



Received 17 June 2020

Accepted 21 September 2020

Edited by T. Ishikawa, Harima Institute, Japan

Keywords: single-particle imaging; three-dimensional virus reconstruction; bacteriophage PR772; XFELs.

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

An advanced workflow for single-particle imaging with the limited data at an X-ray free-electron laser

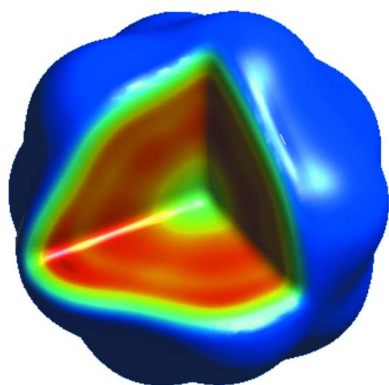
Dameli Assalauova,^a Young Yong Kim,^a Sergey Bobkov,^b Ruslan Khubbutdinov,^{a,c} Max Rose,^a Roberto Alvarez,^{d,e} Jakob Andreasson,^f Eugeniu Balaur,^g Alice Contreras,^{h,i} Hasan DeMirici,^{j,k} Luca Gelisio,^l Janos Hajdu,^{f,m} Mark S. Hunter,ⁿ Ruslan P. Kurta,^o Haoyuan Li,^{p,n} Matthew McFadden,ⁱ Reza Nazari,^{d,q} Peter Schwander,^r Anton Teslyuk,^{b,s} Peter Walter,ⁿ P. Lourdu Xavier,^{l,n,t} Chun Hong Yoon,ⁿ Sahba Zaare,^{d,n} Viacheslav A. Ilyin,^{b,s} Richard A. Kirian,^d Brenda G. Hogue,^{h,i,u} Andrew Aquila^{n,*} and Ivan A. Vartanyants^{a,c,*}

^aDeutsches Elektronen-Synchrotron DESY, Notkestraße 85, Hamburg, D-22607, Germany, ^bNational Research Center 'Kurchatov Institute', Akademika Kurchatova pl. 1, Moscow, 123182 Russian Federation, ^cNational Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe sh. 31, Moscow, 115409, Russian Federation, ^dDepartment of Physics, Arizona State University, Tempe, Arizona AZ 85287, USA, ^eSchool of Mathematics and Statistical Sciences, Arizona State University, Tempe, Arizona AZ 85287, USA, ^fInstitute of Physics, ELI Beamlines, Academy of Sciences of the Czech Republic, Prague, CZ-18221, Czech Republic, ^gAustralian Research Council Centre of Excellence in Advanced Molecular Imaging, Department of Chemistry and Physics, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Melbourne, Victoria 3086, Australia, ^hSchool of Life Sciences, Arizona State University, Tempe, Arizona AZ 85287, USA, ⁱBiodesign Institute Center for Immunotherapy, Vaccines and Virotherapy, Arizona State University, Tempe, Arizona AZ 85287, USA, ^jStanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, ^kDepartment of Molecular Biology and Genetics, Koc University, Istanbul, 34450, Turkey, ^lCenter for Free Electron Laser Science (CFEL), DESY, Notkestraße 85, Hamburg, D-22607, Germany, ^mLaboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Uppsala, SE-75124, Sweden, ⁿSLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA, ^oEuropean XFEL, Holzkoppel 4, Schenefeld, D-22869, Germany, ^pPhysics Department, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305-2004, USA, ^qSchool of Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ 85287, USA, ^rUniversity of Wisconsin Milwaukee, Milwaukee WI 53211, USA, ^sMoscow Institute of Physics and Technology, Moscow, 141700, Russian Federation, ^tMax-Planck Institute for the Structure and Dynamics of Matter, Luruper Chaussee 149, Hamburg, D-22761, Germany, and ^uBiodesign Institute, Center for Applied Structural Discovery, Arizona State University, Tempe, AZ 85287, USA. *Correspondence e-mail: aquila@slac.stanford.edu, ivan.vartanyants@desy.de

An improved analysis for single-particle imaging (SPI) experiments, using the limited data, is presented here. Results are based on a study of bacteriophage PR772 performed at the Atomic, Molecular and Optical Science instrument at the Linac Coherent Light Source as part of the SPI initiative. Existing methods were modified to cope with the shortcomings of the experimental data: inaccessibility of information from half of the detector and a small fraction of single hits. The general SPI analysis workflow was upgraded with the expectation-maximization based classification of diffraction patterns and mode decomposition on the final virus-structure determination step. The presented processing pipeline allowed us to determine the 3D structure of bacteriophage PR772 without symmetry constraints with a spatial resolution of 6.9 nm. The obtained resolution was limited by the scattering intensity during the experiment and the relatively small number of single hits.

1. Introduction

The recent worldwide outbreak of the COVID-19 pandemic has indicated urgency for the development of complementary imaging techniques for the study of virus structures at high resolution. Currently, cryogenic electron microscopy (cryo-EM) outperforms other methods of analysis of single virus particles, although to determine the structure one needs to freeze the samples at cryogenic temperatures. Understanding the nature of the same viruses at room temperature as well as



OPEN ACCESS

in aerosols may provide information on the actual size distribution of virus particles in aerosols, a factor that may affect infection of humans by these viruses.

As suggested about two decades ago, intense femtosecond pulses from X-ray free-electron lasers (XFELs) may outrun conventional radiation damage, found at synchrotron X-ray sources, or the Coulomb explosion, found at the intensities of XFELs. They therefore allow, in principle, atomic resolution structure determination of isolated macromolecules in their native environment at room temperatures (Neutze *et al.*, 2000). After the start of the first hard X-ray FELs (Emma *et al.*, 2010; Ishikawa *et al.*, 2012), it became clear that atomic resolution could be achieved with intense X-ray pulses from an XFEL on micro- and nano-crystals of proteins, leading to the serial femtosecond crystallography (SFX) technique (Chapman *et al.*, 2011; Aquila *et al.*, 2012; Boutet *et al.*, 2012). However, progress toward high-resolution 3D electron-density images of non-crystalline biological particles using the single-particle imaging (SPI) approach has been slow compared with SFX (Bogan *et al.*, 2008; Seibert *et al.*, 2011; Hantke *et al.*, 2014; van der Schot *et al.*, 2015; Ekeberg *et al.*, 2015).

There are several reasons for the lower resolution achieved in SPI experiments in comparison with the SFX technique. The most important include: lack of crystalline periodicity to amplify the signal, weak single-particle signal from non-periodic nanoscale objects compared with instrumental background, limited number of usable data frames collected and heterogeneity of the samples. Radiation-damage processes initiated by X-ray photoionization may also play an important role at high resolution, but this limitation is minimized at XFEL sources by performing experiments in the ‘diffract and destroy’ mode (Chapman *et al.*, 2006). One may expect that in order to further enhance resolution, one could increase the power of XFEL pulses to boost the signal-to-noise ratio for bio-particles, such as single proteins or virus particles that scatter very weakly. Unfortunately, increased XFEL fluence strips electrons from atoms more efficiently and the scattered power from the bound electrons (which contains structural information) does not increase in proportion to the X-ray fluence (Lorenz *et al.*, 2012; Gorobtsov *et al.*, 2015). Decreasing the pulse length below 1 fs may help to outrun Auger decay, but further decrease of the pulse duration may not lead to the desired suppression of all ionization channels (Gorobtsov *et al.*, 2015).

Beside these fundamental limitations, there are a few other limiting factors. Background scattering originating from the beamline obscures the weak scattering signals from biological samples, and fluctuations of the beam position and intensity can cause additional variables that need to be accounted for in reconstruction algorithms. Moreover, single-particle sample delivery remains a challenging topic. In addition, despite strong development of detector technology [see, for example, Allahgholi *et al.* (2019)], the dynamic range of the present detectors is still not sufficient to capture a full diffraction pattern in a strong single shot. To tackle all these limitations of the SPI technique and to push the methodology further, a

dedicated SPI consortium was formed at the Linac Coherent Light Source (LCLS) (Aquila *et al.*, 2015).

Several results were reported from the SPI consortium, covering both hard and soft X-ray experiments and focused primarily on well characterized viruses with sizes of a few tens of nanometres. Different analysis methods have been applied to virus structure determination [see, for example, Munke *et al.* (2016), Reddy *et al.* (2017), Kurta *et al.* (2017), Rose *et al.* (2018) and Shi *et al.* (2019)]. In the present work, we advance the previous studies performed on bacteriophage PR772 (Rose *et al.*, 2018) by extending the general analysis pipeline of the SPI data as first proposed by Gaffney & Chapman (2007). The main upgraded features include: a classification method based on the expectation-maximization (EM) algorithm, developed in cryo-EM (Dempster *et al.*, 1977), and mode decomposition for the final virus structure determination (Vartanyants & Singer, 2010; Thibault & Menzel, 2013). EM-based algorithms were first applied in SPI data analysis for 3D orientations recovery (Loh & Elser, 2009; Ayer *et al.*, 2016). Below we present a detailed description of all the steps that allowed us to obtain the electron density of the PR772 virus with an improved resolution compared with the previous SPI studies on the same virus (Rose *et al.*, 2018).

2. Experiment

The experiment was performed in the Atomic, Molecular and Optical Science (AMO) instrument (Bozek, 2009; Ferguson *et al.*, 2015; Osipov *et al.*, 2018) at the LCLS at SLAC National Accelerator Laboratory using the LAMP end station [see Li *et al.* (2020) for experimental details]. PR772 bacteriophage growth and purification were carried out essentially as previously described (Reddy *et al.*, 2017). Viruses in ammonium acetate volatile buffer were aerosolized in a helium environment using gas dynamic virtual nozzles (GDVN) that were 3D printed via 2 photon polymerization photo-lithography with a Nanoscribe Professional GT printer (Nazari *et al.*, 2020). Laboratory measurements showed reproducible jet diameters in the range of 0.5–2.0 μm (DePonte *et al.*, 2008; Weierstall *et al.*, 2012). The particles then passed through a differentially pumped skimmer for pressure reduction and were then injected/focused into the sample chamber using an aerodynamic lens injector (Benner *et al.*, 2008; Hantke *et al.*, 2014). The focused particle stream intersected the focused and pulsed XFEL beam.

The LCLS had a repetition rate of 120 Hz, for this experiment the average pulse energy was ~ 4 mJ, with a focal diameter of ~ 1.5 μm and a photon energy of 1.7 keV (wavelength 0.729 nm). Diffraction patterns were recorded by a pnCCD detector (Strüder *et al.*, 2010) mounted at a 0.130 m distance from the interaction region. In the experiment, a silver behenate salt was used as a calibration agent to determine sample detector distance and panel position. The scattering signal was only recorded by one of the two detector panels (one panel was not operational during the experiment because of an electronic fault). The size of the working panel was 512 by 1024 pixels with a pixel size of $75 \times 75 \mu\text{m}^2$, with

the long edge closest to the interaction point. In the middle of the experiment (at run 205) the detector panel was moved 1 mm vertically up relative to the incoming X-ray beam to reduce background scattering. Our work is based on the data obtained from this panel covering a part of reciprocal space, as shown in Figs. 1(a)–1(d).

3. Initial classification steps

SPI data analysis involves many subsequent steps leading to the 3D reconstructed particle structure from a large set of 2D diffraction patterns (Gaffney & Chapman, 2007). Improvements in the data-analysis pipeline at early stages can result in a significant enhancement in reconstruction quality. Therefore, several pre-processing methods were developed to avoid experimental artifacts on the collected diffraction patterns. The pre-processing stages include: hit finding, background correction, beam-position refinement and particle-size filtering.

The initial experimental dataset, as collected at the AMO instrument, consists of $\sim 1.2 \times 10^7$ diffraction patterns (9×10^6 patterns before and 3×10^6 patterns after moving the detector panel) (see Li *et al.*, 2020). The hit finding was performed using the software *psocake* in the *psana* framework (Damiani *et al.*,

2016). As a result, 1.9×10^5 diffraction patterns were identified as hits from the initial set of diffraction patterns (see Table 1) and the signal from these hits was converted to photon counts. Examples of selected diffraction patterns are shown in Fig. 1.

Visual inspection of the selected patterns revealed the presence of additional instrumental scattering near the center of the diffraction patterns [see Fig. S1(a) in the Supporting information]. The scattering was caused by the interaction of the tails of the FEL beam with the upstream apertures or from the sample injector. To subtract this additional signal, histograms of intensity for each pixel were analyzed (Figs. S1 and S2). It was assumed that for pixels with zero photon counts, the histogram of intensity distribution should have a Gaussian form and a mean value equal to zero. The mean value of this distribution was subtracted from each diffraction pattern. Subtraction of this additional scattering contribution was an important step prior to finding the beam center position (see the Supporting information for details).

The particle-size filtering was performed in two steps. First, the power spectral density (PSD) function, *i.e.* angular averaged intensity, of each diffraction pattern was fitted with the form factor of a solid sphere in a wide range of sizes from 30 to 300 nm. Next, a fidelity criterion was introduced to distinguish

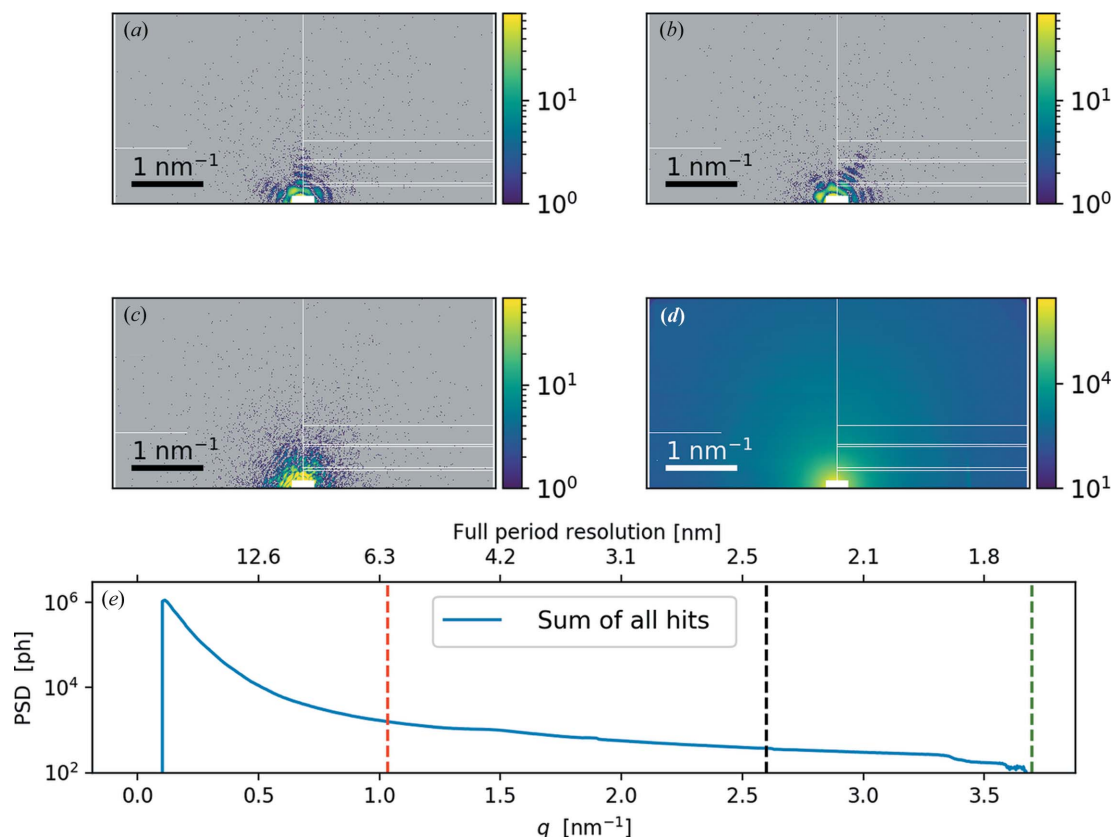


Figure 1

Examples of diffraction patterns from the SPI experiment. (a), (b) Diffraction patterns corresponding to a single PR772 virus hit by an XFEL beam. (c) A diffraction pattern corresponding to a non-single hit which was sorted out from the analysis at the classification step. (d) A sum of 1.9×10^5 diffraction patterns identified as hits. White regions in the center of the diffraction patterns as well as white stripes correspond to a mask introduced to reduce artefacts owing to the signal exceeding the detector capabilities. The mask was the same before and after the move of the detector panel. (e) The PSD function of the scattered intensity for the sum of all diffraction patterns identified as hits collected in the experiment. The signal until the corner and edge of the detector is indicated by the green and black vertical dashed lines, respectively. For orientation determination the data until momentum-transfer values of 1 nm^{-1} (shown by the red dashed line) were used.

Table 1

Datasets selected at different stages of the analysis: hit finding selection, PSD-fitting quality filtering, particle-size filtering and single-hit diffraction pattern selection.

The percentage of the chosen dataset to the initial one S_0 is given in parentheses.

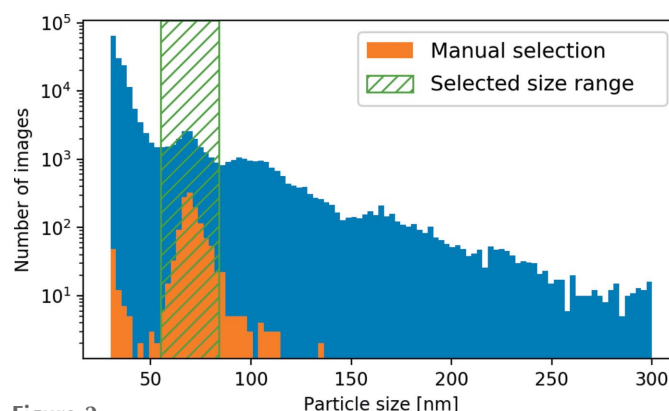
Analysis step	Dataset name	Number of diffraction patterns
Initial dataset	S_0	1.2×10^7
Hit-finding classification	S_{hit}	191183 (1.6%)
PSD-fitting score filtering	S_{fit}	179886 (1.5%)
Particle-size filtering	S_D	18213 (0.1%)
First EM-based classification	S_1^{EM}	1609
Second EM-based classification	S_2^{EM}	1402
Third EM-based classification	S_3^{EM}	1366
Fourth EM-based classification	S_4^{EM}	1401
Fifth EM-based classification	S_5^{EM}	2119
Final EM-based classification	S_{EM}	1085 (0.009%)
Manual selection	S_{man}	1393 (0.01%)

between successful and unsuccessful size findings (see the Supporting information for details). As an outcome of this analysis, the number of selected diffraction patterns was reduced from 1.9×10^5 to 1.8×10^5 , which were used for the final particle-size filtering. A histogram for different particle sizes in the selected size range is shown in Fig. 2. An extended range of sizes observed in this figure corresponds to clusters of particles stuck together and the varying thickness of a hydration layer. One can also identify a peak in this histogram in the range from 55–84 nm. This size range agrees well with the expected virus particle size of ~ 70 nm (Rose *et al.*, 2018; Reddy *et al.*, 2019). The considered range contains 1.8×10^4 diffraction patterns, which were selected for further single-hit classification (see Table 1).

4. Single-hit diffraction patterns classification

The key step of data selection in this work was single-hit classification. The angular X-ray cross-correlation analysis classification that was used in the previous work (Bobkov *et al.*, 2015; Rose *et al.*, 2018) was not as effective with the present experimental data owing to the absence of the scattering signal on one half of the detector and the low single-hit rate.

The single-hit classification approach was based on the EM algorithm (Dempster *et al.*, 1977; Loh & Elser, 2009; Ayer *et al.*, 2016). This algorithm allows for unsupervised clustering of data when neither initial data assignments to clusters nor cluster parameters are known. The dataset is distributed into a pre-defined number of clusters and cluster parameters are retrieved automatically at the same time. Later, manual input is used to classify each of the cluster models, based on symmetry considerations or expected fringe patterns, in order to perform data selection. This algorithm is commonly used in cryo-EM for unsupervised single-particle clustering [see, for example, the software *RELION* (Scheres *et al.*, 2005)]. Low contrast and low signal-to-noise level are common problems for cryo-EM and SPI so we implemented in this work the original EM algorithm developed in cryo-EM for clustering of SPI data.

**Figure 2**

A particle-size histogram after the PSD-function filtering. The blue area corresponds to 1.8×10^5 diffraction patterns, satisfying the fidelity score criterion $FS > 1.05$ (see the Supporting information for details). The orange area corresponds to manually selected single-hit diffraction patterns. A range of particle sizes from 55 to 84 nm (1.8×10^4 diffraction patterns) was selected for the further SPI analysis (green dashed area).

Unsupervised EM clustering is an iterative process; the algorithm starts from a random model for each class. At each iteration the probability for each diffraction pattern to belong to a certain class is calculated. To accommodate for random particle orientations in the SPI experiment, the cluster model is compared with the 2D diffraction pattern rotated in-plane. After probabilities are evaluated, a new cluster model is calculated by weighted averaging of patterns that belong to each cluster in each orientation. The weights are defined by computed probabilities. In fact, the algorithm imitates the expand–maximize–compress (EMC) algorithm (Loh & Elser, 2009) but, instead of rotation in 3D space, an in-plane rotation and cluster distribution are used (Loh, 2014). When the EM algorithm converges, the supervised class selection takes place. The cluster models that correspond to single hits of an investigated particle are selected manually by an expert. As an example, in one of the EM clustering we selected classes 1 and 2 as classes containing six high-contrast fringes that we attribute to scattering from a PR772 virus particle [see Fig. 3(a)].

To make classification more accurate we performed five independent EM clusterings starting from random cluster models. Each EM clustering produced slightly different results, which are summarized in Table 1. The intersection of all the results was considered as a stable single-hit selection [see Fig. 3(b)]. Finally, we ended with the dataset containing 1085 single-particle patterns (see Table 1). The PSD function of an average of all the selected diffraction patterns is shown in Fig. 3(c). The scattering signal from the virus particle extends up to the momentum-transfer value $q_{\text{max}} = 1 \text{ nm}^{-1}$ in reciprocal space, which corresponds to a resolution ($2\pi/q_{\text{max}}$) of 6.3 nm in real space.

A manual search and selection of single hits from the dataset of 1.9×10^5 diffraction patterns produced a new dataset containing 1393 diffraction patterns (Li *et al.*, 2020). From this selection, 574 diffraction patterns are also present in our EM-based selection. The PSD function for the manual selection shown in Fig. 3(c) has a lower contrast without visible fringes. We attribute this mostly to the absence of the

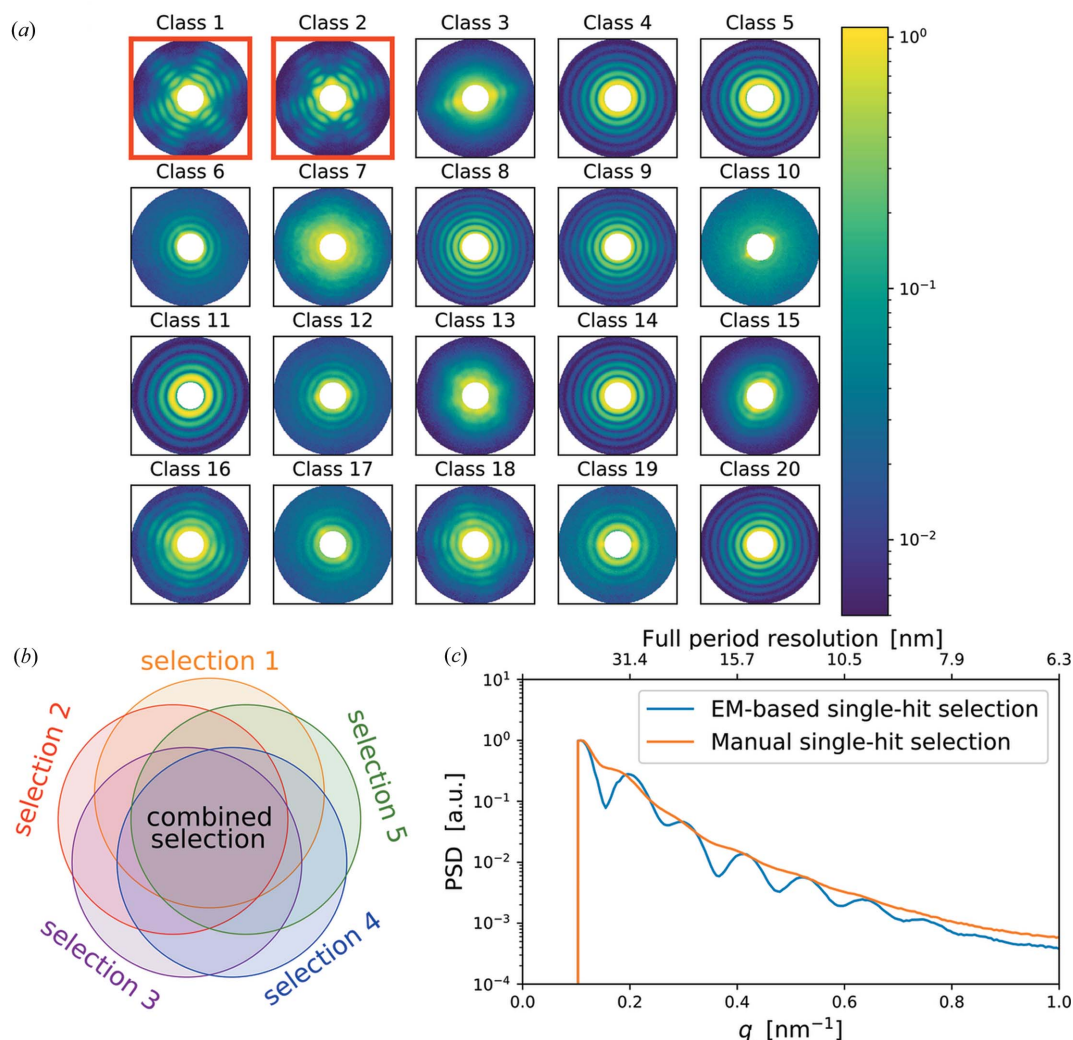


Figure 3

Classification of diffraction patterns by EM clustering. (a) Diffraction patterns are distributed into 20 classes according to their features. Classes 1 and 2 were selected as they clearly contain structural features of the investigated virus and its icosahedral shape. These two classes contain 1609 diffraction patterns. (b) EM-clustering was repeated five times, and intersecting selection with 1085 patterns was considered for further analysis. (c) Averaged PSD functions for EM-based single-hit selection containing 1085 patterns (blue line) and for manual selection containing 1393 patterns (orange line).

size-filtering step in manual selection. The virus size fluctuations could be caused by a slight change in hydration-layer thickness, variation of the position where the particles interact with the X-ray beam, slightly upstream or downstream of the focus, or a real sample size distribution of the PR772 virus. We note that in our previous work (Rose *et al.*, 2018) the number of single-hit diffraction patterns was $\sim 7.3 \times 10^3$, which is about one order of magnitude higher than in this work. The reason for the smaller number of patterns was a combination of downtime owing to detector troubleshooting and time needed for sample-injection optimizations. The GDVN flow and pressures needed to be optimized because of the relatively large initial droplet diameters of $\sim 1\text{--}2\ \mu\text{m}$.

5. Orientation determination and background subtraction

For orientation determination we used the EMC algorithm implemented in the *Dragonfly* software package (Ayyer *et al.*,

2016). This iterative algorithm successfully combines 2D diffraction patterns into the 3D intensity distribution of the PR772 virus shown in Fig. 4(a) (see the Supporting information for further details).

The results presented in Fig. 4(a) clearly show that the recovered 3D diffraction intensity contains a substantial high-momentum background that may be caused by scattering on helium from the carrier gas. To reduce the influence of this background contribution, a background subtraction was applied. Several approaches for the background correction in SPI experiment analysis were developed earlier (Rose *et al.*, 2018; Lundholm *et al.*, 2018; Ayyer *et al.*, 2019) and it was understood that the background-correction method may affect the reconstructed structure. Here we used a combined approach for the background determination, first a constant background was subtracted and then, on the reconstruction step, the contrast of diffraction patterns was additionally enhanced by applying deconvolution algorithms (see the next section for details). In the first step, the background level was

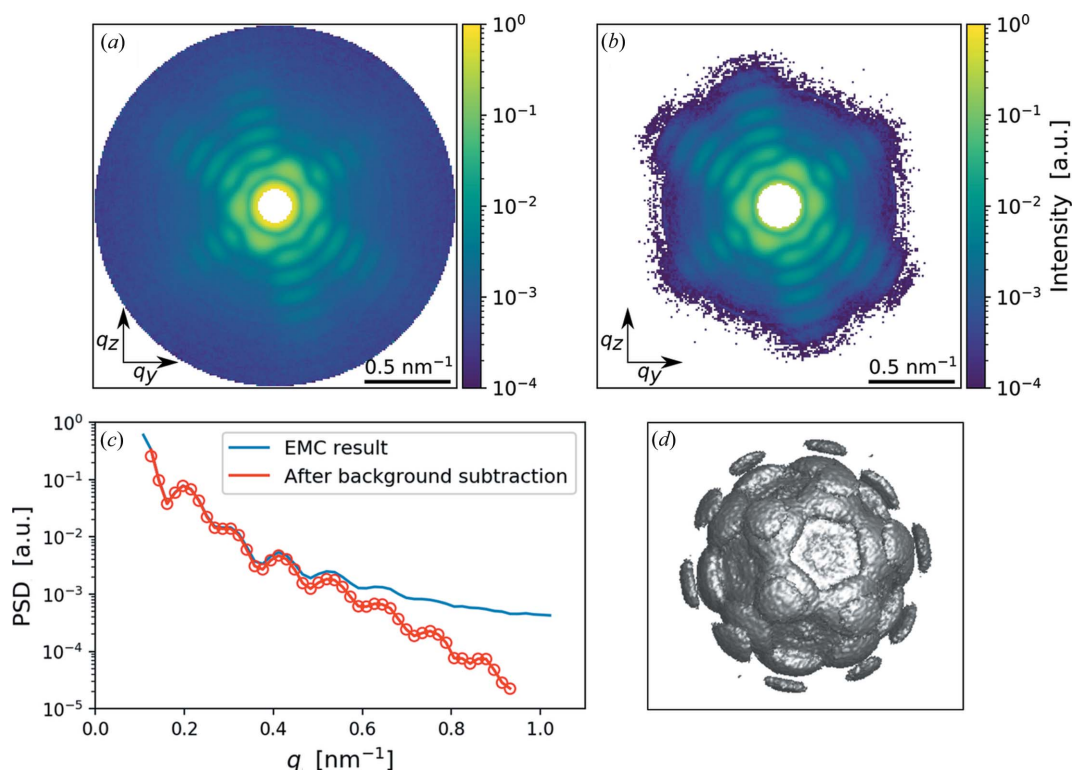


Figure 4

Orientation-determination results. (a) A 2D q_y – q_z cut of the 3D intensity distribution in reciprocal space and (b) the same intensity distribution after the background subtraction. (c) The PSD function for the EMC result (blue line) and after the background subtraction (red line with dots). It is well visible that the background subtraction enhanced structural visibility in the high q region. (d) The 3D intensity distribution in reciprocal space after the background subtraction, shown at 0.5% level of the maximum value.

defined as the mean signal value in the high-momentum region of the 3D intensity distribution, free of particle-scattering contributions (see the Supporting information for details). Reciprocal-space data with and without background subtraction are shown in Figs. 4(a)–4(c). As seen in Fig. 4(c), the power-law dependence of data after the background subtraction is the same up to a high q range of about one inverse nanometre. This is in contrast to the data after EMC orientation determination [the blue curve in Fig. 4(c)], which are clearly saturated at high q range.

As a result of the background subtraction, more features and higher contrast were revealed in the high momentum-transfer region. The fringe visibility or averaged contrast $\langle \gamma \rangle$ was defined as

$$\langle \gamma \rangle = \frac{1}{N} \sum_{i=1}^N \gamma_i, \text{ where } \gamma_i = \frac{I_{\max}^i - I_{\min}^i}{I_{\max}^i + I_{\min}^i}. \quad (1)$$

Here γ_i is the local contrast, $I_{\max}^i$ and $I_{\min}^i$ are the PSD-function values in the local maxima and following minima, respectively, and N is a number of pairs of maxima and minima considered in this analysis. In our case, the average contrast values $\langle \gamma \rangle$ were calculated for the first $N = 6$ pairs. For the experimental data before the background subtraction, we obtained the value $\langle \gamma \rangle = 0.41$. The fringe contrast after the background subtraction showed significant improvement with $\langle \gamma \rangle = 0.58$.

To proceed further with the analysis, low ($q < 0.12 \text{ nm}^{-1}$) and high ($q > 0.93 \text{ nm}^{-1}$) momentum-transfer regions were

eliminated from the final data selection owing to scattering artefacts present in these regions. After the background subtraction, negative values of intensity were set to zero. A 3D intensity distribution in reciprocal space after the background subtraction is shown in Fig. 4(d).

6. Phase retrieval and PR772 virus structure

To determine the electron-density distribution of the PR772 virus, the phase retrieval was performed using the 3D intensity distribution of the virus in reciprocal space described in the previous section [see Fig. 4(d)]. Iterative phase-retrieval algorithms are based on the Fourier transform between real and reciprocal/diffraction space using two constraints: in reciprocal space the amplitude of the signal is set equal to the experimentally measured values and in real space a finite support of the particle is used (Fienup, 1982; Marchesini, 2007).

The electron-density distribution of the PR772 virus was obtained with the following steps. First, the central gap, originating from the initial data masking of the diffraction patterns, was filled [see Figs. 4(a) and 4(b)]. This was accomplished by running multiple 3D reconstructions of the virus and leaving the central gap of the diffraction volume to freely evolve. All reconstructions produced practically identical continuous 3D functions in reciprocal space that smoothly covered the central gap (see Fig. S10). This reciprocal-space distribution with the filled central part was used for the next

reconstruction step. On that step, 50 successive reconstructions using a combination of continuous hybrid input–output (Fienup, 2013), error reduction (Fienup, 1982) and shrink-wrap algorithms (Marchesini *et al.*, 2003) were performed. In addition to these algorithms, we used the Richardson–Lucy deconvolution technique (Clark *et al.*, 2012). Application of this method allowed us to take into consideration additional background and substantially enhance the contrast of the reconstructed diffraction pattern to the value of $\langle \gamma \rangle = 0.87$. As a result, we obtained complex-valued real-space images for each of the 50 reconstructions (see the Supporting information for details).

On the final step, to identify the electron density of the virus we used mode decomposition. We first introduced a density matrix in the form

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \langle \rho^*(\mathbf{r}_1) \rho(\mathbf{r}_2) \rangle, \quad (2)$$

where $\rho(\mathbf{r})$ relates to complex-valued real-space images and the brackets $\langle \dots \rangle$ indicate ensemble averaging over different reconstructions. We further decomposed the density matrix into orthogonal modes $\rho_n(\mathbf{r})$ using principal component analysis (Khubbutdinov *et al.*, 2019):

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \sum_{n=0}^N \beta_n \rho_n^*(\mathbf{r}_1) \rho_n(\mathbf{r}_2), \quad (3)$$

where β_n refers to eigenvalues of this decomposition. This approach is advantageous in comparison with averaging that would make important object features blurry and possibly affect the final resolution. By considering the zero mode only, unique features present in all other reconstructions will be represented. In practice, we used all 50 reconstruction results and performed mode decomposition. The fundamental mode (with the weight factor β_0 of 99%) was considered as the final result of reconstruction. Such a high weight-factor value indicates that all reconstructions converged essentially to the same result with an uncertainty level of only 1%. To determine the electron density, we took an absolute value of this fundamental mode complex-valued image. A three-times up-sampled version of the real-space electron density is shown in Fig. 5 (see the Supporting information for details).

As seen in Figs. 5(a)–5(c), the retrieved electron density of bacteriophage PR772 shows the expected icosahedral structure. A 2D cut of the virus structure is presented in Fig. 5(d), where a higher electron density in a thin outer shell is well resolved and is attributed to the capsid proteins arrangement. As can be seen in Fig. 5(d), the electron density in the center of the reconstructed virus particle is reduced. The reason for this may be the heterogeneity of virus particles present in solution and injected in the X-ray beam.

Bacteriophage PR772, like other members of the *Tectiviridae* family, contains an inner proteolipid membrane that facilitates delivery of the viral genomic DNA during infection (Mäntynen *et al.*, 2019). When the virus binds to a bacterial host cell the inner membrane is extruded from one of the viral vertices to form a nanotube that facilitates genome delivery. Spontaneous release of the viral genome has been reported

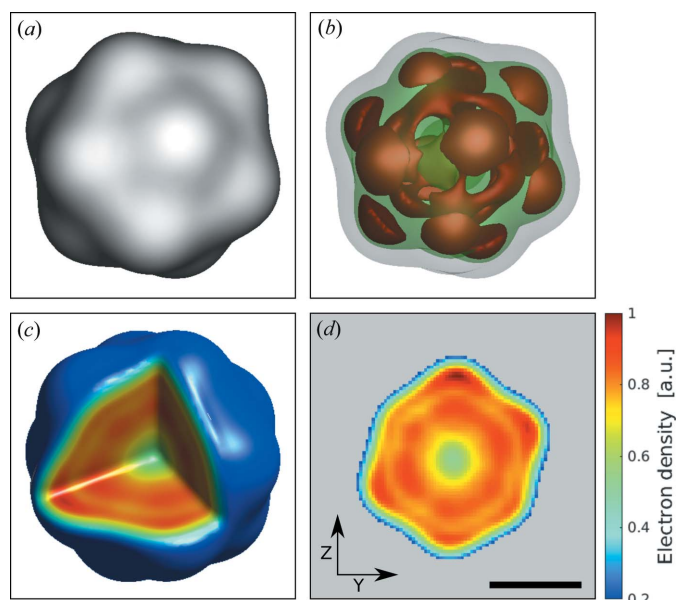


Figure 5

Electron density of the reconstructed PR772 virus normalized to the maximum value. (a) The outer structure of the PR772 virus at the isosurface value of 20% of the maximum electron density. (b) The inner 3D structure of the PR772 virus at the isosurface values of 85% (brown area), 75% (green area) and 20% (gray area). (c) A 3D section of the virus. (d) An electron-density slice of the virus. Amplitude values of less than 0.2 were set to gray color scale. The color map for (c) and (d) is the same. Images were up-sampled three times for better visibility. The scale bar in (d) is 30 nm.

(Peralta *et al.*, 2013; Santos-Pérez *et al.*, 2017). Our bacteriophage PR772 preparations were also analyzed by cryo-EM (Fig. S6). Under the conditions used for plunge freezing, we observed particles that were more rounded than icosahedral. Preliminary volume analysis suggests that some particles had released their DNA. These particles appeared to be ‘triggered’ during the freezing process since transmission electron microscopy imaging did not reveal this. It may be possible from minor differences in virus preparation or upon XFEL sample delivery that some of the virus particles were similarly ‘triggered’ to release their genomes during cryo-freezing, thus resulting in a decrease in inner electron density. Previous analyses of XFEL snapshots on the same virus also suggested that some particles exhibit decreased inner density (Shi *et al.*, 2019; Hosseinizadeh *et al.*, 2017).

To identify the particle size, the electron-density profiles in different directions were investigated. In this work, we used the same approach as developed earlier (Rose *et al.*, 2018) and analyzed the electron-density profiles in the directions from facet to facet and from vertex to vertex of the reconstructed virus particle [see Figs. 6(a) and 6(b)]. For the particle-size estimate, we selected the electron-density threshold value of 0.2 as it was considered in the shrink-wrap algorithm during the phase retrieval. From this criterion we determined the particle size in different directions (see Table 2 and the Supporting information). Thus, the obtained mean particle size was 61 ± 2 nm between facets and 63 ± 2 nm between vertices. These sizes correspond well to the initial range of particle sizes (from 55 to 84 nm) considered at the initial

Table 2

Particle-size analysis from facet to facet and from vertex to vertex for the reconstructed PR772 virus.

All distances between facets and vertices are given in the Supporting information.

		Size (nm)
Facet to facet	D_{mean}	61 ± 2
	D_{max}	64 ± 2
	D_{min}	59 ± 2
Vertex to vertex	D_{mean}	63 ± 2
	D_{max}	67 ± 2
	D_{min}	60 ± 2

classification step (see Section 4) and other XFEL data performed on the same virus (Kurta *et al.*, 2017; Rose *et al.*, 2018).

Similar to previous work (Kurta *et al.*, 2017; Rose *et al.*, 2018), we also observed a certain elongation of the particle shape, which might be inherently present in the viruses in solution or appear owing to the aerosol injection conditions. We defined the elongation of the particle by the following measure,

$$\alpha = \frac{D_{\text{max}} - D_{\text{min}}}{D_{\text{mean}}}, \quad (4)$$

where D_{max} , D_{min} and D_{mean} are maximum, minimum and mean particle sizes, respectively. The virus structure obtained in the current work showed an elongation value of $\alpha = 11\%$ for sizes taken between vertices, which is similar to the result of the previous SPI experiments (Rose *et al.*, 2018) with $\alpha = 9\%$.

The mean capsid thickness was obtained from the Gaussian fitting of the electron-density profiles with well pronounced features (Fig. S14) and was determined to be 7.6 ± 0.3 nm. The thickness of the capsid in recent cryo-EM studies (Reddy *et al.*, 2019) was identified to be below 10 nm, which is in good agreement with the thickness determined in this experiment.

Finally, we determined the resolution of the reconstructed electron density of the PR772 virus. We used the Fourier shell correlation (FSC) approach (Harauz & van Heel, 1986) to determine the resolution of the reconstructed virus. For this method two independent sets of reconstruction were required, usually each is based on half of the available dataset. FSC measures the normalized cross-correlation coefficient between both reconstructions over corresponding shells in Fourier space. The results of the FSC analysis are shown in Fig. 7. To estimate the achieved resolution, we used half-bit threshold criteria (van Heel & Schatz, 2005). Its intersection with the FSC curve gave a resolution value of 6.9 nm (see Fig. 7). The obtained result is better than the previously reported

value of ~ 8 nm for the same virus (Rose *et al.*, 2018), though the number of diffraction patterns used for the final analysis was much lower (15% of the previous dataset). In this case, the resolution was limited by the number of diffraction patterns and moderate scattered intensity, whereas previous reconstructions were limited by the extent of the detector.

7. Summary and outlook

We have presented an implementation of the SPI data-analysis workflow from diffraction patterns measured at the AMO instrument at the LCLS to reconstruct the electron density of bacteriophage PR772. This work is an important methodological step for development of SPI data analysis. We implemented several methods into the workflow including EM-based classification and mode decomposition that were crucial for the high-resolution final reconstruction. Although only half of the detector was operational, implementation of all of these steps allowed us to determine the PR772 virus structure with a higher resolution as compared with previous SPI studies.

From the initial set of 1.2×10^7 experimentally measured diffraction patterns, the final single-hit selection set contained 1085 diffraction patterns. About 53% of this final set was also present in the single-hit selection made manually. The number of diffraction patterns classified for further analysis is substantially lower than in the previous experiment (Rose *et al.*, 2018), improvements in sample delivery to increase the number of diffraction patterns are, therefore, crucial for advancements in reconstruction resolution.

The combination of all methods implemented in the workflow allowed us to obtain the 3D electron density of the virus with a resolution of 6.9 nm based on the FSC analysis, which is better than that obtained in the previous studies (Rose *et al.*, 2018). The obtained mean PR772 virus size in this experiment was 61 nm (between facets) and 63 nm (between vertices), like in the previous SPI experiments performed on the same PR772 virus (Kurta *et al.*, 2017; Rose *et al.*, 2018). We also observed a similar elongation of $\sim 11\%$ of the virus structure as was determined in the previous experiments.

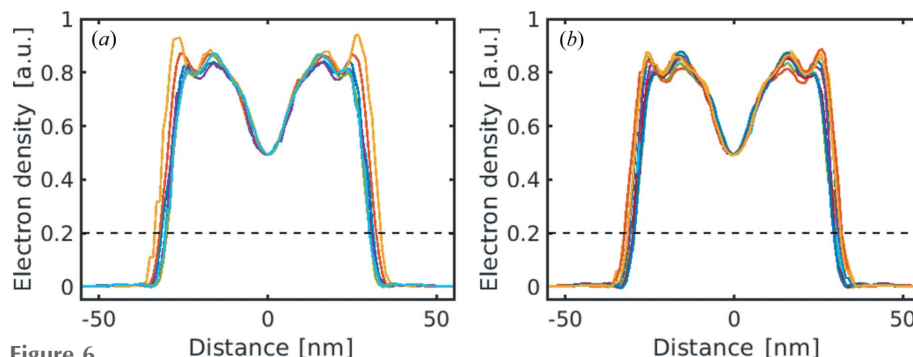


Figure 6 Electron-density profiles of the reconstructed virus PR772 normalized to the maximum value for the cut between vertices (a) and facets (b). The horizontal black dashed lines denote a particle-size threshold of 0.2. The mean virus size is 63 and 61 nm for the distance between vertices and between facets, respectively.

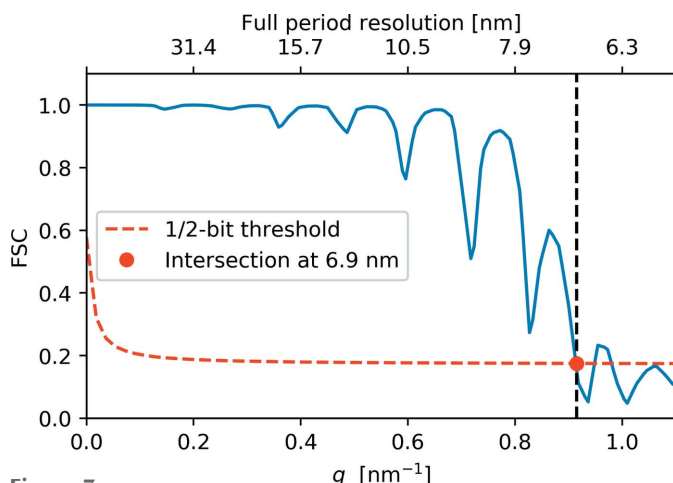


Figure 7
FSC of the final reconstruction (blue line) that shows a 6.9 nm resolution (red dot) with a half-bit threshold (red dashed line).

The presented research is another step forward in SPI data analysis. Implemented steps may become especially important when SPI experiments are performed at high repetition rate XFELs, such as the European XFEL (Decking *et al.*, 2020; Mancuso *et al.*, 2019) and LCLS-II (Schoenlein *et al.*, 2015) facilities. The first experiments performed at the European XFEL demonstrated the possibility of collecting the SPI data at a megahertz rate (Sobolev *et al.*, 2020), which might be crucial for future progress in SPI experiments performed at XFELs.

Acknowledgements

The authors acknowledge the invaluable support of the technical and scientific staff of the LCLS at SLAC National Accelerator Laboratory. The authors also acknowledge support of the project and discussions with E. Weckert and careful reading of the manuscript by A. Schropp.

Funding information

DA, YYK, SB, RK, MR, AT, VAI and IAV acknowledge support from the Helmholtz Associations Initiative Networking Fund grant No. HRSF-0002 and the Russian Science Foundation grant No. 18-41-06001. HD and PS acknowledge the US National Science Foundation (NSF) Science and Technology Center BioXFEL Award 1231306. HD acknowledges the Scientific and Technological Research Council of Turkey (TUBITAK) grant (118C270). JA acknowledges support from the project ‘Advanced research using high-intensity laser-produced photons and particles’ (CZ.02.1.01/0.0/0.0/16_019/0000789) from the European Regional Development Fund (ADONIS) and the Ministry of Education, Youth and Sports as part of targeted support from the National Program of Sustainability II. PLX acknowledges a grant from the European Research Council, ‘Frontiers in Attosecond X-ray Science: Imaging and Spectroscopy (AXSIS)’, ERC-2013-SyG 609920 (2014–2018), the Human Frontier Science Program (RGP0010/2017) and the fellowship

from the Joachim Herz Stiftung. JH acknowledges the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the European Research Council, the Röntgen–Ångström Cluster and the project ‘Structural dynamics of biomolecular systems’ (CZ.02.1.01/0.0/0.0/15_003/0000447) from the European Regional Development Fund. EB acknowledges support by the Australian Research Council Centre of Excellence in Advanced Molecular Imaging (CE140100011). The authors acknowledge use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, where use is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

References

- Allahgholi, A., Becker, J., Delfs, A., Dinapoli, R., Goettlicher, P., Greiffenberg, D., Henrich, B., Hirsemann, H., Kuhn, M., Klanner, R., Klyuev, A., Krueger, H., Lange, S., Laurus, T., Marras, A., Mezza, D., Mozzanica, A., Niemann, M., Poehlsen, J., Schwandt, J., Sheviakov, I., Shi, X., Smoljanin, S., Steffen, L., Sztuk-Dambietz, J., Trunk, U., Xia, Q., Zeribi, M., Zhang, J., Zimmer, M., Schmitt, B. & Graafsma, H. (2019). *J. Synchrotron Rad.* **26**, 74–82.
- 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.
- Aquila, A., Hunter, M. S., Doak, R. B., Kirian, R. A., Fromme, P., White, T. A., Andreasson, J., Arnlund, D., Bajt, S., Barends, T. R. M., Barthelmess, M., Bogan, M. J., Bostedt, C., Bottin, H., Bozek, J. D., Coleman, C., Coppola, N., Davidsson, J., DePonte, D. P., Elser, V., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Frank, M., Fromme, R., Graafsma, H., Grotjohann, I., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, R., Hau-Riege, S., Hauser, G., Hirsemann, H., Holl, P., Holton, J. M., Hömke, A., Johansson, L., Kimmel, N., Kassemeyer, S., Krasniqi, F., Kühnel, K.-U., Liang, M., Lomb, L., Malmerberg, E., Marchesini, S., Martin, A. V., Maia, F. R. N. C., Messerschmidt, M., Nass, K., Reich, C., Neutze, R., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schmidt, C., Schmidt, K. E., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R., Soltau, H., Starodub, D., Stellato, F., Stern, S., Strüder, L., Timneanu, N., Ullrich, J., Wang, X., Williams, G. J., Weidenspointner, G., Weierstall, U., Wunderer, C., Barty, A., Spence, J. C. H. & Chapman, H. N. (2012). *Opt. Express*, **20**, 2706.
- Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. (2016). *J. Appl. Cryst.* **49**, 1320–1335.
- Ayyer, K., Morgan, A. J., Aquila, A., DeMirici, H., Hogue, B. G., Kirian, R. A., Xavier, P. L., Yoon, C. H., Chapman, H. N. & Barty, A. (2019). *Opt. Express*, **27**, 37816.
- Benner, W. H., Bogan, M. J., Rohner, U., Boutet, S., Woods, B. & Frank, M. (2008). *J. Aerosol Sci.* **39**, 917–928.
- Bobkov, S. A., Teslyuk, A. B., Kurta, R. P., Gorobtsov, O. Yu., Yefanov, O. M., Ilyin, V. A., Senin, R. A. & Vartanyants, I. A. (2015). *J. Synchrotron Rad.* **22**, 1345–1352.
- Bogan, M. J., Benner, W. H., Boutet, S., Rohner, U., Frank, M., Barty, A., Seibert, M. M., Maia, F., Marchesini, S., Bajt, S., Woods, B., Riot, V., Hau-Riege, S. P., Svenda, M., Marklund, E., Spiller, E., Hajdu, J. & Chapman, H. N. (2008). *Nano Lett.* **8**, 310–316.
- Boutet, S., Lomb, L., Williams, G. J., Barends, T. R. M., Aquila, A., Doak, R. B., Weierstall, U., DePonte, D. P., Steinbrener, J., Shoeman, R. L., Messerschmidt, M., Barty, A., White, T. A., Kassemeyer, S., Kirian, R. A., Seibert, M. M., Montanez, P. A.,

- Kenney, C., Herbst, R., Hart, P., Pines, J., Haller, G., Gruner, S. M., Philipp, H. T., Tate, M. W., Hromalik, M., Koerner, L. J., van Bakel, N., Morse, J., Ghonsalves, W., Arnlund, D., Bogan, M. J., Coleman, C., Fromme, R., Hampton, C. Y., Hunter, M. S., Johansson, L. C., Katona, G., Kupitz, C., Liang, M., Martin, A. V., Nass, K., Redecke, L., Stellato, F., Timneanu, N., Wang, D., Zatsepin, N. A., Schafer, D., Devever, J., Neutze, R., Fromme, P., Spence, J. C. H., Chapman, H. N. & Schlichting, I. (2012). *Science*, **337**, 362–364.
- Bozek, J. D. (2009). *Eur. Phys. J. Spec. Top.* **169**, 129–132.
- 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., Hult, G., Seibert, M. M., Maia, F. R. N. C., Lee, R. W., Szöke, A., Timneanu, N. & Hajdu, J. (2006). *Nat. Phys.* **2**, 839–843.
- 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., Coleman, 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., Soltan, 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.
- Clark, J. N., Huang, X., Harder, R. & Robinson, I. K. (2012). *Nat. Commun.* **3**, 993.
- Damiani, D., Dubrovin, M., Gaponenko, I., Kroeger, W., Lane, T. J., Mitra, A., O'Grady, C. P., Salnikov, A., Sanchez-Gonzalez, A., Schneider, D. & Yoon, C. H. (2016). *J. Appl. Cryst.* **49**, 672–679.
- Decking, W., Abeghyan, S., Abramian, P., Abramsky, A., Aguirre, A., Albrecht, C., Alou, P., Altarelli, M., Altmann, P., Amyan, K., Anashin, V., Apostolov, E., Appel, K., Auguste, D., Ayvazyan, V., Baark, S., Babies, F., Baboi, N., Bak, P., Balandin, V., Baldinger, R., Baranasic, B., Barbanotti, S., Belikov, O., Belokurov, V., Belova, L., Belyakov, V., Berry, S., Bertucci, M., Beutner, B., Block, A., Blöcher, M., Böckmann, T., Böhm, C., Böhnert, M., Bondar, V., Bondarchuk, E., Bonezzi, M., Borowiec, P., Bösch, C., Bösenberg, U., Bosotti, A., Böspflug, R., Bousonville, M., Boyd, E., Boshko, Y., Brand, A., Branlard, J., Briechle, S., Brinker, F., Brinker, S., Brinkmann, R., Brockhauser, S., Brovko, O., Brück, H., Brüdgam, A., Butkowski, L., Büttner, T., Calero, J., Castro-Carballo, E., Cattalano, G., Charrier, J., Chen, J., Cherepenko, A., Cheskidov, V., Chiodini, M., Chong, A., Choroba, S., Chorowski, M., Churanov, D., Cichalewski, W., Clausen, M., Clement, W., Cloué, C., Cobos, J. A., Coppola, N., Cunis, S., Czuba, K., Czwalińska, M., D'Almagne, B., Dammann, J., Danared, H., de Zubiaurre Wagner, A., Delfs, A., Delfs, T., Dietrich, F., Dietrich, T., Dohlus, M., Dommach, M., Donat, A., Dong, X., Doynikov, N., Dressel, M., Duda, M., Duda, P., Eckoldt, H., Ehsan, W., Eidam, J., Eints, F., Engling, C., Englisch, U., Ermakov, A., Escherich, K., Eschke, J., Saldin, E., Faesing, M., Fallou, A., Felber, M., Fenner, M., Fernandes, B., Fernández, J. M., Feuker, S., Filippakopoulos, K., Floettmann, K., Fogel, V., Fontaine, M., Francés, A., Martin, I. F., Freund, W., Freyermuth, T., Friedland, M., Fröhlich, L., Fusetti, M., Fydrich, J., Gallas, A., García, O., García-Tabares, L., Geloni, G., Gerasimova, N., Gerth, C., Geßler, P., Gharibyan, V., Gloor, M., Glowinkowski, J., Goessel, A., Gołębiewski, Z., Golubeva, N., Grabowski, W., Graeff, W., Grebentsov, A., Grecki, M., Grevsmuehl, T., Gross, M., Grosse-Wortmann, U., Grünert, J., Grunewald, S., Grzegory, P., Feng, G., Guler, H., Gusev, G., Gutierrez, J. L., Hagge, L., Hamberg, M., Hanneken, R., Harms, E., Hartl, I., Hauberg, A., Hauf, S., Hauschildt, J., Hauser, J., Havlicek, J., Hedqvist, A., Heidbrook, N., Hellberg, F., Henning, D., Hensler, O., Hermann, T., Hidvégi, A., Hierholzer, M., Hintz, H., Hoffmann, F., Hoffmann, M., Hoffmann, M., Holler, Y., Hüning, M., Ignatenko, A., Ilchen, M., Iluk, A., Iversen, J., Iversen, J., Izquierdo, M., Jachmann, L., Jardon, N., Jastrow, U., Jensch, K., Jensen, J., Jeżabek, M., Jidda, M., Jin, H., Johansson, N., Jonas, R., Kaabi, W., Kaefer, D., Kammering, R., Kapitzka, H., Karabekyan, S., Karstensen, S., Kasprzak, K., Katalev, V., Keese, D., Keil, B., Kholopov, M., Killenberger, M., Kitaev, B., Klimchenko, Y., Klos, R., Knebel, L., Koch, A., Koepe, M., Köhler, S., Köhler, W., Kohlstrunk, N., Konopkova, Z., Konstantinov, A., Kook, W., Koprek, W., Körfer, M., Korth, O., Kosarev, A., Kosiński, K., Kostin, D., Kot, Y., Kotarba, A., Kozak, T., Kozak, V., Kramert, R., Krasilnikov, M., Krasnov, A., Krause, B., Kravchuk, L., Krebs, O., Kretschmer, R., Kreutzkamp, J., Kröplin, O., Krzysik, K., Kube, G., Kuehn, H., Kujala, N., Kulikov, V., Kuzmynich, V., La Civita, D., Lacroix, M., Lamb, T., Lancetov, A., Larsson, M., Le Pividic, D., Lederer, S., Lensch, T., Lenz, D., Leuschner, A., Levenhagen, F., Li, Y., Liebing, J., Lilje, L., Limberg, T., Lipka, D., List, B., Liu, J., Liu, S., Lorbeer, B., Lorkiewicz, J., Lu, H. H., Ludwig, F., Machau, K., Maciocha, W., Madec, C., Magueur, C., Maiano, C., Maksimova, I., Malcher, K., Maltezopoulos, T., Mamoshkina, E., Manschwetus, B., Marcellini, F., Marinkovic, G., Martinez, T., Martirosyan, H., Maschmann, W., Maslov, M., Matheisen, A., Mavric, U., Meißner, J., Meissner, K., Messerschmidt, M., Meyners, N., Michalski, G., Michelato, P., Mildner, N., Moe, M., Moglia, F., Mohr, C., Mohr, S., Möller, W., Mommerz, M., Monaco, L., Montiel, C., Moretti, M., Morozov, I., Morozov, P., Mross, D., Mueller, J., Müller, C., Müller, J., Müller, K., Munilla, J., Münnich, A., Muratov, V., Napoly, O., Näser, B., Nefedov, N., Neumann, R., Neumann, R., Ngada, N., Noelle, D., Obier, F., Okunev, I., Oliver, J. A., Omet, M., Oppelt, A., Ottmar, A., Oublad, M., Pagani, C., Paparella, R., Paramonov, V., Peitzmann, C., Penning, J., Perus, A., Peters, F., Petersen, B., Petrov, A., Petrov, I., Pfeiffer, S., Pflüger, J., Philipp, S., Pienaud, Y., Pierini, P., Pivovarov, S., Planas, M., Pławski, E., Pohl, M., Polinski, J., Popov, V., Prat, S., Prenting, J., Priebe, G., Pryschecki, H., Przygoda, K., Pyata, E., Racky, B., Rathjen, A., Ratuschni, W., Renaud-Camperdros, S., Rehlich, K., Reschke, D., Robson, C., Roever, J., Roggli, M., Rothenburg, J., Rusiński, E., Rybaniac, R., Sahling, H., Salmani, M., Samoylova, L., Sanzone, D., Saretzki, F., Sawlanski, O., Schaffran, J., Schlarb, H., Schlösser, M., Schlott, V., Schmidt, C., Schmidt-Foehre, F., Schmitz, M., Schmökel, M., Schnautz, T., Schneidmiller, E., Scholz, M., Schöneburg, B., Schultze, J., Schulz, C., Schwarz, A., Sekutowicz, J., Sellmann, D., Semenov, E., Serkez, S., Sertore, D., Shehzad, N., Shemarykin, P., Shi, L., Sienkiewicz, M., Sikora, D., Sikorski, M., Silenzi, A., Simon, C., Singer, W., Singer, X., Sinn, H., Sinram, K., Skvorodnev, N., Smirnov, P., Sommer, T., Sorokin, A., Stadler, M., Steckel, M., Steffen, B., Steinhilber-Kühl, N., Stephan, F., Stodulski, M., Stolper, M., Sulimov, A., Susen, R., Świerblewski, J., Sydlo, C., Syresin, E., Sytchev, V., Szuba, J., Tesch, N., Thie, J., Thiebault, A., Tiedtke, K., Tischhauser, D., Tolkiehn, J., Tomin, S., Tonisch, F., Toral, F., Torbin, I., Trapp, A., Treyer, D., Trowitzsch, G., Trublet, T., Tschentscher, T., Ullrich, F., Vannoni, M., Varela, P., Varghese, G., Vashchenko, G., Vasic, M., Vazquez-Velez, C., Verguet, A., Vilcins-Czvitkovits, S., Villanueva, R., Visentin, B., Viti, M., Vogel, E., Volobuev, E., Wagner, R., Walker, N., Wamsat, T., Weddig, H., Weichert, G., Weise, H., Wendorf, R., Werner, M., Wichmann, R., Wiebers, C., Wiencek, M., Wilksen, T., Will, I., Winkelmann, L., Winkowski, M., Wittenburg, K., Witzig, A., Wlk, P., Wohlenberg, T., Wojciechowski, M., Wolff-Fabris, F., Wrochna, G., Wrona, K., Yakopov, M., Yang, B., Yang, F., Yurkov, M., Zagorodnov, I., Zalden, P., Zavadtsev, A., Zavadtsev, D.,

- Zhirnov, A., Zhukov, A., Ziemann, V., Zolotov, A., Zolotukhina, N., Zummack, F. & Zybin, D. (2020). *Nat. Photon.* **14**, 391–397.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). *J. R. Stat. Soc. Ser. B*, **39**, 1–22.
- DePonte, D. P., Weierstall, U., Schmidt, K., Warner, J., Starodub, D., Spence, J. C. H. & Doak, R. B. (2008). *J. Phys. D Appl. Phys.* **41**, 195505.
- Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R. N. C., Seltzer, V., Clavier, 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.
- Emma, P., Akre, R., Arthur, J., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P., Coffee, R., Decker, F.-J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gilevich, S., Hastings, J., Hays, G., Hering, P., Huang, Z., Iverson, R., Loos, H., Messerschmidt, M., Miahnahri, A., Moeller, S., Nuhn, H.-D., Pile, G., Ratner, D., Rzepiela, J., Schultz, D., Smith, T., Stefan, P., Tompkins, H., Turner, J., Welch, J., White, W., Wu, J., Yocky, G. & Galayda, J. (2010). *Nat. Photon.* **4**, 641–647.
- Ferguson, K. R., Bucher, M., Bozek, J. D., Carron, S., Castagna, J.-C., Coffee, R., Curiel, G. I., Holmes, M., Krzywinski, J., Messerschmidt, M., Minitti, M., Mitra, A., Moeller, S., Noonan, P., Osipov, T., Schorb, S., Swiggers, M., Wallace, A., Yin, J. & Bostedt, C. (2015). *J. Synchrotron Rad.* **22**, 492–497.
- Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758.
- Fienup, J. R. (2013). *Appl. Opt.* **52**, 45.
- Gaffney, K. J. & Chapman, H. N. (2007). *Science*, **316**, 1444–1448.
- Gorobtsov, O. Y., Lorenz, U., Kabachnik, N. M. & Vartanyants, I. A. (2015). *Phys. Rev. E*, **91**, 062712.
- Hantke, M. F., Hasse, D., Maia, F. R. N. C., Ekeberg, T., John, K., Svenda, M., Loh, N. D., Martin, A. V., Timneanu, N., Larsson, D. S. D., van der Schot, G., Carlsson, G. H., Ingelman, M., Andreasson, J., Westphal, D., Liang, M., Stellato, F., DePonte, D. P., Hartmann, R., Kimmel, N., Kirian, R. A., Seibert, M. M., Mühlhig, K., Schorb, S., Ferguson, K., Bostedt, C., Carron, S., Bozek, J. D., Rolles, D., Rudenko, A., Epp, S., Chapman, H. N., Barty, A., Hajdu, J. & Andersson, I. (2014). *Nat. Photon.* **8**, 943–949.
- Harauz, G. & van Heel, M. (1986). *Optik*, **73**, 146–156.
- Heel, M. van & Schatz, M. (2005). *J. Struct. Biol.* **151**, 250–262.
- 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.
- Ishikawa, T., Aoyagi, H., Asaka, T., Asano, Y., Azumi, N., Bizen, T., Ego, H., Fukami, K., Fukui, T., Furukawa, Y., Goto, S., Hanaki, H., Hara, T., Hasegawa, T., Hatsui, T., Higashiya, A., Hirano, T., Hosoda, N., Ishii, M., Inagaki, T., Inubushi, Y., Itoga, T., Joti, Y., Kago, M., Kameshima, T., Kimura, H., Kiriha, Y., Kiyomichi, A., Kobayashi, T., Kondo, C., Kudo, T., Maesaka, H., Maréchal, X. M., Masuda, T., Matsubara, S., Matsumoto, T., Matsushita, T., Matsui, S., Nagasono, M., Nariyama, N., Ohashi, H., Ohata, T., Ohshima, T., Ono, S., Otake, Y., Saji, C., Sakurai, T., Sato, T., Sawada, K., Seike, T., Shirasawa, K., Sugimoto, T., Suzuki, S., Takahashi, S., Takebe, H., Takeshita, K., Tamasaku, K., Tanaka, H., Tanaka, R., Tanaka, T., Togashi, T., Togawa, K., Tokuhisa, A., Tomizawa, H., Tono, K., Wu, S., Yabashi, M., Yamaga, M., Yamashita, A., Yanagida, K., Zhang, C., Shintake, T., Kitamura, H. & Kumagai, N. (2012). *Nat. Photon.* **6**, 540–544.
- Khubbutdinov, R., Menushenkov, A. P. & Vartanyants, I. A. (2019). *J. Synchrotron Rad.* **26**, 1851–1862.
- 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. N. C., Munke, A., Nettelblad, C., Pande, K., Reddy, H. K. N., 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.
- Li, H., Aquila, A., Abbey, B., Alvarez, R., Ayyer, K., Barty, A., Berntsen, P., Bielecki, J., Bucher, M., Carini, G., Chapman, H. N., Contreras, A., Daurer, B. J., DeMirici, H., Flückiger, L., Frank, M., Hogue, B. G., Hosseinizadeh, A., Jönsson, H. O., Kirian, R. A., Kurta, R. P., Loh, D., Maia, F. R. N. C., Mancuso, A., McFadden, M., Muehlig, K., Munke, A., Narayana Reddy, H. K., Nazari, R., Ourmazd, A., Rose, M., Schwander, P., Seibert, M. M., Sellberg, J. A., Sierra, R. G., Sun, Z., Svenda, M., Vartanyants, I., Walter, P., Williams, G., Xavier, P. L. & Zaare, S. (2020). *Sci. Data*. (submitted) (see also CXIDB <https://dx.doi.org/10.11577/1645124>).
- Loh, N. D. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130328.
- Loh, N. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.
- Lorenz, U., Kabachnik, N. M., Weckert, E. & Vartanyants, I. A. (2012). *Phys. Rev. E*, **86**, 051911.
- 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., Mühlhig, 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., Gorkhove, T. & Maia, F. R. N. C. (2018). *IUCr*, **5**, 531–541.
- Mancuso, A. P., Aquila, A., Batchelor, L., Bean, R. J., Bielecki, J., Borchers, G., Doerner, K., Giewekemeyer, K., Graceffa, R., Kelsey, O. D., Kim, Y., Kirkwood, H. J., Legrand, A., Letrun, R., Manning, B., Lopez Morillo, L., Messerschmidt, M., Mills, G., Raabe, S., Reimers, N., Round, A., Sato, T., Schulz, J., Signe Takem, C., Sikorski, M., Stern, S., Thute, P., Vagović, P., Weinhausen, B. & Tschentscher, T. (2019). *J. Synchrotron Rad.* **26**, 660–676.
- Mäntynen, S., Sundberg, L.-R., Oksanen, H. M. & Poranen, M. M. (2019). *Viruses*, **11**, 76.
- 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., Ayyer, 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. N., Lan, T.-Y., Larsson, D. S. D., Liu, H., Loh, N. D., Maia, F. R. N. C., Mancuso, A. P., Mühlhig, 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.
- Nazari, R., Zaare, S., Alvarez, R., Karpos, K., Engelman, T., Madsen, C., Nelson, G., Spence, J. C. H., Weierstall, U., Adrian, R. J. & Kirian, R. A. (2020). *Opt. Express*, **28**, 21749–21765.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
- Osipov, T., Bostedt, C., Castagna, J.-C., Ferguson, K. R., Bucher, M., Montero, S. C., Swiggers, M. L., Obaid, R., Rolles, D., Rudenko, A., Bozek, J. D. & Berrah, N. (2018). *Rev. Sci. Instrum.* **89**, 035112.
- Peralta, B., Gil-Carton, D., Castañón-Díez, D., Bertin, A., Boulogne, C., Oksanen, H. M., Bamford, D. H. & Abrescia, N. G. A. (2013). *PLoS Biol.* **11**, e1001667.
- Reddy, H. K. N., Carroni, M., Hajdu, J. & Svenda, M. (2019). *eLife*, **8**, e48496.
- 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., 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.
- Santos-Pérez, I., Oksanen, H. M., Bamford, D. H., Goñi, F. M., Reguera, D. & Abrescia, N. G. A. (2017). *Biochim. Biophys. Acta*, **1861**, 664–672.
- Scheres, S. H. W., Valle, M., Nuñez, R., Sorzano, C. O. S., Marabini, R., Herman, G. T. & Carazo, J.-M. (2005). *J. Mol. Biol.* **348**, 139–149.
- Schoenlein, R., Abbamonte, P., Abild-Pedersen, F., Adams, P., Ahmed, M., Albert, F., Mori, R. A., Anfinrud, A., Aquila, A. & Armstrong, M. (2015). *New Science Opportunities Enabled By LCLS-II X-ray Lasers*. SLAC report, Menlo Park, California, USA, 1–189.
- 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.-T., Chapman, H. N., Andersson, I., Hajdu, J. & Ekeberg, T. (2015). *Nat. Commun.* **6**, 5704.
- Seibert, M. M., Ekeberg, T., Maia, F. R. N. C., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., Hantke, M., DePonte, D. P., Barty, A., Schulz, J., Gumprecht, L., Coppola, N., Aquila, A., Liang, M., White, T. A., Martin, A., Coleman, C., Stern, S., Abergel, C., Seltzer, V., Claverie, J.-M., Bostedt, C., Bozek, J. D., Boutet, S., Miahnahri, A. A., Messerschmidt, M., Krzywinski, J., Williams, G., Hodgson, K. O., Bogan, M. J., Hampton, C. Y., Sierra, R. G., Starodub, D., Andersson, I., Bajt, S., Barthelmess, M., Spence, J. C. H., Fromme, P., Weierstall, U., Kirian, R., Hunter, M., Doak, R. B., Marchesini, S., Hau-Riege, S. P., Frank, M., Shoeman, R. L., Lomb, L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Schmidt, C., Foucar, L., Kimmel, N., Holl, P., Rudek, B., Erk, B., Hömke, A., Reich, C., Pietschner, D., Weidenspointner, G., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Schlichting, I., Herrmann, S., Schaller, G., Schopper, F., Soltan, H., Kühnel, K.-U., Andrichke, R., Schröter, C.-D., Krasniqi, F., Bott, M., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., Chapman, H. N. & Hajdu, J. (2011). *Nature*, **470**, 78–81.
- Shi, Y., Yin, K., Tai, X., DeMirci, H., Hosseiniadeh, A., Hogue, B. G., Li, H., Ourmazd, A., Schwander, P., Vartanyants, I. A., Yoon, C. H., Aquila, A. & Liu, H. (2019). *IUCrJ*, **6**, 331–340.
- Sobolev, E., Zolotarev, S., Giewekemeyer, K., Bielecki, J., Okamoto, K., Reddy, H. K. N., Andreasson, J., Ayer, K., Barak, I., Bari, S., Barty, A., Bean, R., Bobkov, S., Chapman, H. N., Chojnowski, G., Daurer, B. J., Dörner, K., Ekeberg, T., Flückiger, L., Galzitskaya, O., Gelisio, L., Hauf, S., Hogue, B. G., Horke, D. A., Hosseiniadeh, A., Ilyin, V., Jung, C., Kim, C., Kim, Y., Kirian, R. A., Kirkwood, H., Kulyk, O., Küpper, J., Letrun, R., Loh, N. D., Lorenzen, K., Messerschmidt, M., Mühlig, K., Ourmazd, A., Raab, N., Rode, A. V., Rose, M., Round, A., Sato, T., Schubert, R., Schwander, P., Sellberg, J. A., Sikorski, M., Silenzi, A., Song, C., Spence, J. C. H., Stern, S., Sztuk-Dambietz, J., Teslyuk, A., Timneanu, N., Trebbin, M., Uetrecht, C., Weinhausen, B., Williams, G. J., Xavier, P. L., Xu, C., Vartanyants, I. A., Lamzin, V. S., Mancuso, A. & Maia, F. R. N. C. (2020). *Commun. Phys.* **3**, 97.
- Strüder, L., Epp, S., Rolles, D., Hartmann, R., Holl, P., Lutz, G., Soltan, H., Eckart, R., Reich, C., Heinzinger, K., Thamm, C., Rudenko, A., Krasniqi, F., Kühnel, K.-U., Bauer, C., Schröter, C.-D., Moshhammer, R., Techert, S., Miessner, D., Porro, M., Häcker, O., Meidinger, N., Kimmel, N., Andrichke, R., Schopper, F., Weidenspointner, G., Ziegler, A., Pietschner, D., Herrmann, S., Pietsch, U., Walenta, A., Leitenberger, W., Bostedt, C., Möller, T., Rupp, D., Adolph, M., Graafsma, H., Hirsemann, H., Gärtner, K., Richter, R., Foucar, L., Shoeman, R. L., Schlichting, I. & Ullrich, J. (2010). *Nucl. Instrum. Methods Phys. Res. A*, **614**, 483–496.
- Thibault, P. & Menzel, A. (2013). *Nature*, **494**, 68–71.
- Vartanyants, I. A. & Singer, A. (2010). *New J. Phys.* **12**, 035004.
- Weierstall, U., Spence, J. C. H. & Doak, R. B. (2012). *Rev. Sci. Instrum.* **83**, 035108.