere having a diameter corresponding to the molecular size during the iteration. The condition of phase retrieval is shown in Supplementary Table S2.

To quantify how well molecular structure was restored in real space, we calculated Fourier shell correlations (FSC) and phase retrieval transfer function (PRTF), which are commonly used to evaluate resolutions.

$$\text{FSC}(k) = \frac{\sum_{k_i \in k} \tilde{F}_{restore}(k_i) \cdot \tilde{F}_{answer}(k_i)^*}{\sqrt{\sum_{k_i \in k} |\tilde{F}_{restore}(k_i)|^2} \sqrt{\sum_{k_i \in k} |\tilde{F}_{answer}(k_i)|^2}} \tag{3}$$

$$\text{PRTF}(k_i) = \frac{\langle \tilde{F}_{restore}(k_i) \rangle}{|\tilde{F}_{answer}(k_i)|} \tag{4}$$

where $\tilde{F}_{restore}$ is the structure factor derived from experimental diffraction patterns with retrieved phases ϕ and $\tilde{F}_{answer}$ is the ground truth structure factor, which is the Fourier transform of the original electron density map. $\langle...\rangle$ denotes an average over the independent $\tilde{F}_{restore}$. FSC measures the normalized cross-correlation coefficient between two structures over corresponding shells in Fourier space. PRTF represents the confidence in the retrieved phases as a function of resolution.

Results

Optimization of interpolation parameters

To investigate the most optimal interpolation parameter set for each slice set, we assembled the 3D-diffraction intensity, V_{true}, using the angles used for the creation of slice set. Figure 4 shows the Fourier shell correlation (FSC) curves between restored molecular structures from V_{true} using the different interpolation parameters for each slice set and the electron density map converted from the PDB structure. For easy comparison, we show FSC curves for the restored structures using the most optimal interpolation parameters for each slice set in Figure 5.

As shown in Figure 5, for higher σ, the optimal α is lower and the optimal η is longer. The optimal η were the same between different beam intensity, while optimal α was generally lower for lower beam intensity. Although the differences of FSC curves among the different σ were small, FSC for σ=4 were slightly higher than those for others, and those for σ=8 were the lowest. These differences are more noticeable when the beam intensity is lower. From these results, we determined the optimal interpolation parameters as summarized in Table 2.

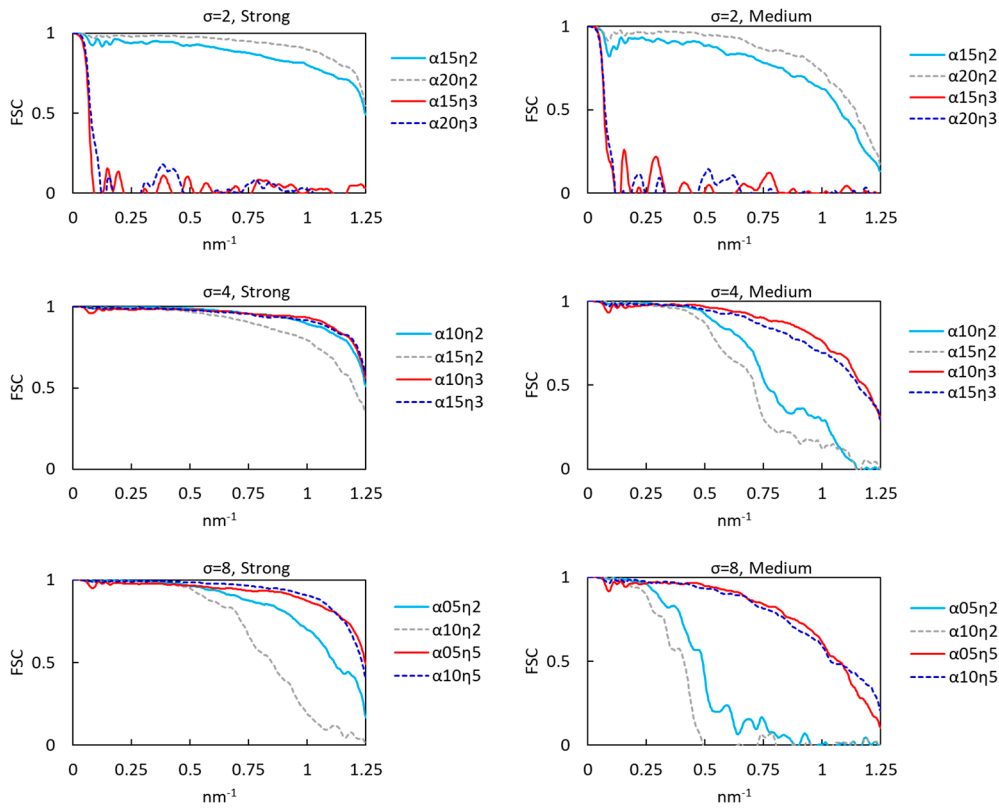### Results of angle estimations using slice matching

As described in the method section, we assembled the 3D-diffraction intensity, V_{match}, using the angles estimated by slice matching using the most optimal interpolation parameters. We performed slice matching started with five different initial random angle sets, called as trial 1 to trial 5.

Figure 6 shows the final assigned angle distribution on the sphere surface estimated by slice matching and Table 3

Table 2 Optimal interpolation parameters of 3D-volume reconstruction in Fourier space and FSC resolution of restored model

Beam intensity	σ	(α, η)	wavenumber at FSC=0.5 [nm⁻¹]	Resolution [nm]
Strong	2	α20η2	>1.25	<0.8
	4	α15η3	>1.25	<0.8
	8	α10η5	1.23	0.81
Medium	2	α20η2	1.14	0.88
	4	α10η3	1.18	0.85
	8	α05η5	1.08	0.93

Figure 4 Fourier shell correlation (FSC) curves between restored molecular structures from V_{true} using the different interpolation parameters and the electron density map converted from the PDB structure. To clarify the comparison, only FSC curves for the underlined combinations in Table 1 are shown.
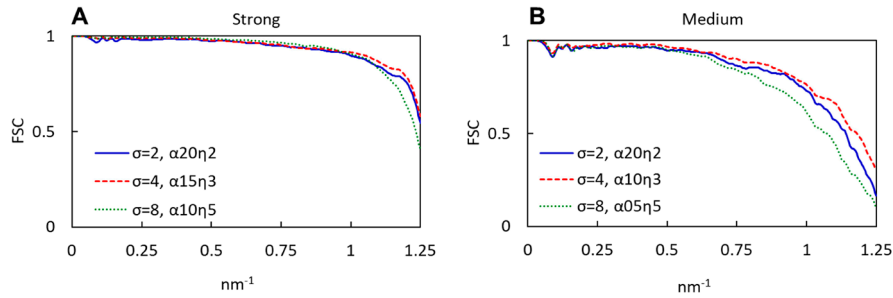

Figure 5 Fourier shell correlation (FSC) curves between restored molecular structures from V_{true} with (A) strong beam intensity, and (B) medium beam intensity, using the optimal interpolation parameters and the electron density map converted from the PDB structure.

shows the Δangle for each trial. Euler angles assigned for each diffraction patterns should be distributed evenly on the sphere to reconstruct good molecular structure in real space. Apparently, for trial 5 of σ=8 with strong beam intensity and for trial 4 of σ=8 with medium beam intensity, assigned angle distributions were quite uneven. We confirmed that Δangles for these trials were too large as shown in Table 3. The results of these two failed trials were eliminated for further analysis.

Figure 7 shows the 3D-electron density maps which were phase retrieved from V_{match}. All molecular structures were

well restored although the maps from the slices created with medium beam intensity were blurred compared to those with strong beam intensity. Figure 8 shows the FSC curves between maps restored from V_{match} and the electron density map converted from the PDB structure for each slice set. We also calculated phase retrieval transfer function, PRTF, which represents the confidence in the retrieved phases as a function of resolution, for each restored map as shown in Supplementary Figure S1. Outlier FSC curves were found for strong beam intensity, trial 4 of σ=2 and trial 3 for σ=4. In our protocol, we performed multiple phase retrieval
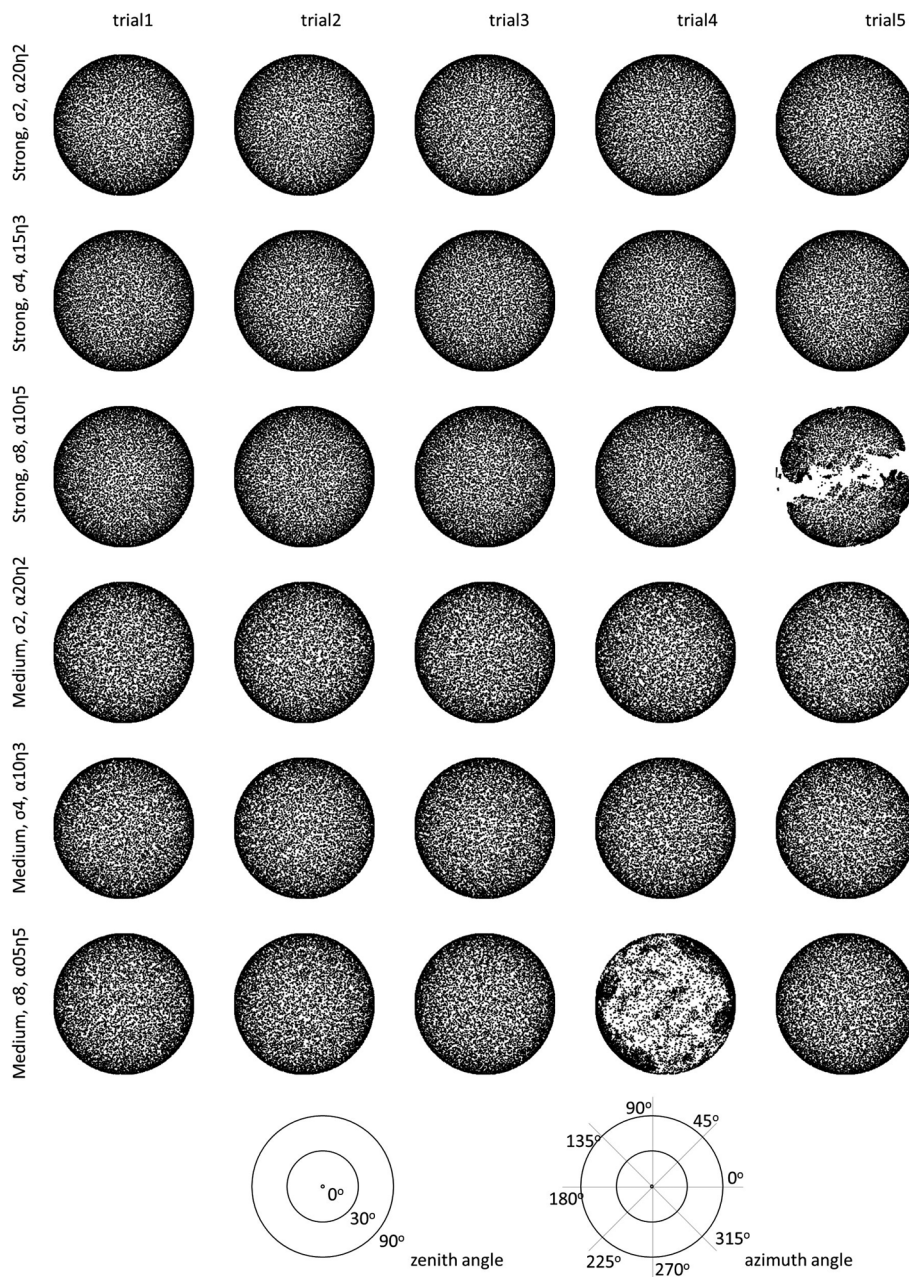
Figure 6 Distribution of the angles estimated by slice matching shown on the sphere surface. Corresponding azimuth and zenith angles are shown below.

Table 3 Angle errors, Δangle, for slice matching

|	trial 1	trial 2	trial 3	trial 4	trial 5
Strong, σ2	0.67±0.51	0.64±0.49	0.64±0.49	0.64±0.49	0.66±0.50
Strong, σ4	0.54±0.41	0.55±0.41	0.54±0.41	0.55±0.41	0.54±0.41
Strong, σ8	0.60±0.46	0.60±0.45	0.60±0.46	0.59±0.45	20.04±30.66
Medium, σ2	1.38±2.75	1.39±2.75	1.39±2.89	1.37±2.71	1.42±3.13
Medium, σ4	1.31±2.19	1.34±2.71	1.31±2.35	1.32±2.27	1.32±2.20
Medium, σ8	1.53±3.54	1.56±3.79	1.54±3.46	37.73±30.66	1.54±3.63
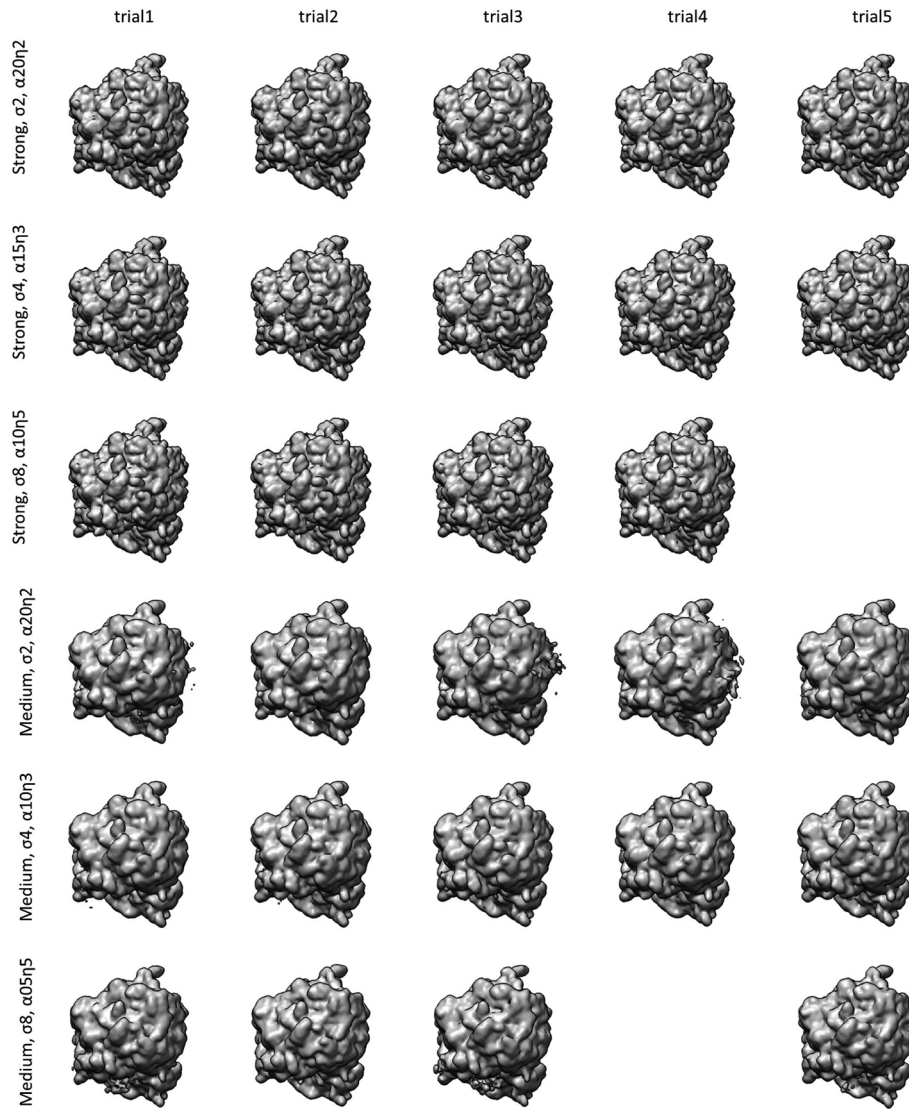
Figure 7 3D-molecular structures phase retrieved from V_{match}. Voxels having the density above 20% of the maximum density within the volume are shown. The restored structures for trial 5 of σ=8 with strong beam intensity and for trial 4 of σ=8 with medium beam intensity are not shown, because we did not perform phase retrieval for them due to their wrong matching results as shown in Figure 6.

starting with different random phases for the resulting the 3D-diffraction intensity from each trial. Therefore, these low resolutions of restored molecular structures are due to some errors in the assembled 3D-diffraction intensity, although Δangles for these trials were almost the same as other trials as shown in Table 3. FSC curves of obtained V_{match} became lower compared to the one created with correct angles, V_{true}, assembled from the same diffraction pattern set with the same interpolation parameters. This reduction of resolution is associated with the errors of angle estimation. Uncertainties of Δangles for medium beam intensity compared to strong beam intensity indicate the difficulty of both slice matching and phase retrieval due to weak signals. As a result, FSC curves for medium beam showed some fluctuation among the trials.

To measure the resolutions of the restored 3D-molecular structures, we used the wavenumbers at PRTF=1/e and FSC=0.5. These values are commonly used to evaluate the resolutions [36–38]. Figure 9 shows the correlation between PRTF resolution and Δangle (Fig. 9A), and between FSC resolution and Δangle (Fig. 9B), for all trials. Regardless of σ, Δangles for strong beam intensity were smaller than those for medium beam intensity. These smaller angle errors contribute to higher resolutions for restored 3D-molecular structures in terms of both PRTF and FSC resolutions. Pearson's correlation coefficient between Δangle and PRTF was −0.981 and between Δangle and FSC was −0.992.

Regarding the effect of σ parameter on the resulting structure resolution, Δangles for σ=4 were the smallest among all σ for both strong and medium beam intensities and the
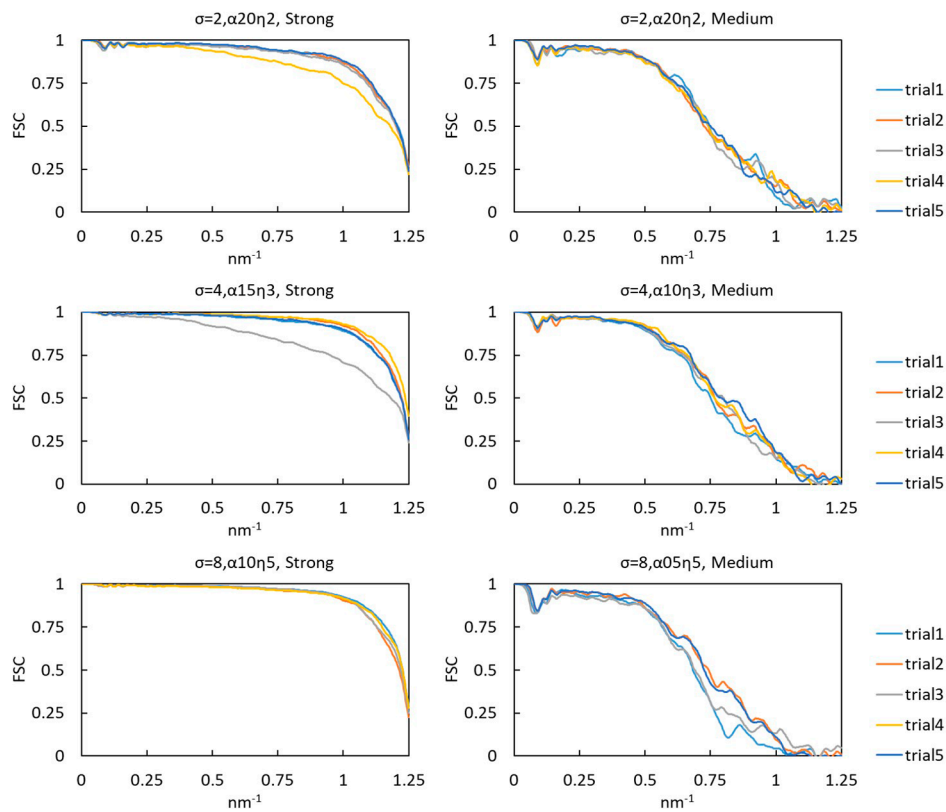
Figure 8 Fourier shell correlation (FSC) curves between restored molecular structures from V_{match} and the electron density map converted from the PDB structure.
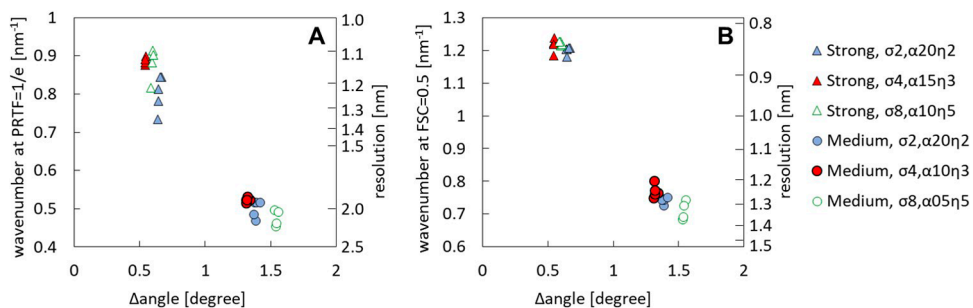
Figure 9 (A) PRTF resolution of restored electron density maps against Δangle and (B) FSC resolution against Δangle for all restored electron density maps from V_{match}.

restored molecular structures also showed the best FSC and PRTF resolutions. For strong beam intensity, σ=2 was the worst among the tested conditions. In particular, PRTF resolutions for σ=2 shows large deviation compared to those for higher σ. For medium beam intensity, both FSC and PRTF of restored molecular structures from the diffraction patterns of σ=8 were significantly worse than others because of larger Δangle. Finer binning in these diffraction patterns may be the cause of errors in the angle estimations.

Discussion

In this study, we investigated how the XFEL beam intensity and the binning of diffraction patterns affect the resolutions of restored structures, through extensive optimization of the parameters for the 3D-diffraction intensity assembly algorithm.

To assemble the 3D-diffraction intensity with finer detector pixel size (large oversampling ratio, σ), longer interpolation radius η is required using small decreasing rate α. This result is expected because experimental diffraction pattern with large σ has fine detector pixels, corresponds to smaller

wavenumber width, Δk, in Fourier space. Thus, each voxel within the 3D-diffraction intensity needs to be interpolated using more pixels from the corresponding 2D-diffraction patterns. We found that the optimal cutoff length, η, is not necessarily proportional to σ, presumably because interpolation over too many voxels can cause errors. By using the optimized interpolation parameters, restored molecular structures, V_{true}, with $\sigma=4$ showed the best FSC resolutions for both strong and medium beam intensity (Fig. 5). Another factor that would affect the resulting resolution is the accuracy of the angle estimation. Δangle for $\sigma=4$ showed the minimum values compared to other σ for both strong and medium beam intensity (Fig. 9). These small angle errors help to restore molecular structures with higher FSC resolution. For trials using slices with large $\sigma=8$, the slice matching protocol occasionally assigns heavily biased angle distributions (Fig. 6). The relationship between Δangle and σ appears to be non-trivial. It is expected that higher photon counts increases the accuracy of pattern matching. However, it is not the case in our results. A lower number of pixels are used for pattern matching for the tests with lower σ values, which may be one factor that increases the angle estimation error for $\sigma=2$.

Often, the oversampling ratio of an experimental diffraction pattern is lowered by binning from the original data. The main purpose of binning is to reduce noise, and to reduce the number of pixels which have no photon detection. Experimental diffraction pattern of a single molecule is sparse and has many non-photon-detected pixels even on the center region as shown in Figure 1. To restore such faint information, taking the average of a large number of pixels would be beneficial to increase SNR. On the other hand, this procedure makes Δk larger and is accompanied by the loss of the detailed information found within binned pixels. Thus, the balance between these two factors should affect the final resolution of the reconstructed 3D-molecular structure.

Our results indicate that effect of binning is subtle, but the optimal choice requires careful examination. The choice of the oversampling ratio, σ, affected both the accuracy of the angle estimation and phase recovery procedure and the qualities of restored structures measured by PRTF and FSC correlate with the accuracies of angle estimation (Fig. 9). The quality of phase retrieval usually improves with larger oversampling ratio [27]. The actual oversampling ratio for a 3D-diffraction intensity, σ^3, is larger than its linear oversampling ratio, σ, and thus a 3D-diffraction intensity has more information used for phase retrieval. In our test case, $\sigma=4$ was the best choice to obtain higher resolution structure. This ratio could raise SNR effectively with keeping the amount of information for phase retrieval.

Practically, handling large size diffraction patterns with large σ is computationally expensive. In our slice matching protocol, we use small cropped size diffraction patterns to reduce computation time. However, even using cropped size, the diffraction patterns with large σ require more calcu-

lation time because they require a larger matching region defined by R_i and R_o, and a longer interpolation radius, η. The computational time to calculate correlations between the diffraction patterns (10242 experimental diffraction patterns and 1692 reference diffraction patterns) using the annular region defined by $R_i=20$ and $R_o=40$ is 2 times more than using the region defined by $R_i=5$ and $R_o=10$. Also, the calculation time to assemble a 3D-diffraction intensity from 10242 2D diffraction patterns with $\eta=5$ requires about 7 times the calculation time to assemble it with $\eta=2$. Moreover, the calculation time of the phase retrieval was proportional to the cube of the pixel size. The intensive memory requirement is also a problem, because all voxels within the volume in both Fourier and real space have complex numbers.

Conclusion

We investigated the optimal parameter set for slice matching algorithm with different linear oversampling ratios σ, and how σ affects the resolutions of restored structures. We found that the resolution of a restored molecular structure is sensitive to the interpolation parameters. For the diffraction patterns with larger σ, the more distant pixels need to be interpolated with sufficient strength. By using the optimal parameter set, resolutions of restored structures are almost the same regardless of the σ values. Our results suggest that for SPA by XFEL, linear oversampling ratio of four is sufficient and the corresponding binning procedure is beneficial for reducing both the Poisson noise and the computational time for the analysis.

Acknowledgement

Conflicts of Interest

All authors declare that they have no conflict of interest.

Author Contribution

M. N. performed the calculations and wrote the manuscript. O. M. and F. T. directed the entire project and wrote the manuscript.

References

[1] Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
[2] Huldt, G., Szoke, A. & Hajdu, J. Diffraction imaging of single

particles and biomolecules. *J. Struct. Biol.* **144**, 219–227 (2003).

[3] Chapman, H. N., Barty, A., Bogan, M. J., Boutet, S., Frank, M., Hau-Riege, S. P., *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat. Phys.* **2**, 839–843 (2006).

[4] Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., *et al.* Femtosecond X-ray protein nano-crystallography. *Nature* **470**, 73–77 (2011).

[5] Gaffney, K. J. & Chapman, H. N. Imaging atomic structure and dynamics with ultrafast X-ray scattering. *Science* **316**, 1444–1448 (2007).

[6] Aquila, A., Barty, A., Bostedt, C., Boutet, S., Carini, G., dePonte, D., *et al.* The linac coherent light source single parti-cle imaging road map. *Struct. Dyn.* **2**, 41701 (2015).

[7] Rose, M., Bobkov, S., Ayyer, K., Kurta, R. P., Dzhigaev, D., Kim, Y. Y., *et al.* Single-particle imaging without symmetry constraints at an X-ray free-electron laser. *IUCrJ* **5**, 727–736 (2018).

[8] Miyashita, O. & Joti, Y. X-ray free electron laser single-particle analysis for biological systems. *Curr. Opin. Struct. Biol.* **43**, 163–169 (2017).

[9] Hirata, K., Shinzawa-Itoh, K., Yano, N., Takemura, S., Kato, K., Hatanaka, M., *et al.* Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. *Nat. Methods* **11**, 734–736 (2014).

[10] Suga, M., Akita, F., Hirata, K., Ueno, G., Murakami, H., Nakajima, Y., *et al.* Native structure of photosystem II at 1.95 A resolution viewed by femtosecond X-ray pulses. *Nature* **517**, 99–103 (2015).

[11] Spence, J. C. H. XFELs for structure and dynamics in biology. *IUCrJ* **4**, 322–339 (2017).

[12] Gallagher-Jones, M., Bessho, Y., Kim, S., Park, J., Kim, S., Nam, D., *et al.* Macromolecular structures probed by combin-ing single-shot free-electron laser diffraction with synchrotron coherent X-ray imaging. *Nat. Commun.* **5**, 3798 (2014).

[13] Kimura, T., Joti, Y., Shibuya, A., Song, C., Kim, S., Tono, K., *et al.* Imaging live cell in micro-liquid enclosure by X-ray laser diffraction. *Nat. Commun.* **5**, 3052 (2014).

[14] Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R. N. C., Seltzer, V., Claverie, J.-M., *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an x-ray free-electron laser. *Phys. Rev. Lett.* **114**, 98102 (2015).

[15] Takayama, Y., Inui, Y., Sekiguchi, Y., Kobayashi, A., Oroguchi, T., Yamamoto, M., *et al.* Coherent X-Ray Diffraction Imaging of Chloroplasts from Cyanidioschyzon merolae by Using X-Ray Free Electron Laser. *Plant Cell Physiol.* **56**, 1272–1286 (2015).

[16] van der Schot, G., Svenda, M., Maia, F. R. N. C., Hantke, M., DePonte, D. P., Seibert, M. M., *et al.* Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nat. Commun.* **6**, 5704 (2015).

[17] Tokuhisa, A., Taka, J., Kono, H. & Go, N. Classifying and assembling two-dimensional X-ray laser diffraction patterns of a single particle to reconstruct the three-dimensional diffrac-tion intensity function: Resolution limit due to the quantum noise. *Acta Crystallogr. A* **68**, 366–381 (2012).

[18] Tegze, M. & Bortel, G. Atomic structure of a single large bio-molecule from diffraction patterns of random orientations. *J. Struct. Biol.* **179**, 41–45 (2012).

[19] Tegze, M. & Bortel, G. Selection and orientation of different particles in single particle imaging. *J. Struct. Biol.* **183**, 389–393 (2013)

[20] Sorzano, C. O. S., Jonić, S., El-Bez, C., Carazo, J. M., De Carlo, S., Thévenaz, P., *et al.* A multiresolution approach to orientation assignment in 3D electron microscopy of single particles. *J. Struct. Biol.* **146**, 381–392 (2004).

[21] Loh, N.-T. D. & Elser, V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E Stat.*

Nonlin. Soft Matter Phys. **80**, 26705 (2009).

[22] Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* **3**, 977–990 (2008).

[23] Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 (2015).

[24] Nakano, M., Miyashita, O., Jonic, S., Tokuhisa, A. & Tama, F. Single-particle XFEL 3D reconstruction of ribosome-size particles based on Fourier slice matching: requirements to reach subnanometer resolution. *J. Synchrotron Radiat.* **25**, 1010–1021 (2018).

[25] Shechtman, Y., Eldar, Y. C., Cohen, O., Chapman, H. N., Miao, J. & Segev, M. Phase Retrieval with Application to Optical Imaging: A contemporary overview. *IEEE Signal Process. Mag.* **32**, 87–109 (2015).

[26] Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **15**, 1662 (1998).

[27] Miao, J., Ishikawa, T., Anderson, E. H. & Hodgson, K. O. Phase retrieval of diffraction patterns from noncrystalline samples using the oversampling method. *Phys. Rev. B* **67**, 174104 (2003).

[28] Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., *et al.* Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl. Acad. Sci. USA* **105**, 19684–19689 (2008).

[29] de la Rosa-Trevín, J. M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M., *et al.* Xmipp 3.0: an improved soft-ware suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).

[30] Reddy, H. K. N., Yoon, C. H., Aquila, A., Awel, S., Ayyer, K., Barty, A., *et al.* Coherent soft X-ray diffraction imaging of coliphage PR772 at the Linac coherent light source. *Sci. Data* **4**, 170079 (2017).

[31] Yabashi, M., Tanaka, H. & Ishikawa, T. Overview of the SACLA facility. *J. Synchrotron Radiat.* **22**, 477–484 (2015).

[32] Filipe, M. Single particle diffraction simulator, spsim. (2008).

[33] Song, C., Ramunno-Johnson, D., Nishino, Y., Kohmura, Y., Ishikawa, T., Chen, C. C., *et al.* Phase retrieval from exactly oversampled diffraction intensity through deconvolution. *Phys. Rev. B* **75**, 012102 (2007).

[34] Nakano, M., Miyashita, O., Jonic, S., Song, C., Nam, D., Joti, Y., *et al.* Three-dimensional reconstruction for coherent dif-fraction patterns obtained by XFEL. *J. Synchrotron Radiat.* **24**, 727–737 (2017).

[35] Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758-2769 (1982).

[36] Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., *et al.* High-resolution ab initio three-dimensional x-ray diffraction microscopy. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **23**, 1179–1200 (2006).

[37] Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., *et al.* Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* **470**, 78–81 (2011).

[38] Steinbrener, J., Nelson, J., Huang, X., Marchesini, S., Shapiro, D., Turner, J. J., *et al.* Data preparation and evaluation tech-niques for x-ray diffraction microscopy. *Opt. Express* **18**, 18598–18614 (2010).