

Experimental 3D coherent diffractive imaging from photon-sparse random projections

K. Giewekemeyer,^{a*} A. Aquila,^{a‡} N.-T. D. Loh,^{b,c,d} Y. Chushkin,^e K. S. Shanks,^f J.T. Weiss,^f M. W. Tate,^f H. T. Philipp,^f S. Stern,^{a,g} P. Vagovic,^{a,g} M. Mehrjoo,^a S. C. Teo,^{b,d} M. Barthelmess,^g F. Zontone,^e C. Chang,^h R. C. Tiberio,ⁱ A. Sakdinawat,^h G. J. Williams,^{h,q} S. M. Gruner,^{f,j,k} and A. P. Mancuso^{a,l}

Received 9 November 2018

Accepted 24 February 2019

Edited by T. Ishikawa, Harima Institute, Japan

‡ Present address: SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

§ Present address: Deutsches Elektronen-Synchrotron, Notkestraße 85, 22607 Hamburg, Germany.

¶ Present address: NSLS-II, Brookhaven National Laboratory, PO Box 5000, Upton, NY 11973, USA.

Keywords: coherent X-ray diffractive imaging (CXDI); X-ray free-electron lasers; XFELs; phase problem; single particles.

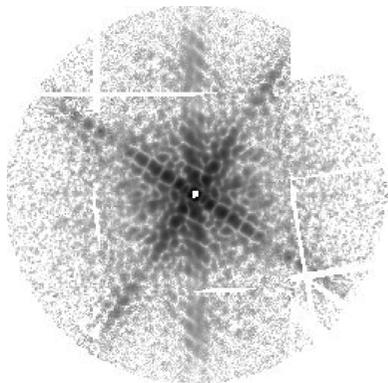
Supporting information: this article has supporting information at www.iucrj.org

^aEuropean XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany, ^bCentre for Bio-imaging Sciences, National University of Singapore, 14 Science Drive 4, 117557 Singapore, ^cDepartment of Physics, National University of Singapore, 2 Science Drive 3, 117551 Singapore, ^dDepartment of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117557 Singapore, ^eESRF – The European Synchrotron, 71 avenue des Martyrs, 38000 Grenoble, France, ^fLaboratory for Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA, ^gCenter for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, 22607 Hamburg, Germany, ^hSLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, ⁱStanford Nano Shared Facilities, Stanford University, 348 Via Pueblo, Stanford, CA 94305, USA, ^jCornell High Energy Synchrotron Source (CHESS), Cornell University, Ithaca, NY 14853, USA, ^kKavli Institute at Cornell for Nanoscale Science, Cornell University, Ithaca, NY 14853, USA, and ^lDepartment of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia. *Correspondence e-mail: klaus.giewekemeyer@xfel.eu

The routine atomic resolution structure determination of single particles is expected to have profound implications for probing structure–function relationships in systems ranging from energy-storage materials to biological molecules. Extremely bright ultrashort-pulse X-ray sources – X-ray free-electron lasers (XFELs) – provide X-rays that can be used to probe ensembles of nearly identical nanoscale particles. When combined with coherent diffractive imaging, these objects can be imaged; however, as the resolution of the images approaches the atomic scale, the measured data are increasingly difficult to obtain and, during an X-ray pulse, the number of photons incident on the 2D detector is much smaller than the number of pixels. This latter concern, the signal ‘sparsity’, materially impedes the application of the method. An experimental analog using a conventional X-ray source is demonstrated and yields signal levels comparable with those expected from single biomolecules illuminated by focused XFEL pulses. The analog experiment provides an invaluable cross check on the fidelity of the reconstructed data that is not available during XFEL experiments. Using these experimental data, it is established that a sparsity of order 1.3×10^{-3} photons per pixel per frame can be overcome, lending vital insight to the solution of the atomic resolution XFEL single-particle imaging problem by experimentally demonstrating 3D coherent diffractive imaging from photon-sparse random projections.

1. Introduction

A major motivation for advancing coherent diffractive imaging (CDI) using X-rays has always been its potential application to the imaging of individual nanoscale objects. A specific case concerns biological macromolecules, where the structure can be determined without the need for crystallization (Neutze *et al.*, 2000; Shenoy, 2003), representing an early potential application of X-ray free-electron lasers (XFELs). The high peak flux, of the order of 10^{12} photons per pulse, and the shorter than 100 fs-scale pulse duration are prerequisites for generating a diffraction signal from a single macromolecule, which is destroyed by the Coulomb explosion of the sample (Neutze *et al.*, 2000) resulting from ionization during the measurement. This scheme of ‘diffraction



before destruction' is routinely used for serial femtosecond crystallography, in which a randomly oriented crystal is illuminated by the X-ray beam (Chapman *et al.*, 2011; Schlichting, 2015). The crystal size has typically been large enough such that sufficient numbers of X-rays are diffracted into a recordable pattern prior to destruction to be able to determine the orientation of the crystal. This is repeated with new microcrystals until there is an adequate sampling of reciprocal space to determine the 3D structure of the unit cell of the crystal.

To obtain 3D structural information from single particles, such as virus particles or single proteins, serial diffraction data from many identical or nearly identical objects has to be measured with sufficient orientational variation. Usually, this is achieved by randomly injecting particles into the FEL beam, relying on the statistical coincidence of a single particle being hit by an FEL pulse (Barty, 2016; Spence, 2017). Due to the random nature of this process, the sample's orientation for each diffraction pattern is generally unknown and has to be recovered *a posteriori* in order to build up a continuous 3D diffraction volume in reciprocal space. This can then be inverted by iterative phase retrieval (Marchesini, 2007) into a real-space electron-density distribution, the last step of single-particle CDI.

A first complete demonstration of the method was provided by the 3D structure determination of the Giant Mimivirus, approximately 450 nm in diameter, to a resolution of 125 nm (Ekeberg *et al.*, 2015). More recently, the same type of experiment has been successfully performed using the much smaller Melbourne virus (diameter around 230 nm) to a resolution of 28 nm (Lundholm *et al.*, 2018). At the same time, a step towards even smaller viruses – the Rice Dwarf Virus (RDV) and bacteriophage PR772, both with a diameter of around 70 nm – has been made, resulting in images at a resolution of around 17 nm (Kurta *et al.*, 2017) and more recently, for PR772, slightly below 10 nm (Rose *et al.*, 2018).¹ Those data were a result of the Single-Particle-Imaging Initiative at the Linac Coherent Light Source (Aquila *et al.*, 2015; Munke *et al.*, 2016; Reddy *et al.*, 2017). This also resulted in the collection of a few hundred high-resolution diffraction frames from RDV at a photon energy of 7 keV, showing that a useful diffraction signal can be collected at 5.9 Å resolution from single hits (Munke *et al.*, 2016).

Reconstructing biological macromolecules to 3 Å resolution or better has previously been set as the ultimate goal of single-particle CDI (Aquila *et al.*, 2015). However, when the particle is sufficiently small, only a few hundred X-rays are scattered elastically prior to destruction of the particle (Yoon *et al.*, 2016; Fortmann-Grote *et al.*, 2017). In this case the recorded scattering pattern contains too few X-rays to be able to determine the orientation of the particle from the single pattern alone. Such patterns are called 'sparse', *i.e.* the

features necessary to interpret the pattern are totally dominated by Poisson noise. For example, a protein of at least *ca* 10 nm diameter is necessary to scatter, on average, 50 photons outside the central speckle, at a photon energy of 8 keV in a nanoscale FEL focus (see below for further details). In this case, hundreds of thousands of diffraction patterns have to be collected to build up the 3D reciprocal-space intensity, *i.e.* to assemble an invertible data set (Loh & Elser, 2009). To date, no such experimental data set exists and considerable method development is still required towards the realization of single-particle CDI as an independent method of macromolecular structure determination.

An important branch of this method development addresses the problem of orientation recovery in the case of very weak diffraction which is often not only sparse, but also contaminated by background signal, originating from sources such as the instrument or the particle beam. In recent years, several methods for orientation recovery have been devised (Huldt *et al.*, 2003; Loh & Elser, 2009; Fung *et al.*, 2009; Bortel & Tegze, 2011; Tegze & Bortel, 2012; Giannakis *et al.*, 2012; Yefanov & Vartanyants, 2013; Kassemeyer *et al.*, 2013; Zhou *et al.*, 2014; Donatelli *et al.*, 2015, 2017; Flamant *et al.*, 2016; Nakano *et al.*, 2017, 2018; von Ardenne *et al.*, 2018) and also applied to experimental single-particle FEL (Loh *et al.*, 2010; Kassemeyer *et al.*, 2013; Ekeberg *et al.*, 2015; Kurta *et al.*, 2017; Lundholm *et al.*, 2018; von Ardenne *et al.*, 2018) or similar data (Nakano *et al.*, 2017).

However, the important case of sparse diffraction from a 3D object has only been solved experimentally in a setting different from the classical CDI problem. For example, one of the methods of orientation recovery, a statistical technique based on expectation maximization, the expand–maximize–compress (EMC) algorithm (Loh & Elser, 2009; Ayyer *et al.*, 2016), has been applied successfully to real-space sparse radiographic data in two (Philipp *et al.*, 2012) and three dimensions (Ayyer *et al.*, 2014), to sparse crystallographic data limited to one (Wierman *et al.*, 2016) and two rotation axes (Lan *et al.*, 2017), and very recently also to synchrotron-based serial protein crystallographic data for random crystal orientations (Lan *et al.*, 2018).

Here, we demonstrate 3D CDI from sparse random projections in the same geometry as that used for FEL-based single-particle imaging experiments. Using a synchrotron beam on a micron-scale sample, we show that, with as few as about 50 scattered photons per diffraction pattern, and without explicit knowledge of the sample's orientation for a given data frame, it is possible to robustly reconstruct the scattering distribution of the sample in reciprocal space and to invert this diffraction volume into a high-resolution 3D electron density distribution. We show that, despite strong sparsity in the data, it is possible to reconstruct a sample to a complexity of more than 40 resolution elements within the largest diameter of the sample. For ease of comparison, the term 'resolution element' is defined by analogy with the work of Loh & Elser (2009) and Ayyer *et al.* (2016), *i.e.* two resolution elements equal the full-period (crystallographic) resolution. While in simulations it is often defined by the edge

¹ Unless otherwise noted, we define resolution as the crystallographic (full-period) resolution, in contrast with the half-period resolution which is also often used in the literature on CDI. For comparability, resolution values obtained according to an analysis of the phase retrieval transfer function are given here.

of the detector, we define it here in terms of the resolution as obtained after phasing.

2. Experimental

The experiment was performed on the undulator beamline ID10, end station EH2, of the European Synchrotron Radiation Facility (ESRF) (Chushkin *et al.*, 2014). The photon energy was set to 8.1 keV using a water-cooled Si(111) pseudo-channel-cut monochromator with an intrinsic energy resolution of $\Delta E/E \simeq 1.4 \times 10^{-4}$. The sample was a solid gold object with a largest diagonal length of about $1.1 \mu\text{m}$, fabricated by electroplating and supported by a silicon nitride membrane. It was placed in the beam on a high-precision tomographic stage at a distance of 4.0 m from the detector. Defined by several sets of slits (Chushkin *et al.*, 2014), the lateral beam size at the sample plane was approximately $10 \times 10 \mu\text{m}$. A schematic of the setup is shown in Fig. 1. For further details of all the major aspects of our data treatment and the algorithms employed in the reciprocal- and real-space reconstruction of our data, we refer the reader to the supporting information.

For data collection, the mixed-mode pixel array detector (MM-PAD) was used. It is a wide dynamic range integrating detector developed at Cornell University that is capable of

collecting data at a kilohertz frame rate with a high signal-to-noise ratio (SNR), ranging from one X-ray photon per pixel to a maximum rate exceeding 10^8 photons per pixel per second at 8 keV photon energy (Tate *et al.*, 2013). The beam was attenuated using a polished single-crystal Si attenuator with a transmission of *ca* 0.3. As a result, the overall flux reaching the detector was approximately 10^8 photons s^{-1} .

The data set analyzed here comprises diffraction patterns from 227 unique 3D orientations. With 2000 collected frames per orientation, this amounts to a total of $M_{\text{data}} = 454\,000$ data frames. Each frame was collected with an illumination time of 25 ms. The orientations were obtained from two independent tomographic series (see Fig. 1), during each of which only θ , the angle about the tomographic rotation axis y , was varied.

3. Analysis

3.1. Detector calibration and further data treatment

As the collected data were very sparse, reliable identification of single- and few-photon events becomes crucial for the analysis which explicitly takes into account the Poissonian nature of the noise in the data (see below). Calibration of the detector data, *i.e.* the transformation of the raw detector

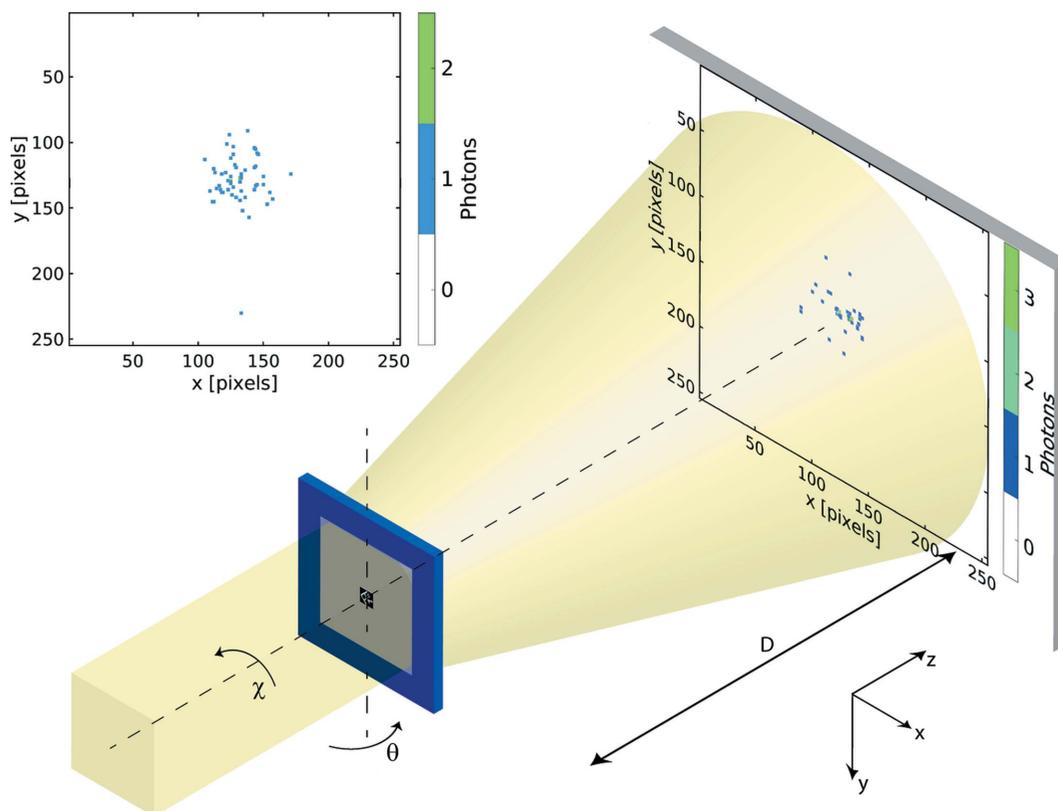


Figure 1

A schematic diagram of the experiment. The sample, a gold nanostructure supported on a silicon nitride membrane, was rotated about the y axis by an angle θ to obtain diffraction patterns at different orientations with respect to the optical axis z . A first rotation series about the y axis was followed by an in-plane rotation of the sample about z (angle χ) and a subsequent second rotation series about y . The beam attenuation and illumination time were adjusted so that each data frame contains only about 50 scattered photons. An example of a single diffraction pattern is shown in the inset on the upper left-hand side (see also animation in the supporting information). The area of a single pixel in comparison with the field-of-view area has been enlarged for visualization.

output into photon counts, is therefore an essential step of the analysis.

The applied calibration procedure consisted of several steps. A binary mask was used to reject all pixels that were inactive or were within gaps between detector modules. Second, the dark signal was subtracted for each pixel. The gain of the detector was determined from the histogram of pixel values over many frames within a region of interest where the maximum intensity per pixel is only a few photons per frame. The resulting histogram shows discrete peaks corresponding to zero, one, two *etc.* photons per pixel. In terms of analog-to-digital units (ADU), a gain of 11.1 ADU photon⁻¹ at 8.1 keV was determined from the spacing of these peaks. This gain, together with the widths of these peaks, yields an SNR of 5.2 for single photons. For the transformation of ADUs into single photons, a threshold energy of $E_t = E_\gamma - \text{HWHM}_n = 0.77E_\gamma$ was used, with the single-photon energy E_γ and the noise peak half-width at half-maximum HWHM_n .

With around 40 000 active detector pixels in a region of interest (ROI) of 255×255 pixels around the beam center, this leads to a false-positive probability $P(1|0) \simeq 3 \times 10^{-5}$ or between 1 and 2 events per frame. At the sparsity level of the data and with more than 400 000 frames comprising the complete data set, it was found that cosmic radiation made up a considerable contribution to the total recorded intensity. We therefore applied a threshold-based removal procedure. The sparsity level of the single frames is reflected by an average of 49.3 scattered photons detected per frame (median 48 and standard deviation 9.6), after masking out pixels dominated by empty-beam scattering. Due to the low level of background scattering from the instrument, *e.g.* optical components or apertures, no background or ‘empty-beam’ subtraction was performed on the data.

3.2. Orientation determination

The goal of orientation determination is to obtain the 3D reciprocal-space intensity $W(\mathbf{q})$ that is proportional to the modulus-squared Fourier transform of the 3D electron density of the sample. Here, \mathbf{q} denotes the 3D Cartesian reciprocal-space coordinate. To reconstruct $W(\mathbf{q})$, the EMC algorithm correlates each data frame, K_d ($d = 1, \dots, M_{\text{data}}$), with tomographic slices W_j ($j = 1, \dots, M_{\text{rot}}$) of W , corresponding to M_{rot} possible sample orientations, based on the current iterate of our model of $W(\mathbf{q})$. Each iteration comprises expanding the current model W into slices W_j , an update $W_j \rightarrow W'_j$ by maximizing a log-likelihood function $Q(W')$, and compressing slices W'_j into a new 3D model $W'(\mathbf{q})$. The update itself consists of forming the weighted sum

$$W'_{ij} = \frac{\sum_{d=1}^{M_{\text{data}}} P_{jd}(W) K_{id}}{\sum_{d=1}^{M_{\text{data}}} P_{jd}(W)}. \quad (1)$$

The index i specifies a pixel and $P_{jd}(W)$ is the probability of a frame K_d having been collected at an orientation j , based on the slice W_j . For very noisy diffraction data from weak scatterers this probability can be well described using Poisson statistics (Loh & Elser, 2009).

Each orientation j of the sample is represented by a unit quaternion, q_j . For the present experiment, the set $\{q_j\}$ of possible orientations was derived from the experimental setup and procedure (for details, see supporting information). As a result, the optimized set $\{q_j\}$ of orientations could be used to generate a Fourier intensity $W_{\text{ref}}(\mathbf{q})$, assembled using full knowledge of the orientations, as a reference for the EMC-based intensity reconstruction. EMC received the same set of orientations as an input, together with all data frames, but without explicit knowledge of any frame’s orientation. As a further input, a geometry file was included that contained the reciprocal-space coordinate of each detector pixel and a 3D binary mask, S , identifying those voxels in the cubic domain of $W(\mathbf{q})$ that are to be excluded from the analysis process, *e.g.* because they are never reached by any Ewald sphere slice. As S is non-symmetric with respect to $\mathbf{q} = 0$, the Friedel symmetrization step included in EMC (Loh & Elser, 2009) was modified accordingly. To account for the fact that some pixels in each frame K_d contain a large fraction of sample scattering, but still have a significant contribution of ‘parasitic’ or beamline scatter, we defined a binary mask on the detector ROI to identify those pixels to be included in the update rule $W \rightarrow W'$, but not into the calculation of probabilities $P_{jd}(W)$ (Ayyer *et al.*, 2016).

As EMC was always initiated with a random intensity distribution, independent runs of the algorithm show some statistical variation. To reduce the associated uncertainty, the EMC algorithm was run 20 times for 500 iterations, followed by an averaging procedure similar to that described by Yoon *et al.* (2016). A small fraction of the ensemble, 2 out of 20 reconstructions, exhibited artifacts due to localized over-weighting of certain orientations. These could be automatically discarded by rejecting highly non-homogeneous distributions of orientations. Within the remaining results, two main classes could be observed which are related by an overall rotation of about 180° around an axis close to one of the coordinate axes. This is in accordance with a previous study for an isotropic orientational distribution (Yoon *et al.*, 2016). After manual attribution to one of the two classes, the results were averaged and their relation was verified by orientational registration. The final averaged 3D reciprocal-space volume $\langle W(\mathbf{q}) \rangle$ was obtained as an average of 13 individual results in the same orientation as $W_{\text{ref}}(\mathbf{q})$.

4. Results

4.1. Reciprocal space (intensity)

A comparison of $\langle W(\mathbf{q}) \rangle$ with W_{ref} is shown in Fig. 2. Visually, the orthogonal slices through the reconstructed and reference intensities are in very good agreement. This observation is reflected by an overall Pearson correlation coefficient $r = C(W_{\text{ref}}(\mathbf{q}), \langle W(\mathbf{q}) \rangle)$ of 97.1% for $\mathbf{q} \in S$ and $r = 99.0\%$ for $20 \leq q = |\mathbf{q}| \leq 50$ [see the solid lines in Figs. 2(c) and 2(f)].

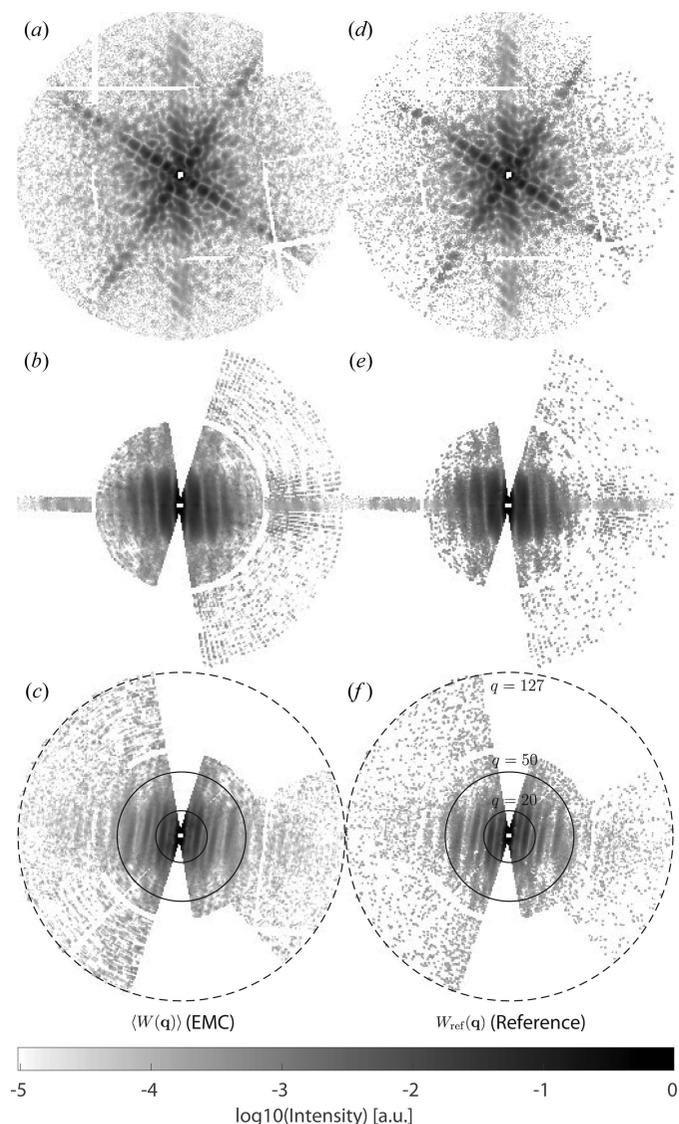


Figure 2 Orthogonal slices through (a)–(c) the EMC-reconstructed and (d)–(f) the reference 3D diffraction volumes. The EMC-reconstructed diffraction volume $W(\mathbf{q})$ results from averaging the results of 13 independent EMC runs, each starting with a random intensity distribution. The reference diffraction volume $W_{\text{ref}}(\mathbf{q})$ was constructed based on the known orientations of the sample for each frame during the measurement. The dashed circles in panels (c) and (f) indicate a radius of 127 voxels, whereas the solid circles indicate radii of 20 and 50, respectively. All slices are drawn on the same scale with dimensionless lateral coordinates in units of $\Delta Xk/D$.

4.2. Validation (intensity reconstruction)

To assess the reliability of the reconstructed 3D intensity in reciprocal space, we randomly assigned the frames of the data set to two independent half-data sets and reconstructed two independent 3D reciprocal-space volumes as described before. A Fourier shell correlation (FSC) curve (van Heel & Schatz, 2005), obtained here directly from the two reciprocal-space volumes, is shown in Fig. 3 (red line). It intersects the half-bit threshold curve (van Heel & Schatz, 2005) at a value beyond $q = 120$, indicating a self-consistent reconstruction of the reciprocal-space volume close to the Nyquist limit. Here, q is

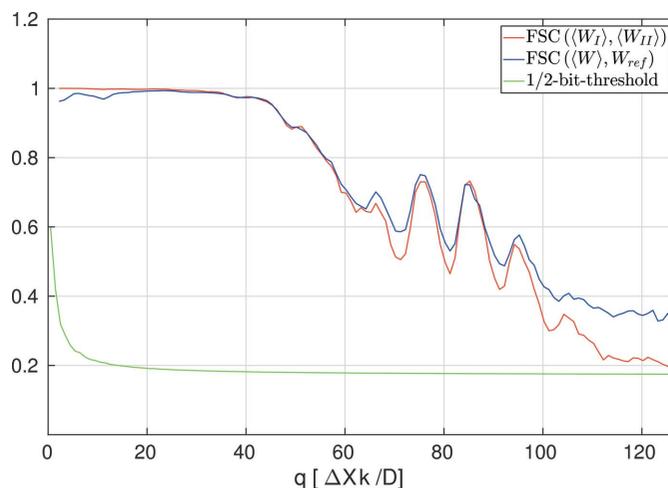


Figure 3 The red line illustrates the Fourier shell correlation (FSC) between two EMC-retrieved reciprocal-space volumes resulting from splitting the data set into two equal halves and performing the same analysis on them as on the whole data set. The blue line indicates the FSC between $W(\mathbf{q})$ and W_{ref} , *i.e.* the reciprocal-space volume resulting from analyzing the whole data set using EMC and the reference intensity assembled using known orientations. The green line denotes the half-bit threshold curve, used as a common criterion for resolution determination in analysis of FSC curves.

measured in units of $\Delta Xk/D$, where $\Delta X = 150 \mu\text{m}$ is the detector pixel pitch, k is the wave number and D is the sample-to-detector distance. (Note that, in the limit of a flat Ewald sphere patch, which is well fulfilled here, q then corresponds to the distance to the central beam in units of detector pixels. For further details, see the supporting information.) This analysis shows the effect of EMC alone rather than merging the effects of orientation determination and phasing, as would result for a traditional FSC analysis on phased results in real space. For comparison, the FSC curve resulting from correlating $W(\mathbf{q})$ with W_{ref} , as obtained from known orientations, is also shown (blue line). The high degree of similarity between these two curves indicates strong agreement between the result obtained by EMC and the reference intensity distribution, assembled using full knowledge of the orientations.

4.3. Real space (density)

To obtain the real-space electron-density distribution, the missing phases of the 3D Fourier intensity need to be determined. To this end, we applied standard iterative phase retrieval to the 3D diffraction data, *i.e.* a combination of the hybrid input–output (HIO) algorithm and the error reduction (ER) algorithm (Fienup & Wackerman, 1986; Marchesini *et al.*, 2003; Xiong *et al.*, 2014). In total, 600 iterations were applied, *i.e.* 420 iterations of HIO with a feedback parameter $\beta = 0.9$, followed by 180 iterations of ER. For further details see the supporting information.

To ensure the reproducibility of the obtained result, 60 reconstructions were performed in total. The results were filtered in a two-step selection process. In the first, manual, step, we discarded images which visually deviated from the most abundant reconstruction result. In the second step, the

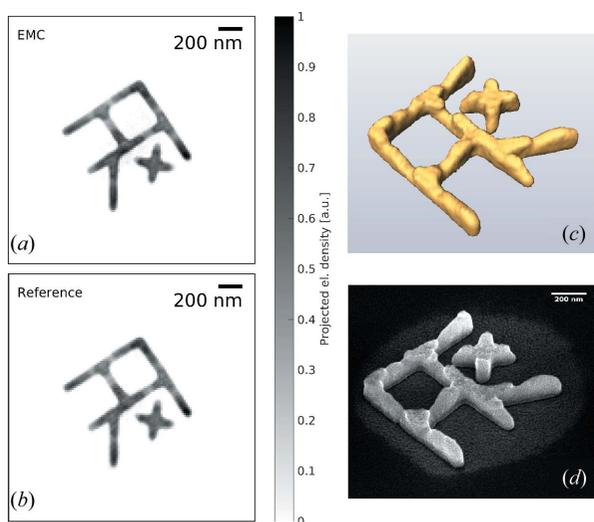


Figure 4 Reconstruction of the 3D electron density. (a) Reconstruction from the result derived by EMC. The electron density projected along an axis perpendicular to the drawing plane is shown here. (b) Reconstruction from the reference Fourier volume. Again, the projected electron density is shown. (c) 3D iso-surface rendering of the reconstructed electron density shown in panel (a). The threshold of the iso-surface has been set to 0.2, given a normalized density with values between 0 and 1 (see also animation in the supporting information). (d) Scanning electron micrograph from the original sample.

remaining reconstructions (43 for the reference and 37 for EMC data set) were aligned with sub-pixel precision (Guizar-Sicairos *et al.*, 2008) and averaged. Then the 20 reconstructions showing the highest correlation with this average were selected for the final average. Note that the procedure applied in the second step could also be used to avoid any manual intervention. However, in such a case, several iterations would likely be required in order to avoid bias by strong outliers in the average. The resulting real-space reconstructions, *i.e.* the real part of the final average, from both reciprocal-space intensities are shown in Fig. 4.

All details of the reference reconstruction are reproduced in the EMC-based reconstruction down to a resolution level of very few pixels. A comparison of a scanning electron microscopy image of the sample with an iso-surface rendering of the EMC-based reconstruction shows that height variations due to imperfections in the fabrication process are well reproduced by the reconstruction. This identifies the sample as a true 3D structure with features in all coordinate directions.

4.4. Validation (density reconstruction)

The resolution of the final image was estimated via the phase retrieval transfer function (PRTF) according to a procedure similar to the one described by Chapman *et al.* (2006). More specifically, before summation of the complex-valued reconstructions, their constant phases were adjusted so that the real part of each reconstruction was maximized. The PRTF curves for the results shown in Fig. 4 are shown in Fig. 5. The full-period resolution, as determined by the spatial frequency corresponding to a PRTF value of $1/e$, amounts to a

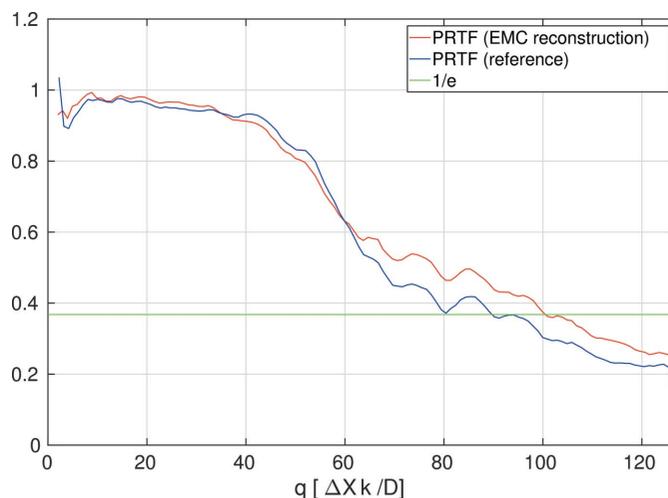


Figure 5 Phase retrieval transfer functions for the reconstructions from the EMC-generated Fourier space intensity ($W(\mathbf{q})$) and the reference intensity W_{ref} . The curves decay to a value of $1/e$ between $q = 90$ and $q = 100$, corresponding to a half-period resolution between 20 and 23 nm.

value between 40 and 45 nm. This corresponds to 24 to 26 (full-period) resolution elements within the largest linear extension of the particle, as given by the smallest sphere completely containing the particle (diameter $\simeq 1.1 \mu\text{m}$).

5. Discussion

5.1. Significance of the observed level of sparsity

To assess the significance of the sparsity level for a single detector frame in the present experiment, we have calculated the expected average total number of scattered photons outside the central speckle for a selection of 35 000 human protein structures from the RCSB Protein Data Bank (PDB; <http://www.rcsb.org>) (Berman *et al.*, 2000). Here, a focal-spot diameter of 300 nm was assumed, at a photon energy of 8 keV and a pulse energy of 1 mJ (with 20% beamline transmission). As a result, it could be shown that, under these realistic conditions, the minimum diameter for a protein to scatter 50 photons outside the central speckle amounts to 10.6 nm. For further details, see the supporting information.

This clearly shows that the signal level in the present experiment, obtained at a synchrotron source from a nano-fabricated gold structure, is comparable with what can be expected under realistic conditions from a relevant protein structure at an FEL source.

5.2. Particle complexity, rotation group sampling, SNR and sparsity

Another parameter to be discussed is the particle complexity R , as measured in half-resolution units per particle radius (Loh & Elser, 2009). Reconstructing a particle with 10 nm diameter down to a resolution of 3 Å results in a complexity of $R \simeq 33$, far beyond the current state of the art for FEL-based SPI (Lundholm *et al.*, 2018) which is $R \simeq 8$ for a globular virus particle. The present structure reaches a

complexity >20 in two dimensions, being constrained in the height direction to a value between 2 and 3. A comparison between a scanning electron microscopy (SEM) image and an isosurface rendering of the reconstructed particle density shows that all features in the height direction are very well reproduced. Despite its flat shape, this clearly identifies the particle as a true 3D structure and underlines the significance of the present result as a step forward towards the complexity level required for real protein structures.

Furthermore, compared with a serial imaging experiment at an FEL, where thousands of particles in random 3D orientations contribute to a full data set, the number of unique 3D orientations contributing to the present data set seems relatively low ($M_{\text{rot}} = 227$). It can be shown, however, that at the given resolution and complexity this does not restrict the relevance of the result. The required minimum angular separation between adjacent orientations for a sufficient sampling of the 3D rotation group is linked to the complexity R of the particle (Loh *et al.*, 2010): $\delta\theta = 1/R$. In case of a non-globular shape, the maximum complexity in a given coordinate direction allows for a conservative estimate, leading here to $\delta\theta = 1/R_{\text{max}} = 2.2^\circ$. This shows that a finer sampling for the tomographic series contributing to the present data set would not have added more information, at the obtained resolution.

Evidently, the present experiment profits from a high SNR (see the supporting information) which would have been impossible without a setup well optimized for forward-scattering CDI (Chushkin *et al.*, 2014). Most importantly, this consists of a set of accurately placed apertures upstream of the sample which are adjusted to define the beam incident on the sample and, at the same time, to suppress scattering arising from upstream apertures by those further downstream. Similar schemes can be applied at FEL sources to make them compatible with CDI experiments (Munke *et al.*, 2016). Even though in the FEL case the aerosol jet in which sample particles are injected through a stream of carrier gas causes an additional source of background scatter (Daurer *et al.*, 2017), the present data set gives an experimental benchmark for a signal-to-noise level which would likely allow FEL-based single-particle imaging.

To provide some understanding of how orientation information can be retrieved despite the sparsity of photon measurements per pattern, consider the two key attributes of a scheme that succeeds in this retrieval. First, this scheme must recognize when it is presented with the correct 3D Fourier intensity distribution, and second, it must contain instructions to improve the compatibility between a candidate intensity distribution and the set of measured diffraction patterns. To satisfy the first attribute of recognizability, the EMC algorithm has to model the detection noise and the randomness of the orientations as accurately as possible. In the present case, this is achieved by a Poissonian noise model and an equal weight for each possible orientation. If a compatible 3D intensity distribution is recognized, then the most likely orientations of each photon pattern within this distribution can be estimated naturally. To fulfill the second attribute, EMC uses this noise model to modify any candidate 3D intensity distribution to

increase its likelihood of generating the set of measured diffraction patterns. A derivation of the corresponding maximization rule is given by Loh & Elser (2009). If there are a sufficient number of photons per pattern, rotation group sampling, signal-to-background ratio and total number of measurements, then EMC can seek and recognize a family of compatible 3D intensity distributions even if the initial guess were completely random (Loh & Elser, 2009). A more quantitative definition of sufficiency for this orientation problem, beyond what is already in the literature (*e.g.* Loh & Elser, 2009; Elser, 2009; Philipp *et al.*, 2012; Ayyer *et al.*, 2014; Loh, 2014), is the subject of a future publication by some of the co-authors here.

6. Summary and conclusions

In summary, we have demonstrated experimentally, in the same geometry as used for FEL-based single-particle imaging, the reconstruction of a complex three-dimensional object using coherent diffractive imaging (CDI) from photon-sparse random projections at a sparsity level to be expected for a typical protein at an FEL source. To this end, we have collected 454 000 data frames with about 50 scattered photons per frame, evenly distributed over 227 unique orientations, and reconstructed a consistent 3D reciprocal-space volume without explicit knowledge of the orientation of a given frame.

It has been shown that, by application of the expand–maximize–compress (EMC) algorithm, both the reconstructed reciprocal-space intensity and the real-space density of the sample agree to a high level with reconstructions obtained using complete knowledge of the frame orientations.

The data set is made freely available in the CXI data bank (Maia, 2012; <http://cxidb.org/>), entry 84, to be used as a testbed for algorithm development for CDI-based single-particle imaging, *e.g.* by alternative methods of orientation determination. In addition, the data set can serve as a target for a signal-to-noise level enabling FEL-based single-particle imaging in the future.

Acknowledgements

Jerome Kieffer and Pierre Paleo are acknowledged for GPU implementation of the phase iterative algorithm. The authors are grateful to Pascal Dideron, Pascal Voisin and Laurent Claustre for technical support during the experiment.

Funding information

Detector research at Cornell is supported by US Department of Energy awards DE-SC0004079, DE-SC0016035 and DE-SC0017631 and by the Cornell High Energy Synchrotron Source (CHESS), supported by the US National Science Foundation (DMR-1332208). This research used resources of the National Synchrotron Light Source II, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The experi-

ments were performed on beamline ID10 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. The work was done partially while K. Giewekemeyer was visiting the Institute for Mathematical Sciences, National University of Singapore, in 2018. The visit was supported by the Institute. Duane Loh and Colin Teo would like to acknowledge funding support from the Singapore National Research Foundation's Competitive Research Program funding (NRF-CRP16-2015-05). This research used resources from the DOE Early Career Program and Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number DE-AC02-76SF00515. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152.

References

- Aquila, A., Barty, A., Bostedt, C., Boutet, S., Carini, G., dePonte, D., Drell, P., Doniach, S., Downing, K. H., Earnest, T., Elmlund, H., Elser, V., Gühr, M., Hajdu, J., Hastings, J., Hau-Riege, S. P., Huang, Z., Lattman, E. E., Maia, F. R. N. C., Marchesini, S., Ourmazd, A., Pellegrini, C., Santra, R., Schlichting, I., Schroer, C., Spence, J. C. H., Vartanyants, I. A., Wakatsuki, S., Weis, W. I. & Williams, G. J. (2015). *Struct. Dyn.* **2**, 041701.
- Ardenne, B. von, Mechelke, M. & Grubmüller, H. (2018). *Nat. Commun.* **9**, 2375.
- Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. (2016). *J. Appl. Cryst.* **49**, 1320–1335.
- Ayyer, K., Philipp, H. T., Tate, M. W., Elser, V. & Gruner, S. M. (2014). *Opt. Express*, **22**, 2403–2413.
- Barty, A. (2016). *Curr. Opin. Struct. Biol.* **40**, 186–194.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Bortel, G. & Tegze, M. (2011). *Acta Cryst.* **A67**, 533–543.
- Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006). *J. Opt. Soc. Am. A*, **23**, 1179.
- 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.-U., 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.-D., 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*, **470**, 73–77.
- Chushkin, Y., Zontone, F., Lima, E., De Caro, L., Guardia, P., Manna, L. & Giannini, C. (2014). *J. Synchrotron Rad.* **21**, 594–599.
- Daurer, B. J., Okamoto, K., Bielecki, J., Maia, F. R. N. C., Mühlig, K., Seibert, M. M., Hantke, M. F., Nettelblad, C., Benner, W. H., Svenda, M., Timneanu, N., Ekeberg, T., Loh, N. D., Pietrini, A., Zani, A., Rath, A. D., Westphal, D., Kirian, R. A., Awel, S., Wiedorn, M. O., van der Schot, G., Carlsson, G. H., Hasse, D., Sellberg, J. A., Barty, A., Andreasson, J., Boutet, S., Williams, G., Koglin, J., Andersson, I., Hajdu, J. & Larsson, D. S. D. (2017). *IUCrJ*, **4**, 251–262.
- Donatelli, J. J., Sethian, J. A. & Zwart, P. H. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 7222–7227.
- Donatelli, J. J., Zwart, P. H. & Sethian, J. A. (2015). *Proc. Natl Acad. Sci. USA*, **112**, 10286–10291.
- Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R., 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., Martin, A. V., Andersson, I., Loh, N. D., Martin, A. V., Chapman, H., Bostedt, C., Bozek, J. D., Ferguson, K. R., Krzywinski, J., Epp, S. W., Rolles, D., Rudenko, A., Hartmann, R., Kimmel, N. & Hajdu, J. (2015). *Phys. Rev. Lett.* **114**, 098102.
- Elser, V. (2009). *IEEE Trans. Inf. Theory*, **55**, 4715–4722.
- Fienup, J. R. & Wackerman, C. C. (1986). *J. Opt. Soc. Am. A*, **3**, 1897–1907.
- Flamant, J., Le Bihan, N., Martin, A. V. & Manton, J. H. (2016). *Phys. Rev. E*, **93**, 053302.
- Fortmann-Grote, C., Buzmakov, A., Jurek, Z., Loh, N.-T. D., Samoylova, L., Santra, R., Schneidmiller, E. A., Tschentscher, T., Yakubov, S., Yoon, C. H., Yurkov, M. V., Ziaja-Motyka, B. & Mancuso, A. P. (2017). *IUCrJ*, **4**, 560–568.
- Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. (2009). *Nat. Phys.* **5**, 64–67.
- Giannakis, D., Schwander, P. & Ourmazd, A. (2012). *Opt. Express*, **20**, 12799–12826.
- Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. (2008). *Opt. Lett.* **33**, 156–158.
- Heel, M. van & Schatz, M. (2005). *J. Struct. Biol.* **151**, 250–262.
- Huldt, G., Szóke, A. & Hajdu, J. (2003). *J. Struct. Biol.* **144**, 219–227.
- Kassemeyer, S., Jafarpour, A., Lomb, L., Steinbrener, J., Martin, A. V. & Schlichting, I. (2013). *Phys. Rev. E*, **88**, 042710.
- Kurta, R. P., Donatelli, J. J., Yoon, C. H., Berntsen, P., Bielecki, J., Daurer, B. J., DeMirici, H., Fromme, P., Hantke, M. F., Maia, F. R., Munke, A., Nettelblad, C., Pande, K., Reddy, H. K., Sellberg, J. A., Sierra, R. G., Svenda, M., van der Schot, G., Vartanyants, I. A., Williams, G. J., Xavier, P. L., Aquila, A., Zwart, P. H. & Mancuso, A. P. (2017). *Phys. Rev. Lett.* **119**, 158102.
- Lan, T.-Y., Wierman, J. L., Tate, M. W., Philipp, H. T., Elser, V. & Gruner, S. M. (2017). *J. Appl. Cryst.* **50**, 985–993.
- Lan, T.-Y., Wierman, J. L., Tate, M. W., Philipp, H. T., Martin-Garcia, J. M., Zhu, L., Kissick, D., Fromme, P., Fischetti, R. F., Liu, W., Elser, V. & Gruner, S. M. (2018). *IUCrJ*, **5**, 548–558.
- Loh, N. D. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130328.
- Loh, N. D., Bogan, M. J., Elser, V., Barty, A., Boutet, S., Bajt, S., Hajdu, J., Ekeberg, T., Maia, F. R. N. C., Schulz, J., Seibert, M. M., Iwan, B., Timneanu, N., Marchesini, S., Schlichting, I., Shoeman, R. L., Lomb, L., Frank, M., Liang, M. & Chapman, H. N. (2010). *Phys. Rev. Lett.* **104**, 225501.
- Loh, N. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.
- Lundholm, I. V., Sellberg, J. A., Ekeberg, T., Hantke, M. F., Okamoto, K., van der Schot, G., Andreasson, J., Barty, A., Bielecki, J., Bruza, P., Bucher, M., Carron, S., Daurer, B. J., Ferguson, K., Hasse, D., Krzywinski, J., Larsson, D. S. D., Morgan, A., Mühlig, K., Müller, M., Nettelblad, C., Pietrini, A., Reddy, H. K. N., Rupp, D., Sauppe, M., Seibert, M., Svenda, M., Swiggers, M., Timneanu, N., Ulmer, A., Westphal, D., Williams, G., Zani, A., Faigel, G., Chapman, H. N., Möller, T., Bostedt, C., Hajdu, J., Gorkhover, T. & Maia, F. R. N. C. (2018). *IUCrJ*, **5**, 531–541.
- Maia, F. R. N. C. (2012). *Nat. Methods*, **9**, 854–855.

- Marchesini, S. (2007). *Rev. Sci. Instrum.* **78**, 011301.
- Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., Weierstall, U. & Spence, J. C. H. (2003). *Phys. Rev. B*, **68**, 140101.
- Munke, A., Andreasson, J., Aquila, A., Awel, S., Ayer, K., Barty, A., Bean, R. J., Berntsen, P., Bielecki, J., Boutet, S., Bucher, M., Chapman, H. N., Daurer, B. J., DeMirici, H., Elser, V., Fromme, P., Hajdu, J., Hantke, M. F., Higashiura, A., Hogue, B. G., Hosseinizadeh, A., Kim, Y., Kirian, R. A., Reddy, H. K., Lan, T.-Y., Larsson, D. S., Liu, H., Loh, N. D., Maia, F. R., Mancuso, A. P., Mühlig, K., Nakagawa, A., Nam, D., Nelson, G., Nettelblad, C., Okamoto, K., Ourmazd, A., Rose, M., van der Schot, G., Schwander, P., Seibert, M. M., Sellberg, J. A., Sierra, R. G., Song, C., Svenda, M., Timneanu, N., Vartanyants, I. A., Westphal, D., Wiedorn, M. O., Williams, G. J., Xavier, P. L., Yoon, C. H. & Zook, J. (2016). *Sci. Data*, **3**, 160064.
- Nakano, M., Miyashita, O., Jonic, S., Song, C., Nam, D., Joti, Y. & Tama, F. (2017). *J. Synchrotron Rad.* **24**, 727–737.
- Nakano, M., Miyashita, O., Jonic, S., Tokuhisa, A. & Tama, F. (2018). *J. Synchrotron Rad.* **25**, 1010–1021.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
- Philipp, H. T., Ayer, K., Tate, M. W., Elser, V. & Gruner, S. M. (2012). *Opt. Express*, **20**, 13129–13137.
- Reddy, H. K. N., Yoon, C. H., Aquila, A., Awel, S., Ayer, K., Barty, A., Berntsen, P., Bielecki, J., Bobkov, S., Bucher, M., Carini, G. A., Carron, S., Chapman, H., Daurer, B., DeMirici, 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.
- Rose, M., Bobkov, S., Ayer, K., Kurta, R. P., Dzhigaev, D., Kim, Y. Y., Morgan, A. J., Yoon, C. H., Westphal, D., Bielecki, J., Sellberg, J. A., Williams, G., Maia, F. R. N. C., Yefanov, O. M., Ilyin, V., Mancuso, A. P., Chapman, H. N., Hogue, B. G., Aquila, A., Barty, A. & Vartanyants, I. A. (2018). *IUCrJ*, **5**, 727–736.
- Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.
- Shenoy, G. (2003). Technical Report SLAC-R-611. SLAC National Accelerator Laboratory, Menlo Park, California, USA.
- Spence, J. C. H. (2017). *IUCrJ*, **4**, 322–339.
- Tate, M. W., Chamberlain, D., Green, K. S., Philipp, H. T., Purohit, P., Strohman, C. & Gruner, S. M. (2013). *J. Phys. Conf. Ser.* **425**, 062004.
- Tegze, M. & Bortel, G. (2012). *J. Struct. Biol.* **179**, 41–45.
- Wierman, J. L., Lan, T.-Y., Tate, M. W., Philipp, H. T., Elser, V. & Gruner, S. M. (2016). *IUCrJ*, **3**, 43–50.
- Xiong, G., Moutanabbir, O., Reiche, M., Harder, R. & Robinson, I. (2014). *Adv. Mater.* **26**, 7747–7763.
- Yefanov, O. M. & Vartanyants, I. A. (2013). *J. Phys. B At. Mol. Opt. Phys.* **46**, 164013.
- Yoon, C. H., Yurkov, M. V., Schneidmiller, E. A., Samoylova, L., Buzmakov, A., Jurek, Z., Ziaja, B., Santra, R., Loh, N. D., Tschentscher, T. & Mancuso, A. P. (2016). *Sci. Rep.* **6**, 24791.
- Zhou, L., Zhang, T.-Y., Liu, Z.-C., Liu, P. & Dong, Y.-H. (2014). *Acta Cryst. A* **70**, 364–372.