Single-particle XFEL 3D reconstruction of ribosome-size particles based on Fourier slice matching: requirements to reach subnanometer resolution

Miki Nakano,[a] Osamu Miyashita,[a] Slavica Jonic,[b] Atsushi Tokuhisa[a,c] and Florence Tama[a,d,e]*

[a]Advanced Institute of Computational Science, RIKEN, 6-7-1 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan, [b]IMPMC, Sorbonne Universités – CNRS UMR 7590, UPMC Université Paris 6, MNHN, IRD UMR 206, Paris 75005, France, [c]RCH, RIKEN, 6-7-1 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan, [d]Department of Physics, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan, and [e]Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan.
*Correspondence e-mail: florence.tama@riken.jp

Three-dimensional (3D) structures of biomolecules provide insight into their functions. Using X-ray free-electron laser (XFEL) scattering experiments, it was possible to observe biomolecules that are difficult to crystallize, under conditions that are similar to their natural environment. However, resolving 3D structure from XFEL data is not without its challenges. For example, strong beam intensity is required to obtain sufficient diffraction signal and the beam incidence angles to the molecule need to be estimated for diffraction patterns with significant noise. Therefore, it is important to quantitatively assess how the experimental conditions such as the amount of data and their quality affect the expected resolution of the resulting 3D models. In this study, as an example, the restoration of 3D structure of ribosome from two-dimensional diffraction patterns created by simulation is shown. Tests are performed using the diffraction patterns simulated for different beam intensities and using different numbers of these patterns. Guidelines for selecting parameters for slice-matching 3D reconstruction procedures are established. Also, the minimum requirements for XFEL experimental conditions to obtain diffraction patterns for reconstructing molecular structures to a high-resolution of a few nanometers are discussed.

1. Introduction

The determination of the three-dimensional (3D) structure of biomolecules is of great importance for the understanding of their biological functions, which leads to the development of disease treatment and drug discovery. Single-particle imaging using femtosecond X-ray pulses from free-electron lasers (XFELs) is a new technique for observing the structure of biological samples in a state close to nature (Neutze *et al.*, 2000; Huldt *et al.*, 2003; Chapman *et al.*, 2006a, 2011; Gaffney & Chapman, 2007; Aquila *et al.*, 2015). The pulse of an XFEL beam is 10^9 times brighter than present-day third-generation X-ray synchrotron facilities (Miao *et al.*, 2015). This bright and coherent beam allows us to obtain diffraction data without crystallization, and their short femtosecond pulses enables measurements without radiation damage by recording the diffraction patterns before the specimen is destroyed, referred to as 'diffraction before destruction' (Neutze *et al.*, 2000; Gaffney & Chapman, 2007; Chapman *et al.*, 2011; Hirata *et al.*,

Beam intensity → / Number of images ↑

S02 M02 W02
S05 M05 W05
S10 M10 W10
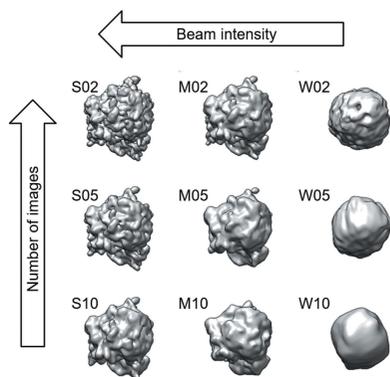
2014; Suga *et al.*, 2015). In addition, the XFEL beam can illuminate the inner structure of samples thicker than 500 nm without multiple-scattering events, which is an unavoidable problem in cryo-electron microscopy (cryo-EM).

The volume of XFEL experimental data is increasing and several low-resolution structures from the single-particle approach have been reported (Seibert *et al.*, 2011; Gallagher-Jones *et al.*, 2014; Kimura *et al.*, 2014; Xu *et al.*, 2014; Ekeberg *et al.*, 2015; Takayama *et al.*, 2015; van der Schot *et al.*, 2015; Hosseinizadeh *et al.*, 2017). Ekeberg *et al.* presented the 3D molecular structure of the giant mimivirus particle, reconstructed at 125 nm resolution from diffraction patterns obtained by XFEL experiments (Ekeberg *et al.*, 2015). Gallagher-Jones *et al.* observed the nanostructure formation of RNA interference microsponges using a combination of XFEL and synchrotron X-rays (Gallagher-Jones *et al.*, 2014). Kimura *et al.* demonstrated two-dimensional (2D) imaging of live cells using XFEL diffraction data at 28 nm full-period resolution (Kimura *et al.*, 2014). Recently, Hosseinizadeh *et al.* reported 3D reconstructions of PR772 virus structure at 9 nm resolution (Hosseinizadeh *et al.*, 2017). It has also been shown, theoretically, that high-resolution 3D structures could be obtained using millions of diffraction patterns (Tegze & Bortel, 2012; Tokuhisa *et al.*, 2012; Hosseinizadeh *et al.*, 2014) and that their dynamic properties could be directly interpreted from the 2D data (Tokuhisa *et al.*, 2016).

However, many challenging problems remain when constructing high-resolution 3D structures of biomolecules from XFEL diffraction data. Diffraction intensities from biomolecules are still low even with the bright XFEL pulse. Additionally, a large number of diffraction patterns need to be combined to obtain high-resolution structural information. Significant efforts are being devoted to increase the quantity and quality of data, such as beam focus, sample delivery method, signal-to-noise ratio (SNR) improvement, and the collection and selection of diffraction patterns (Miao *et al.*, 2015; Yabashi *et al.*, 2015). In addition, computational algorithms are required to estimate the laser beam incidence angles to the particle in each diffraction pattern, and the phase information in order to restore 3D molecular structure.

Currently, there are three major strategies to estimate the orientations of XFEL diffraction patterns: (i) methods based on 'maximum correlation coefficients' (Penczek *et al.*, 1994; Sorzano *et al.*, 2004; Yang & Penczek, 2008; Tegze & Bortel, 2012, 2013), (ii) the 'expand, maximize and compress (EMC)' algorithm (Loh & Elser, 2009), and (iii) the 'manifold-embedding' method (Schwander *et al.*, 2014). In the first two approaches, 3D volumes in Fourier space are reconstructed through iterative procedures. At each iteration step, experimental diffraction patterns are compared against a set of reference diffraction patterns that are created from a tentative 3D model, in order to estimate beam angles for each diffraction pattern. In the 'maximum correlation coefficient' approach, a single orientation is assigned to each diffraction pattern (Penczek *et al.*, 1994; Sorzano *et al.*, 2004; Yang & Penczek, 2008; Tegze & Bortel, 2012, 2013). Once the approximate orientations are obtained, the subsequent

searches for the correct beam angles can be restricted to a range close to the approximate values to reduce computational cost (Scheres *et al.*, 2008). In the EMC algorithm, a number of angular assignments are considered for each target diffraction pattern, which are used concurrently in the 3D volume reconstruction in Fourier space with relative weights based on the similarities between the target and reference diffraction patterns. This strategy significantly improves the convergence of slice matching, but the resolution of the resulting 3D structure may be overestimated (Cheng *et al.*, 2015). An approach to estimate the orientation and phase simultaneously has been proposed recently (Donatelli *et al.*, 2017). In the manifold-embedding method, each snapshot from a specific object orientation is projected onto a 3D hypersurface, by analysing similarities between diffraction patterns (Schwander *et al.*, 2014). In this method, diffraction patterns are compared only when they are classified as neighbors and consistency among the patterns in the assembled 3D volume is not imposed (Ayyer *et al.*, 2016).

A maximum cross-correlation algorithm for 3D reconstruction has been demonstrated to have better scalability when it is applied to a large number of diffraction patterns. Using the maximum cross-correlation algorithm, Tegze & Bortel assigned incident beam angles for 100000 diffraction patterns of NapAB protein simulated with 4×10^{14} photons μm^{-2} laser beam intensity, starting from random orientations (Tegze & Bortel, 2012). They demonstrated that the restored molecular structure had a good agreement with the Protein Data Bank (PDB) structure. Previously, we implemented our maximum cross-correlation method in *Xmipp* (de la Rosa-Trevín *et al.*, 2013), which is an image-processing software package primarily aimed at single-particle 3D cryo-EM, by extending it to treat diffraction data, and performed a study on experimental diffraction patterns of an aerosol nanoparticle obtained by tomographic coherent X-ray diffraction microscopy (CXDM) (Nakano *et al.*, 2017). Although the diffraction patterns obtained by CXDM and XFEL have some differences (Miao *et al.*, 2006; Barty *et al.*, 2008; Jiang *et al.*, 2010), they can be treated similarly. We could estimate the incident beam angles close to those used for tomographic experiment through careful calibration of parameters for analysing the correlations between the experimental and simulated diffraction patterns (Nakano *et al.*, 2017).

In the study presented here, we performed the reconstruction of the structure of a large biological molecule, ribosome, from the simulated diffraction data using our maximum cross-correlation approach. We tested multiple reconstructions using different diffraction data sets simulated with different beam intensities and with different numbers of diffraction patterns to examine how the quantity and quality of the diffraction patterns affect the resulting 3D model. We calibrated the parameters for the estimation of incident beam angles, such as matching region, number of reference patterns and interpolation parameters, in order to obtain a reliable 3D structure, and to suggest some guidelines for selecting suitable parameters. In addition, we discuss the conditions required to obtain the diffraction patterns from XFEL experiments that

are necessary to restore the molecular structure at certain resolutions.

2. Materials and methods

2.1. Method: reconstruction of 3D volume in Fourier space from diffraction patterns

To determine the orientation angles of the samples against the incident beam captured in each diffraction pattern, we have developed the 'slice matching' iteration protocol based on the projection matching protocol included in *Xmipp* (de la Rosa-Trevín *et al.*, 2013; Nakano *et al.*, 2017). Here we briefly summarize the slice-matching procedure (Fig. 1).

(1) Create the initial reference diffraction volume, I_{ref}^{init}, from experimental diffraction patterns using randomly assigned Euler angles (in this study, we call 3D structure in Fourier space 'volume').

(2) Create the 2D diffraction reference pattern library from a reference volume, I_{ref}, by discretizing a sphere in evenly distributed angular steps (Bunge & Baumgardner, 1995) using the central slice theorem. In this theoretical study, the slices are approximated as planes. Each experimental diffraction pattern is compared against reference patterns by calculating the cross-correlation coefficient (CC) (Appendix A), so that the angles of the best matching reference pattern can be treated as new estimated angles for the experimental pattern.

(3) Update the reference volume using experimental patterns and the current estimates of their orientations. To reconstruct I_{ref} from 2D diffraction patterns, a weight function based on the Kaiser–Bessel window is used for Fourier-space interpolation (Appendix B) (Lewitt, 1990; Abrishami *et al.*, 2015).

(4) Iterate steps (2) and (3) until the angle estimation reaches convergence. At the early stage of the slice matching iteration, each experimental pattern is compared against all reference patterns in the library created with a large sampling interval. As the iteration progresses, the reference pattern sampling interval becomes small (Nakano *et al.*, 2017) and only reference patterns created with angles close to the one currently assigned are examined.

(5) Reconstruct 3D volume using the angles estimated by slice matching with large-size diffraction patterns. The outer area of diffraction patterns is not used for angle estimation because the photon count is relatively low in this area and the use of diffraction patterns of a smaller size reduces the computational time. However, this high wavenumber area contains finer structural information in real space and we can expect to obtain reliable diffraction intensities in this area by averaging a large number of large-size diffraction patterns. Thus, we reconstructed the final 3D volume using the large-size diffraction patterns and the angles estimated by slice matching using the cropped diffraction patterns.

Because the diffraction intensities from biological molecules are weak, obtaining a sufficient photon count is a serious challenge, especially at high wavenumber pixels. Conversely, diffraction intensities at low wavenumber pixels are strong and sometimes saturate the detection range, hindering the determination of the overall shape of the molecule by phase retrieval procedures. Therefore we have to carefully set the matching region that is used for the calculation of CC between the diffraction patterns. In addition, to create the reference volume, I_{ref}, we have to adjust the interpolation parameters to map the diffraction intensity on 2D patterns to 3D volume. Furthermore, the comparison of 2D diffraction patterns between experimental and reference diffraction patterns can be performed using diffraction 'amplitude' or diffraction 'intensity' distributions. We performed multiple reconstruction trials using several combinations of the different parameters to examine their impact on 3D structure reconstruction.
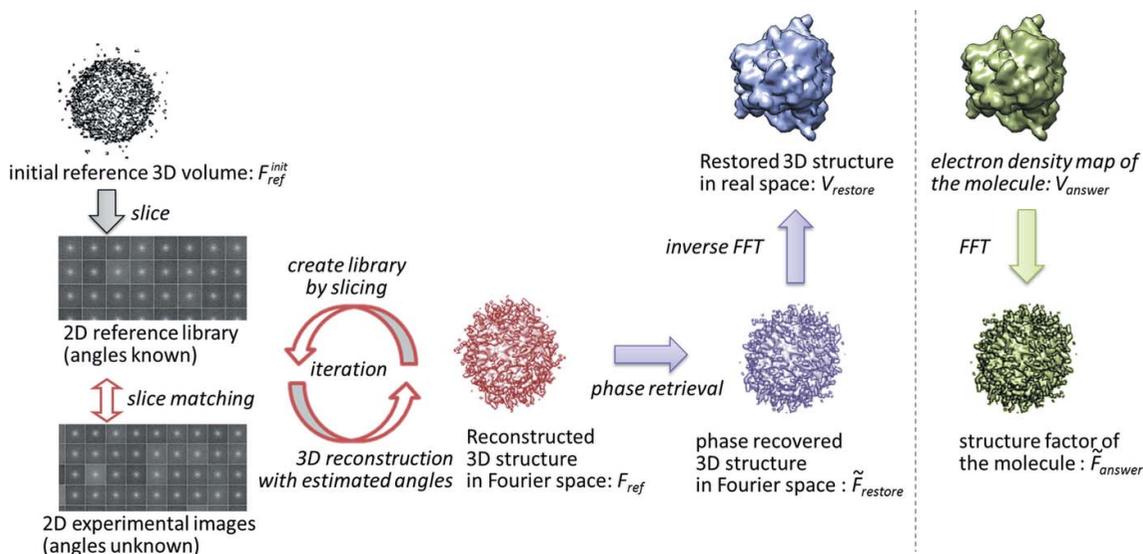

Figure 1
Schematic view of the 3D reconstruction protocol from XFEL experimental diffraction patterns. Each experimental pattern is compared against reference patterns created from the reference volume, where the best match is used to estimate its orientation and to update the reference volume. After the reference volume has converged, phase recovery is performed to restore the molecular structure in real space.

2.2. Phase retrieval

Phase retrieval is performed on the diffraction volumes reconstructed with large-size diffraction patterns after their orientation was estimated by slice matching using cropped patterns. We used the hybrid input–output phase retrieval approach. The support region was set to be a sphere having a diameter corresponding to the molecular size and this support region was the same for all iterations of the phase retrieval.

2.3. Evaluation of reconstructions

To assess the agreement between the reconstructed 3D structure factor amplitude, $F_{ref} = (I_{ref})^{1/2}$, and the ground-truth structure factor, $\tilde{F}_{answer}$, which is the Fourier transform of the electron density map, we calculated the R-factor as follows,

$$R\text{-factor} = \frac{\sum\limits_{i \in \text{volume}} \left| c|\tilde{F}_{answer}(i)| - F_{ref}(i) \right|}{\sum\limits_{i \in \text{volume}} F_{ref}(i)},$$

(1)

$$c = \frac{\sum\limits_{i \in \text{volume}} F_{ref}(i)}{\sum\limits_{i \in \text{volume}} |\tilde{F}_{answer}(i)|},$$

where c is the normalization factor to adjust the amplitude ranges between two diffraction volumes. We also calculated the wavenumber dependence of the discrepancy between these two volumes, R-factor(k), in the same manner,

$$R\text{-factor}(k) = \frac{\sum\limits_{k_i \in k} \left| c|\tilde{F}_{answer}(k_i)| - F_{ref}(k_i) \right|}{\sum\limits_{k_i \in k} F_{ref}(k_i)},$$

(2)

$$c(k) = \frac{\sum\limits_{k_i \in k} F_{ref}(k_i)}{\sum\limits_{k_i \in k} |\tilde{F}_{answer}(k_i)|},$$

where k_i is the voxel within the corresponding shell k in each volume.

To quantify how well molecular structure was restored in real space, we calculated the Fourier shell correlation (FSC) and phase retrieval transfer function (PRTF), which are commonly used to evaluate resolutions (Chapman et al., 2006b; Steinbrener et al., 2010; Seibert et al., 2011), as follows,

$$\text{FSC}(k) = \frac{\sum\limits_{k_i \in k} \tilde{F}_{restore}(k_i) \tilde{F}_{answer}(k_i)^*}{\left[\sum\limits_{k_i \in k} |\tilde{F}_{restore}(k_i)|^2 \sum\limits_{k_i \in k} |\tilde{F}_{answer}(k_i)|^2 \right]^{1/2}}$$

(3)

$$\text{PRTF}(k_i) = \frac{\left\langle |\tilde{F}_{restore}(k_i)| \exp(i\varphi_{k_i}) \right\rangle}{F_{ref}(k_i)}$$

(4)

where $\tilde{F}_{restore}$ is the structure factor derived from converged I_{ref} with retrieved phases φ, and $\langle \ldots \rangle$ denotes an average over independent reconstructions. FSC measures the normalized cross-correlation coefficient between two 3D volumes over

Table 1
Definitions of the representation for 3D volumes in Fourier space.

Name	Phase information	Definition
I_{ref}	No	3D diffraction intensity distribution used for the creation of reference pattern library
F_{ref}	No	Reconstructed 3D structure factor amplitude, square root of I_{ref}
$\tilde{F}_{restore}$	Yes	Structure factor, with phase recovered from converged I_{ref}
$\tilde{F}_{answer}$	Yes	Structure factor of the molecule, Fourier transformed from electron density map

corresponding shells in Fourier space. PRTF represents the confidence in the retrieved phases as a function of resolution. We calculated the wavenumber dependence of PRTF(k) in this study by averaging over shells of among the constant wavenumber k. Notations for 3D volumes used in this study are shown in Table 1.

2.4. Simulated diffraction pattern dataset

In this study, we performed a 3D structure reconstruction of the ribosome from 2D coherent diffraction patterns as an example application of the XFEL method for large biological molecules. The ribosome's biological importance as a complex responsible for biological protein synthesis (Selmer et al., 2006; Polikanov et al., 2015; Sierra et al., 2016) and its large molecular size makes it a suitable XFEL analysis target.

We chose the crystal structure of Thermus thermophilus 70S ribosome bound with release factor RF2 from the Protein Data Bank [PDB ID 4v67 (Korostelev et al., 2008), molecular size $\simeq$ 32 nm] as the target structure, and converted it to an electron density map with a 0.4 nm pixel^{-1} resolution in real space by using Xmipp (de la Rosa-Trevín et al., 2013). The map was first converted to structure factor, $\tilde{F}_{answer}$, by Fourier transform with a padding factor of 4 to leave sufficient space around the support region in real space. The 2D diffraction intensity distribution patterns were generated by taking slices of 3D diffraction intensity distribution obtained from the square-modulus of $\tilde{F}_{answer}$. Poisson noise corresponding to the tested beam intensities was applied onto the simulated diffraction patterns, and used as 'experimental' XFEL diffraction patterns.

Finally, we prepared nine sets of diffraction patterns combining three different beam intensities and three different numbers of patterns per set. Diffraction pattern sets were created with three different angle sampling intervals (Bunge & Baumgardner, 1995), 2, 5 and 10°, with Gaussian noise to shift the slice angle from the grid points, producing 10242, 1692 and 362 patterns, respectively. Three different beam intensities (S: strong; M: medium; W: weak) were considered. The intensities, as estimated by comparing against outputs from spsim (Filipe, 2008) with wavelength 0.1 nm, quantum efficiency of the detector 0.8 and over-sampling ratio 4, were: S, 5.5×10^{13}; M, 5.5×10^{12}; W, 5.5×10^{11} photons µm^{-2} (summarized in Table 2). The diffraction pattern size was 320 pixel × 320 pixel (wavenumber at the edge is 1.25 nm^{-1}).

Table 2
Conditions for slice creation.

Name	S (Strong)	M (Medium)	W (Weak)
Beam intensity (photons μm^{-2})	5.5×10^{13}	5.5×10^{12}	5.5×10^{11}
Angle sampling interval (°)	2	5	10
Number of diffraction patterns	10242	1692	362

Table 3
Matching region and interpolation parameters which resulted in the best restored structures.

Beam intensity	Matching and interpolation parameters
Strong	$q_{in} = 20$, $q_{out} = 30$, $\alpha = 15$, $\eta = 2$ for S02 and S05
	$q_{in} = 10$, $q_{out} = 20$, $\alpha = 10$, $\eta = 2$ for S10
Medium	$q_{in} = 10$, $q_{out} = 20$, $\alpha = 5$, $\eta = 2$ for M02, M05 and M10
Weak	$q_{in} = 5$, $q_{out} = 10$, $\alpha = 5$, $\eta = 2$ for W02, W05 and W10

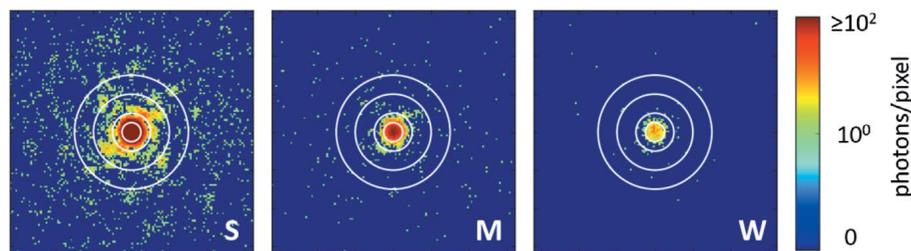To reduce computational time for the slice-matching iteration, these patterns were cropped to 128 pixel × 128 pixel (the wavenumber at the edge is 0.5 nm^{-1}). Representative diffraction patterns created with three beam intensities are shown in Fig. 2. Hereafter, we denote the pattern set with the combination of beam intensity and sampling interval, such as S05, denoting that the slice set was created with the strong beam intensity (S) and 5° angle sampling interval (05).

3. Results

3.1. Selection of the slice-matching region

In the calculation of the CC between experimental diffraction patterns and those in the reference library, we excluded the center and outer regions of the diffraction patterns to improve the sensitivity of the slice matching (Nakano *et al.*, 2017). In the center region of the diffraction patterns, intensities are often too strong to be measured, and often protected by a beam stopper. On the outer region, intensities are usually quite weak and the SNR is low. Therefore, we only calculated the CC for the annular regions defined by the inner and outer radii, q_{in} and q_{out}, as shown in Fig. 3(a).
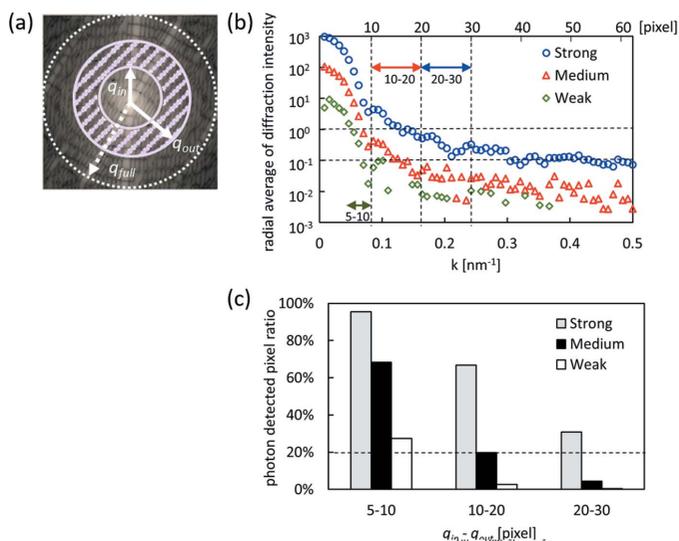
Fig. 3(b) shows the radial average of diffraction intensities on a slice created with different beam intensities. Diffraction intensities decreased with increasing wavenumber, and the range exceeded four orders of magnitude. Fig. 3(c) shows the percentage of the number of pixels where more than one photon was detected within the region between q_{in} and q_{out}. Combinations of two radius parameters were tested (Table 3) and we found that the matching region where the ratio of pixels with photon counts was around 20% worked well for each beam intensity: $q_{in} = 20$ and $q_{out} = 30$ for the strong beam intensity, $q_{in} = 10$ and $q_{out} = 20$ for the medium beam intensity, and $q_{in} = 5$ and $q_{out} = 10$ for the weak beam intensity. For S10,
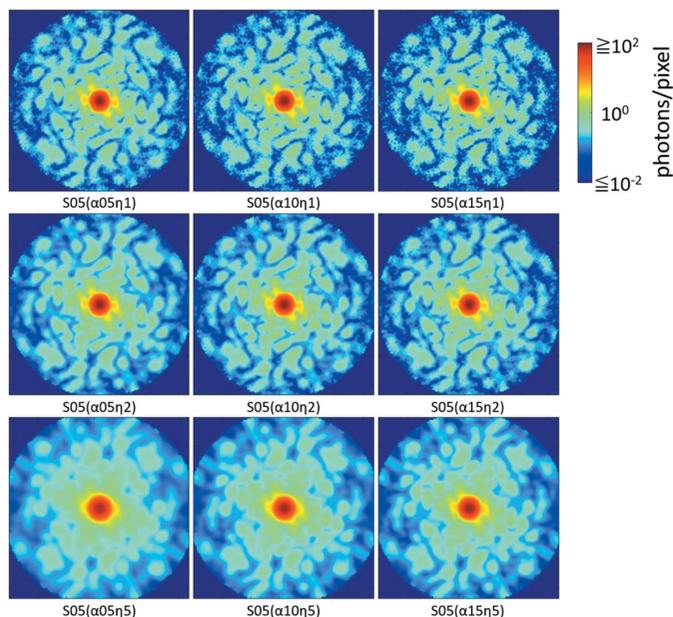
Figure 3
(*a*) Definition of matching region on the diffraction pattern used for slice matching. Only the region filled with pink stripes was used for calculation of CC. q_{in} and q_{out} show the inner and outer radius in pixels from the center of the pattern. (*b*) Radial average of diffraction intensity on the diffraction patterns for each of the three beam intensities tested. Blue, red and green arrows indicate the matching region with $q_{in} - q_{out} = 20$–30 pixels, 10–20 pixels and 5–10 pixels for strong, medium and weak beam intensities, respectively. (*c*) Photon detected pixel ratio within the region between q_{in} and q_{out} on the diffraction pattern.

the restoration of molecular structure performed better with $q_{in} = 10$ and $q_{out} = 20$ instead of $q_{in} = 20$ and $q_{out} = 30$. This would probably be because with a smaller number of diffraction patterns the diffraction intensity distribution at higher wavenumber voxels in the reconstructed diffraction volume becomes sparse. We note that the annular regions defined here correspond to where the diffraction intensity was between 0.1 and 1.0 on average (Fig. 3b).

3.2. Choice of interpolation parameters

To map the diffraction intensities on 2D patterns to 3D volume, we used a weight function based on the Kaiser–Bessel window, $w(\alpha, \eta; d_{kj})$. The value of $w(\alpha, \eta; d_{kj})$ depends on the distance d_{kj} between the position of k and j within the volume; k is the center of the voxel where the diffraction intensity is being calculated, and j is the mapped position of pixel i on the 2D pattern. With a large η, diffraction intensity would be inter-

Figure 2
Representative diffraction patterns for each beam intensity. The pattern size is 128 pixel × 128 pixel, and the wavenumber at the edge corresponds to 0.5 nm^{-1}. The radii of the circles on the patterns are 5, 10, 20 and 30 pixels.

Figure 4
Cross-section views of 3D diffraction intensity distribution reconstructed using ground-truth angles from 1692 diffraction patterns created with strong beam intensity, S05. α and η are the interpolation parameters used for the reconstruction.

polated using the pixels from 2D patterns that are mapped farther in the volume. With a large α, weight for interpolation would be decreased quickly as d_{kj} increases. The detailed procedure of this interpolation was described in our previous study (Nakano *et al.*, 2017).

To examine the effect of the various parameters on the reconstruction process, we performed multiple reconstruction trials with different matching and interpolation parameters. Fig. 4 shows the cross-section views of 3D diffraction intensity distribution reconstructed from the diffraction pattern sets, S05, using ground-truth angles that were used for the creation of experimental diffraction patterns, with different interpolation parameters. The cross-section views became blurred with

increasing η, and the smoothness between adjacent pixels was increased with decreasing α.

The combinatorial effect of parameters and its correlation to the quality of the restored structures as compared with the ground-truth structure is summarized in Table 3. Interpolation parameters α and η were determined empirically. In this study, the parameter $\eta = 2$ produced the best restored structure for all pattern sets. The parameter α was decreased with decreasing beam intensity and the number of patterns contained in the pattern set.

3.3. Amplitude distributions *versus* intensity distributions when comparing diffraction patterns

For all experimental pattern sets, we performed 30 iterations of slice matching, which ensured volume convergence. The angle search parameters used in this study are shown in Table S1 of the supporting information. We reconstructed 3D volume in Fourier space using diffraction patterns of a larger size (320 pixel × 320 pixel) with the angles estimated by slice-matching iteration with patterns of a smaller size (128 pixel × 128 pixel). The wavenumbers at the pattern edges correspond to 1.25 nm^{-1} and 0.5 nm^{-1}, respectively.

The CC can be calculated using either the 2D diffraction amplitude distribution or 2D diffraction intensity distribution. To assess which CC evaluation scheme performs best, we calculated the R-factors and the average angle errors between ground truth and the estimated angles for each slice after the mentioned 30 iterations of slice matching (Fig. 5). For both calculations, F_{ref} was aligned to $|\tilde{F}_{answer}|$ to maximize the correlation coefficient between those two volumes.

In Fig. 5, for strong and medium beam intensities, comparison of patterns worked better using 2D diffraction amplitude distributions instead of intensity distributions, especially regarding the angle error shown in Fig. 5(*b*). For the weak beam intensity, the angle errors were too large using both intensity and amplitude distributions, indicating that diffraction intensities were too weak for accurate slice matching. Therefore, we use the diffraction amplitude distribution in our 2D diffraction pattern comparisons.
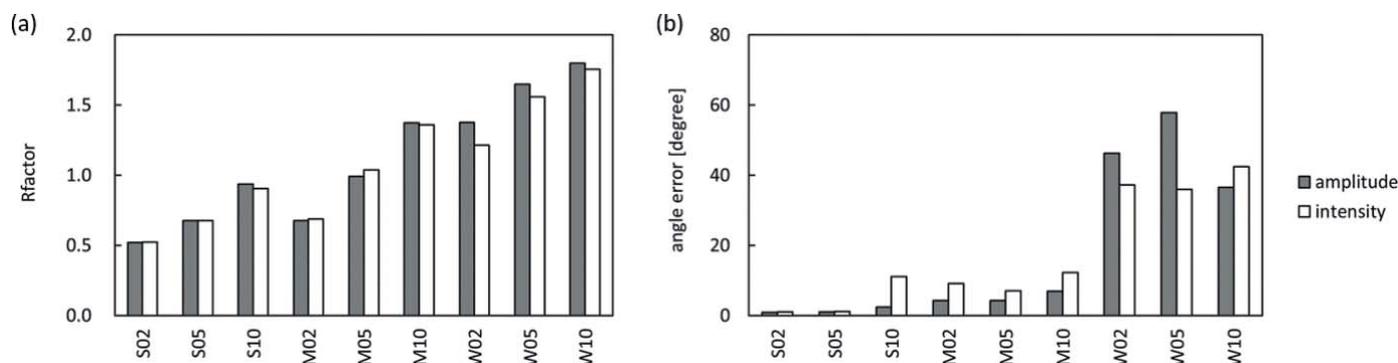
Figure 5
Differences in the slice-matching results between using the 2D diffraction amplitude distribution and the 2D diffraction intensity distribution. (*a*) R-factor differences between the ground-truth volume and reconstructed 3D volume using estimated angles from slice matching. (*b*) Angle error average between the ground-truth and estimated angles for each slice.

3.4. Effect of beam intensity and number of patterns on the accuracy of 3D reconstruction

Using the selected parameters, 3D diffraction volumes are reconstructed after the slice matching. In Fig. 6, we observe the dependence of the *R*-factor on the wavenumber for the reconstructed volumes. A comparison of the resulting 3D volumes against the ground-truth model shows the importance of beam intensity for 3D structure reconstruction. Using the datasets constructed with weak beam intensity, the resulting *R*-factor and angle errors are significantly higher as shown in Figs. 5 and 6(*c*). The number of patterns used in the reconstruction is also important, and using a larger number of diffraction patterns reduced the errors. The heights observed at 0.08 nm^{-1} (12.5 nm in real space) shown in Fig. 6 correspond to the basins observed in the radial average of diffraction intensity on the diffraction patterns (Fig. 3*b*), corresponding to the edge of the molecule in real space.

3.5. Phase retrieval and 3D structure reconstruction in real space

After slice matching, phase recovery was performed on the diffraction volumes reconstructed from the larger-size patterns. We used the hybrid input–output phase retrieval
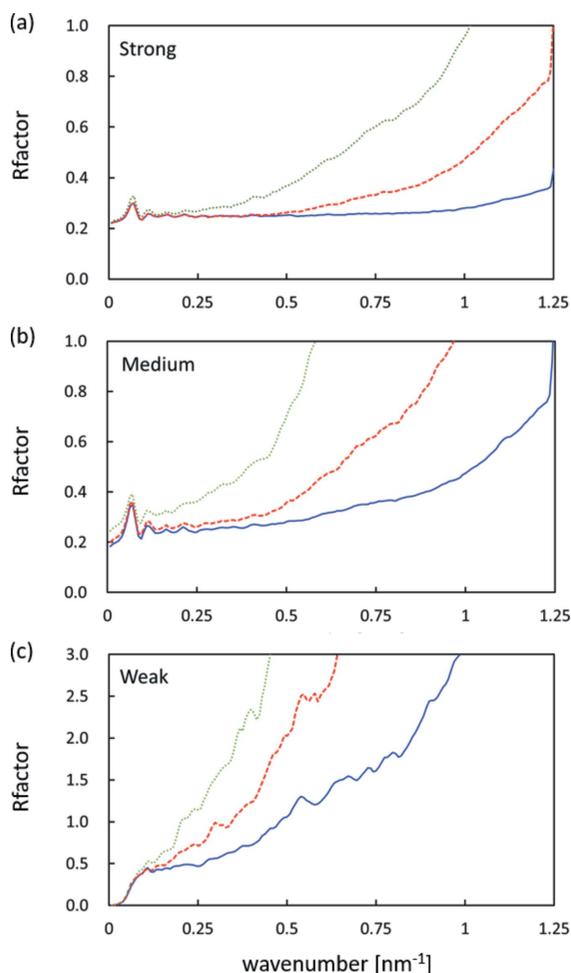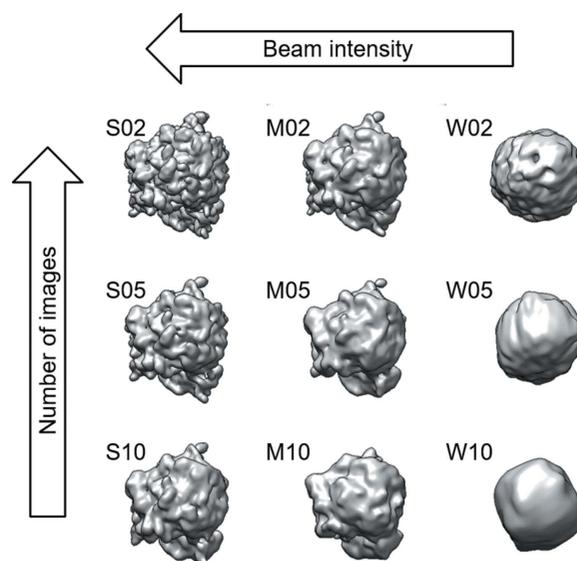

Figure 7
Restored 3D structures in real space. The figures were drawn using *Chimera* (Pettersen *et al.*, 2004).

approach, starting with ten random phases, where a sphere with a diameter of 32 nm (to include the entire molecular complex) was used to fix the support region. The parameters used for phase retrieval are shown in Table S2.

Fig. 7 shows the restored 3D structures in real space, which represent the aligned and averaged results of ten phase retrieval trials. The quality of the recovered 3D structures in real space was assessed by FSC and PRTF [equations (3) and (4)] for the reconstructions from all nine diffraction pattern sets (Fig. 8). Table 4 shows the resolution of restored 3D molecular structures at FSC = 0.5 resolution cut off and PRTF = 1/e cut off, which are commonly used thresholds in the cryo-EM and XFEL literature (Böttcher *et al.*, 1997; Rosenthal & Henderson, 2003; Ekeberg *et al.*, 2015). FSC and PRTF were strongly correlated, and both FSC and PRTF increase with increasing beam intensity and number of diffraction patterns.

With strong and medium beam intensities, the molecular structures were well restored, with the resolutions in the range ~1–4 nm. With weak beam intensity, structural details could not be recovered, and FSC and PRTF quickly decreased even with large numbers of diffraction patterns. These results are expected from the angle estimation errors (Fig. 5*b*), which shows that the slice matching did not work well for the slices created by weak beam intensity. With weak beam intensity data, diffraction intensities were too small to restore the molecular shape even from larger numbers of patterns. These results suggest the requirements for a subnanometer-resolution 3D structure restoration from XFEL data, as we will discuss later.

3.6. Cross-validation of the reconstructions of molecular structure

Our reconstruction protocol depends on two initial conditions: one is the initial reference volume in Fourier space,

Figure 6
Wavenumber dependence of the *R*-factor.

Table 4
Resolution of restored 3D molecular structures in nm.

Number of patterns [angle sampling interval (°)]	Resolution in nm at FSC = 0.5			Resolution in nm at PRTF = 1/e		
	Strong	Medium	Weak	Strong	Medium	Weak
10242 (02)	0.9	1.7	12.5	1.3	2.0	10.0
1692 (05)	1.5	2.3	8.3	1.7	2.9	10.0
362 (10)	2.1	3.7	10	2.2	4.0	10.0
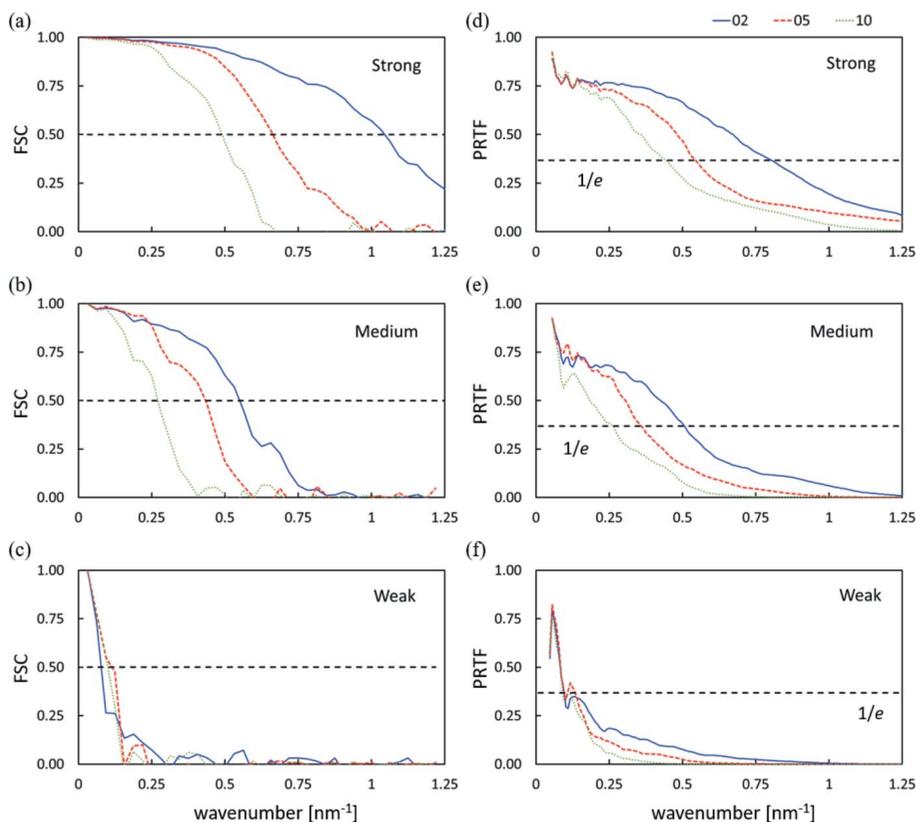
Figure 8
(a)–(c) Fourier shell correlation (FSC) between restored molecular structures and electron density map created from PDB registered structure. (d)–(f) Phase retrieval transfer function (PRTF) among the restored molecular structures obtained by each phase retrieval trial.

which is generated using random angular assignment, used for slice matching, and the other is the randomly assigned initial phase angles used for phase retrieval. To evaluate the reproducibility and the reliability of the restored molecular structure, diffraction patterns (S02, 10242 patterns) were split into two subsets (even and odd), and the slice matching was performed for each subset independently, with parameters determined previously (Table 3). FSCs between $\tilde{F}_{answer}$ and $\tilde{F}_{restore}$ for each restored structure were rather similar, indicating good reproducibility of the slice matching protocol (Fig. 9). Fig. 9 also shows the FSC curve between the 3D structures restored from even and odd pattern subsets. The FSC between two restored structures was higher than compared with those between each $\tilde{F}_{restore}$ and $\tilde{F}_{answer}$, indicating that there were no contradictions among our restored structures using different pattern subsets obtained under the same experimental conditions. We note here that the resolutions estimated from the restored structures were over-

estimated when compared with the resolution of the ground-truth model.

4. Discussion

4.1. Determining the parameters for successful slice matching

To estimate the incident beam angles, three parameters need to be selected during the slice-matching procedure: location of the matching regions, interpolation parameters, and the use of diffraction amplitude or intensity for pattern comparisons.

Since the diffraction intensity changes widely from the center to the outer region of the patterns, it is difficult to find subtle differences between two patterns if the whole pattern is used for CC calculations. Thus, by excluding certain areas on the patterns we increase the sensitivity of the pattern comparisons. Reducing the matching region is also effective in reducing the calculation time. We found that slice matching works well using the regions where the ratio of the pixels with photon counts are around 20%. For a smaller number of patterns in a dataset, better results were obtained when the regions close to center were used for matching.

To map the diffraction intensity on 2D pattern to 3D volume, we used the weight function based on the Kaiser–Bessel window, $w(\alpha, \eta; d_{kj})$, which depends on the distance between the center position k of the calculated voxel and the mapped position j in the 3D volume of the pixel i on the 2D pattern. The term α regulates the decreasing rate of the
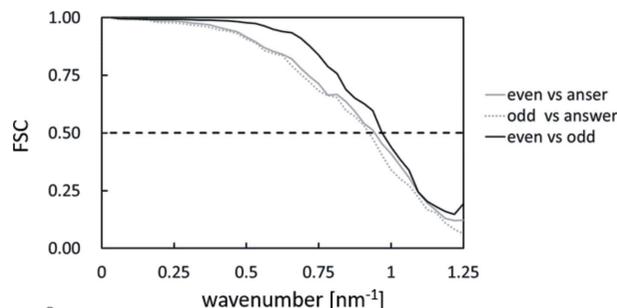
Figure 9
Cross validation of the reconstruction of the molecular structure. The diffraction pattern set of S02 is split into two subsets, even and odd. Slice-matching iterations were performed with $q_{in} = 20$, $q_{out} = 30$, $\alpha = 15$, $\eta = 2$. Gray solid and dotted lines represent the FSC curve between $\tilde{F}_{restore}$ from each set and $\tilde{F}_{answer}$. The black solid line represents the FSC curve between $\tilde{F}_{restore}$ from even and odd pattern sets.

weight. With a smaller α value, further voxels are taken into account for diffraction intensity estimation. In our study, smaller values of α worked better for the pattern set with the fewer photon counts or lower diffraction intensities, because small α can compensate low photon counts in the reconstructed volume in Fourier space. The term η determines the maximum interpolation length from the mapped position j in the volume. Because a larger η value makes the diffraction pattern blurred within the reconstructed volume, the differences in the CC values between the reference slices decreases. In this study, $\eta = 2$ was the best value for all pattern sets. In addition, we found that the 2D diffraction pattern matching works better using diffraction amplitude distributions instead of intensity distributions.

Because our slice matching protocol only uses the annular region following the criteria described above, the smaller dimensional size of the diffraction pattern is sufficient for angle estimation. After the convergence of the slice matching, the reference volume can be updated with larger size diffraction patterns using estimated incident beam angles with smaller size of patterns, and this volume can then be used for the phase retrieval. This protocol significantly reduces the computational cost and uses the information stored at the high wavenumber region effectively.

4.2. Accuracy of the estimated angles and the resolution of the retrieved 3D structure

In order to evaluate the effects of the angle estimation error on the resolution of the reconstructed 3D volumes, we estimated how the angle errors, e (Fig. 5), affect the positions of diffraction pattern pixels d at q_{out} in the 3D diffraction volume, such that $d = 2\pi q_{out} e/360°$, in pixels (Table S3). Using the diffraction pattern set consisting of a large number of patterns created by the strong beam intensity, we achieved an angle error of 0.89°, which translates to a position error of 0.46 pixels at $q_{out} = 30$ pixel. For the medium beam intensity, the average position error was about 1.5 pixels at $q_{out} = 20$ pixels. For the weak beam intensity, the position error was significantly larger.

In addition, we calculated the distances between the angles used to create the experimental diffraction patterns and the angles of the closest reference patterns (the angle sampling interval of the reference patterns at the last iteration was 1°), and calculated the average of these distances. This value would be the theoretical minimum angular error for our data, and they were about 0.28° for all image sets. Although the average angle errors resulting from our slice-matching algorithm are larger than this theoretical minimum error, for the reconstruction with the diffraction pattern set with a large number of patterns created by the strong beam intensity the achieved angle error is approaching the theoretical limit. On the other hand, for the medium beam intensity, the angle errors are larger because of the limited signal at high angle regions, which results in 3D reconstructions with lower resolutions.
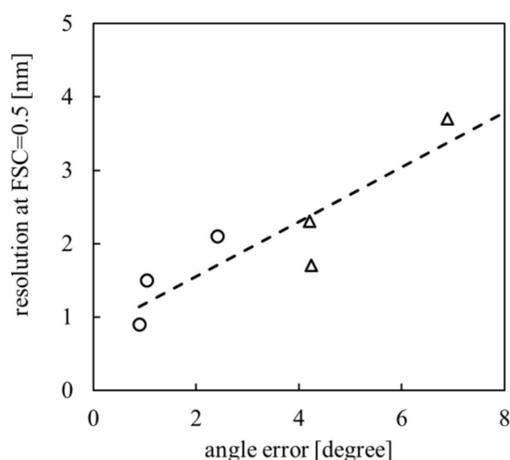
Figure 10
Relationship between the angle error and the resolution of the 3D reconstruction of the molecule estimated by FSC.

We also examined the correlation between the average angle errors and the FSC (0.5) based on the resolutions of the reconstructed real space structures (Fig. 10). The final resolution linearly increased as the angle error increased for the strong and medium beam intensities. The reconstructions from data with the weak beam intensity are not included in this analysis since they do not have sufficient quality. We estimated 'the best resolution' that could be obtained if there was no error in the angle estimation by performing reconstruction from S02 dataset using ground-truth angles. The FSC between this 'ground truth' and the original 3D volume from the PDB was 0.87 nm. This resulting resolution includes the error from the phase recovery procedure, and it approaches the highest frequency in the diffraction patterns (1/0.8 nm). This analysis clearly shows that the slice-matching protocol can estimate the angles accurately, and, at the same time, demonstrates the importance of accurate angle estimation for 3D reconstruction.

4.3. Requirements for experimental conditions to achieve molecular structure resolution

Beam intensity dominantly affects the resolution of the restored molecular structures. At the same time, a larger number of patterns is required to improve the resolution. Averaging a large number of patterns can improve the SNR as high-resolution regions on the diffraction pattern have low photon counts and are noisy. We showed that the medium wavenumber regions are sufficient for accurate angle estimation, and by averaging the photon count at high wavenumber regions from a large number of diffraction patterns we could obtain a higher resolution for the restored molecular structure.

Recently, Ekeberg et al. reconstructed a low-resolution structure of the giant mimivirus from XFEL diffraction patterns obtained using LCLS XFEL at SLAC (Ekeberg et al., 2015). The beam intensity that they used in the experiment was 1.2×10^{10} photons μm^{-2} in the center of the beam. More recently, Reddy et al. obtained the diffraction patterns of

Coliphage PR772 also using LCLS XFEL apparatus, using a 1.5 µm-diameter beam containing about 10^{13} photons µm^{-2} brilliance (Reddy *et al.*, 2017). XFEL beam intensity has been greatly improved, and the beam intensity used for the more recent experiments is closer to our parameter for 'strong' beam intensity (5.5×10^{13} photons µm^{-2}), although this estimated value was obtained by assuming high detector quantum efficiency (0.8 in this case).

We note that our computational experiments presented here are an idealized case. It is still difficult to obtain more than a thousand diffraction patterns of the sample under the same conditions by XFEL. In addition, our computer-simulated diffraction patterns only consider Poisson noise. The raw diffraction patterns obtained by experiments contain several kinds of experimental noise, and often have missing regions due to the limitations of the equipment. These issues are not trivial, and require significant efforts for the development of computational algorithms and the improvement of experimental techniques to achieve molecular structures that are less than 1 nm in resolution.

5. Conclusion

We performed the 3D reconstruction of ribosome structures from 2D diffraction patterns created by simulations under various experimental conditions in order to assess how the quantity and quality of the data affect the resolution of the resulting 3D model. We have confirmed that our protocol showed good reproducibility, and have provided some guidelines for selecting parameters to perform slice matching. We also estimated the experimental conditions required to obtain 1 nm resolution for a recovered molecular structure; above 10000 diffraction patterns created using a beam intensity of above 10^{13} photons µm^{-2} are required. The experimental conditions we describe have yet to be achieved but developments in the field indicate that it would be achievable in the near future. Insights from this study could be significant in the application of 3D reconstruction algorithms in XFEL single-particle experiments.

APPENDIX A
Calculating the zero-mean normalized cross-correlation coefficient

The zero-mean normalized cross-correlation coefficient between each experimental diffraction intensity distribution pattern and all reference diffraction intensity distribution patterns, CC$_{\mathrm{intensity}}$, was calculated using the following equation,

$$\mathrm{CC}_{\mathrm{intensity}} =$$

$$\frac{(1/N_{\mathrm{pix}}) \sum_{i}^{N_{\mathrm{pix}}} \left\{ \left[M_{\mathrm{exp},p}(i) - \overline{M_{\mathrm{exp},p}} \right] \left[M_{\mathrm{ref},q}^{\psi}(i) - \overline{M_{\mathrm{ref},q}} \right] \right\}}{\sigma_{M_{\mathrm{exp},p}} \sigma_{M_{\mathrm{ref},q}}}. \quad (5)$$

$M_{\mathrm{exp},p}(i)$ and $M_{\mathrm{ref},q}(i)$ are the diffraction intensities at pixel i of the pth experimental and qth reference diffraction patterns,

respectively (p = 1 to N_{exp}, q = 1 to N_{ref}). N_{pix} is the number of pixels within q_{in} and q_{out} in each diffraction pattern. $\overline{M_{\mathrm{exp},p}}$ and $\overline{M_{\mathrm{ref},q}}$ are the average intensities of pth experimental and qth reference diffraction patterns, and $\sigma_{M_{\mathrm{exp},p}}$ and $\sigma_{M_{\mathrm{ref},q}}$ are their standard deviations, respectively. $M_{\mathrm{ref},q}^{\psi}(i)$ is the diffraction intensity of the qth reference pattern rotated with angle ψ in the plane to maximize the CC. CC$_{\mathrm{amplitude}}$ was also calculated with diffraction amplitudes of experimental and reference diffraction patterns in the same manner,

$$\mathrm{CC}_{\mathrm{amplitude}} = \left[(1/N_{\mathrm{pix}}) \sum_{i}^{N_{\mathrm{pix}}} \left(\left\{ \left[M_{\mathrm{exp},p}(i) \right]^{1/2} - \overline{\left[M_{\mathrm{exp},p} \right]^{1/2}} \right\} \right. \right.$$
$$\left. \left. \times \left\{ \left[M_{\mathrm{ref},q}^{\psi}(i) \right]^{1/2} - \overline{\left[M_{\mathrm{ref},q} \right]^{1/2}} \right\} \right) \right]$$
$$\Big/ \sigma_{[M_{\mathrm{exp},p}]^{1/2}} \; \sigma_{[M_{\mathrm{ref},q}]^{1/2}}. \quad (6)$$

The incident beam angles used to create the reference pattern having the maximum CC$_{\mathrm{intensity}}$ or CC$_{\mathrm{amplitude}}$ were assigned to the experimental pattern.

APPENDIX B
Calculating the diffraction intensity

To calculate the diffraction intensity at voxel k in the reconstructed volume, $I_{\mathrm{ref}}(k)$, a weight function based on the Kaiser–Bessel window is used (Lewitt, 1990; Abrishami *et al.*, 2015),

$$I_{\mathrm{ref}}(k) = \sum_{p=1}^{N_{\mathrm{exp}}} \sum_{i=1}^{N_{\mathrm{pix}}} w(d_{kj}) \, M_{\mathrm{exp},p}(i) \Big/ \sum_{p=1}^{N_{\mathrm{exp}}} \sum_{i=1}^{N_{\mathrm{pix}}} w(d_{kj}), \quad (7)$$

$$w(d_{kj}) = \frac{\xi(\alpha, \eta)}{I_0(\pi\alpha)} \, I_0 \left(\pi\alpha \sqrt{1 - \left(\frac{2d_{kj} - \eta}{\eta - 1} - 1 \right)^2} \right) \quad (8)$$

for $0 \le d_{kj} \le \eta$. d_{kj} is the distance between the position k and j within the reconstructed volume, $I_{\mathrm{ref}}(k)$ is the center position of the voxel k, and j is the position of the $M_{\mathrm{exp},p}(i)$ map in the 3D volume. I_0 is the zeroth-order modified Bessel function, η is the maximum radius for interpolation, and α is a variable which determines the decreasing rate of $w(d_{kj})$. $\xi(\alpha, \eta)$ is the normalization factor determined by α and η.

research papers

References

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

Aquila, A. *et al.* (2015). *Struct. Dyn.* **2**, 041701.

Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. (2016). *J. Appl. Cryst.* **49**, 1320–1335.

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

Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.

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

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

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

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

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

Donatelli, J. J., Sethian, J. A. & Zwart, P. H. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 7222–7227.

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.

Filipe, M. (2008). *spsim, single particle diffraction simulator*, http://xray.bmc.uu.se/~filipe/?q = hawk/spsim/.

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

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

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

Hosseinizadeh, A., Mashayekhi, G., Copperman, J., Schwander, P., Dashti, A., Sepehr, R., Fung, R., Schmidt, M., Yoon, C. H., Hogue, B. G., Williams, G. J., Aquila, A. & Ourmazd, A. (2017). *Nat. Methods*, **14**, 877–881.

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

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

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

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

Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W. G. & Noller, H. F. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 19684–19689.

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

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

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

Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. (2015). *Science*, **348**, 530–535.

Nakano, M., Miyashita, O., Jonic, S., Song, C., Nam, D., Joti, Y. & Tama, F. (2017). *J. Synchrotron Rad.* **24**, 727–737.

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

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

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Polikanov, Y. S., Melnikov, S. V., Söll, D. & Steitz, T. A. (2015). *Nat. Struct. Mol. Biol.* **22**, 342–344.

Reddy, H. K. N., Yoon, C. H., Aquila, A., Awel, S., Ayyer, K., Barty, A., Berntsen, P., Bielecki, J., Bobkov, S., Bucher, M., Carini, G. A., Carron, S., Chapman, H., Daurer, B., DeMirci, H., Ekeberg, T., Fromme, P., Hajdu, J., Hanke, M. F., Hart, P., Hogue, B. G., Hosseinizadeh, A., Kim, Y., Kirian, R. A., Kurta, R. P., Larsson, D. S. D., Duane Loh, N., Maia, F. R. N. C., Mancuso, A. P., Mühlig, K., Munke, A., Nam, D., Nettelblad, C., Ourmazd, A., Rose, M., Schwander, P., Seibert, M., Sellberg, J. A., Song, C., Spence, J. C. H., Svenda, M., Van der Schot, G., Vartanyants, I. A., Williams, G. J. & Xavier, P. L. (2017). *Sci. Data*, **4**, 170079.

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

Rosenthal, P. B. & Henderson, R. (2003). *J. Mol. Biol.* **333**, 721–745.

Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. (2008). *Nat. Protoc.* **3**, 977–990.

Schot, G. van der, Svenda, M., Maia, F. R. N. C., Hantke, M., DePonte, D. P., Seibert, M. M., Aquila, A., Schulz, J., Kirian, R., Liang, M., Stellato, F., Iwan, B., Andreasson, J., Timneanu, N., Westphal, D., Almeida, F. N., Odic, D., Hasse, D., Carlsson, G. H., Larsson, D. S. D., Barty, A., Martin, A. V., Schorb, S., Bostedt, C., Bozek, J. D., Rolles, D., Rudenko, A., Epp, S., Foucar, L., Rudek, B., Hartmann, R., Kimmel, N., Holl, P., Englert, L., Duane Loh, N., Chapman, H. N., Andersson, I., Hajdu, J. & Ekeberg, T. (2015). *Nat. Commun.* **6**, 5704.

Schwander, P., Fung, R. & Ourmazd, A. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130567.

Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A.,

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

Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R. & Ramakrishnan, V. (2006). *Science*, **313**, 1935–1942.

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

Steinbrener, J., Nelson, J., Huang, X., Marchesini, S., Shapiro, D., Turner, J. J. & Jacobsen, C. (2010). *Opt. Express*, **18**, 18598–18614.

Suga, M., Akita, F., Hirata, K., Ueno, G., Murakami, H., Nakajima, Y., Shimizu, T., Yamashita, K., Yamamoto, M., Ago, H. & Shen, J. (2015). *Nature (London)*, **517**, 99–103.

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

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

Tegze, M. & Bortel, G. (2013). *J. Struct. Biol.* **183**, 389–393.

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

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

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.

Yabashi, M., Tanaka, H. & Ishikawa, T. (2015). *J. Synchrotron Rad.* **22**, 477–484.

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