

# SCIENTIFIC DATA

OPEN

## Data Descriptor: Coherent soft X-ray diffraction imaging of coliphage PR772 at the Linac coherent light source

Received: 9 January 2017

Accepted: 27 April 2017

Published: 27 June 2017

Hemanth K.N. Reddy *et al.*<sup>#</sup>

Single-particle diffraction from X-ray Free Electron Lasers offers the potential for molecular structure determination without the need for crystallization. In an effort to further develop the technique, we present a dataset of coherent soft X-ray diffraction images of Coliphage PR772 virus, collected at the Atomic Molecular Optics (AMO) beamline with pnCCD detectors in the LAMP instrument at the Linac Coherent Light Source. The diameter of PR772 ranges from 65–70 nm, which is considerably smaller than the previously reported ~600 nm diameter Mimivirus. This reflects continued progress in XFEL-based single-particle imaging towards the single molecular imaging regime. The data set contains significantly more single particle hits than collected in previous experiments, enabling the development of improved statistical analysis, reconstruction algorithms, and quantitative metrics to determine resolution and self-consistency.

Design Type(s)	virus particle imaging objective
Measurement Type(s)	X-ray diffraction
Technology Type(s)	X-ray free electron laser
Factor Type(s)	
Sample Characteristic(s)	Enterobacteria phage PR772

Correspondence and requests for materials should be addressed to A.A. (email: Aquila@slac.stanford.edu) or to J.B. (email: johanb@xray.bmc.uu.se).

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

## Background & Summary

Theoretical studies predict X-ray Free Electron Lasers (XFELs) can potentially image biomolecules and hence determine their structures without crystallization<sup>1</sup>. To realize this in practice has proved a significant experimental challenge. The Single Particle Imaging (SPI) Initiative<sup>2</sup> was formed as a collaborative effort to identify and solve the experimental challenges to achieving high-resolution imaging of single molecules with X-rays.

Coliphage PR772 is a virus of approximately 70 nm in diameter, which infects *Escherichia coli* (*E. coli*). It was selected as the sample for this experiment due to its high structural homogeneity, uniform size distribution, suitable particle concentration in solution, having a known structure, and the ability to be aerosolized by Gas Dynamic Virtual Nozzle (GDVN)<sup>3</sup> for injection into the XFEL beam using an aerosol injector. Having a known structure of PR772 (unpublished) enables validation of any subsequent data analysis steps.

For the data presented here, Coliphage PR772 was aerosolized and delivered into the focus of the LAMP instrument at the Atomic Molecular Optics (AMO) beamline<sup>4</sup> of the Linac Coherent Light Source (LCLS) X-ray laser<sup>5,6</sup>. The data include clear diffraction snapshots from single PR772 virus particles.

## Methods

### Sample preparation

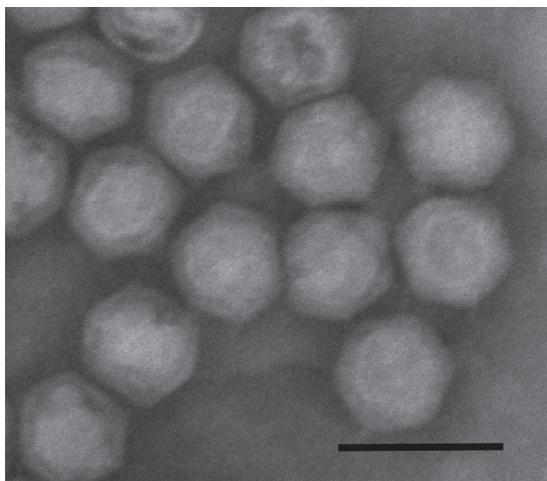
Bacteriophage PR772 (ATCC BAA-769-B1) infects *E. coli* K12 J53-1 (ATCC BAA-769). PR772 was cultured on agar by the overlay method. Tryptic Soy Agar (Difco Tryptic Soy Agar) was prepared and poured into petri dishes to form a hard agar support layer. The overlay medium was soft agar prepared with Tryptic Soy Broth (Bacto Tryptic Soy Broth) along with 0.5% Agar-Agar (Microbiology, Merck). *E. coli* was cultured in Tryptic Soy Broth to reach an OD<sub>600</sub> of 0.7. Soft agar was melted and cooled to 45 °C. *E. coli* and viral stock (10<sup>8</sup> pfu ml<sup>-1</sup>) were added to the soft agar in a volume ratio of 10:1 and mixed well. The mixture was immediately poured onto the hard agar in the petri dishes and incubated overnight at 37 °C.

The soft agar layer from the overnight culture was scraped off and collected in a sterile container, diluted with 100 ml of sterile storage buffer (TRIS 50 mM, NaCl 100 mM, MgSO<sub>4</sub> 1 mM, EDTA 1 mM, pH 8.0) and mixed overnight at 4 °C. The mixture was centrifuged at 8,000 g for 30 min to remove agar and cell debris. The supernatant was collected and filtered through a 0.45 micron filter (Filtropur S 0.45, Sarstedt). The viral particles were separated from the permeate solution by PEG precipitation (9% w/v PEG 8000 and 5.8% w/v NaCl). After mixing overnight at 4 °C, the precipitate was centrifuged for 90 min at 8,000 g. The supernatant was discarded and the viral pellet was suspended in 1 ml storage buffer. The viral suspension was then applied to a Capto-Q anion exchange column. The sample was eluted by varying the concentration of NaCl (100 mM–1.5 M). The fractions representing the elution profile peak was collected and observed with a electron microscopy (EM) (Quanta FEG 650, FEI) to confirm the presence of intact viral particles. (Figure 1)

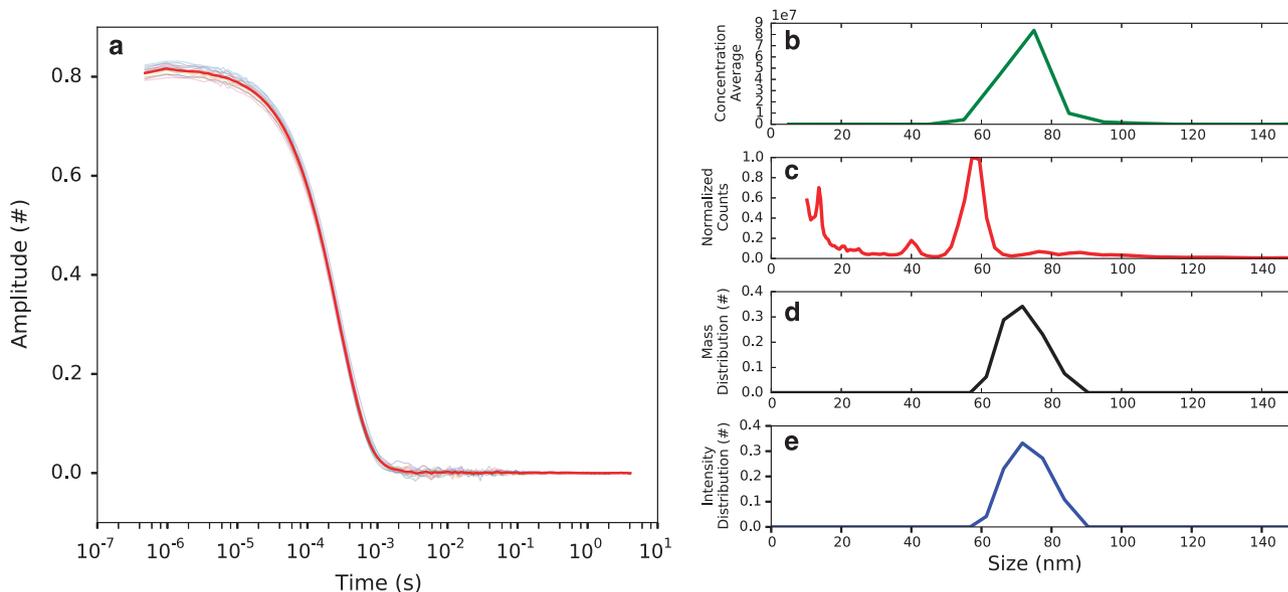
Before sample injection, the PR772 was transferred from storage buffer into volatile ammonium acetate buffer (250 mM, pH 7.5) using PD10 desalting columns (GE Healthcare). Hence, most assays were performed with sample in ammonium acetate buffer instead of storage buffer.

### Sample characterization

**Infectivity.** Virus titers were measured by plaque assay following purification purification<sup>7</sup>. Serial 10-fold dilutions of purified virus were plated on a mat of *E. coli* and incubated overnight at 37 °C.



**Figure 1.** Electron micrograph of purified bacteriophage PR772. Scale bar represents 100 nm.



**Figure 2. Sample characterization of PR772.** NTA measurements (b) of bacteriophage PR772 in ammonium acetate buffer (250 mM, pH 7.5). The curve shows the distribution of the different sized particles in the solution. We observed a single large peak at 75 nm with concentration of  $10^8$  particles/ml, which is similar to the dilution used for measurement. The diameter peak from the DMA measurement (c) was approximately 60 nm. The actual gas flow during the DMA measurement was lower than the set value, which resulted in a diameter low by about 10%. The gas flow was disregarded since the purpose of the DMA measurement was to assess sample purity and not size. DLS measurements of bacteriophage PR772 in ammonium acetate buffer (250 mM, pH 7.5). Curve in red shows the average of 18 correlation data collected over time for the sample (a). Curves in black and blue show the size of the particles based on mass (d) and intensity (e) distribution respectively. The mean diameter of the particle was 75.99 nm with a polydispersity index (PDI) of 0.012.

Plaque forming units per ml (pfu ml<sup>-1</sup>) were calculated using the formula: average # plaques/volume plated x dilution.

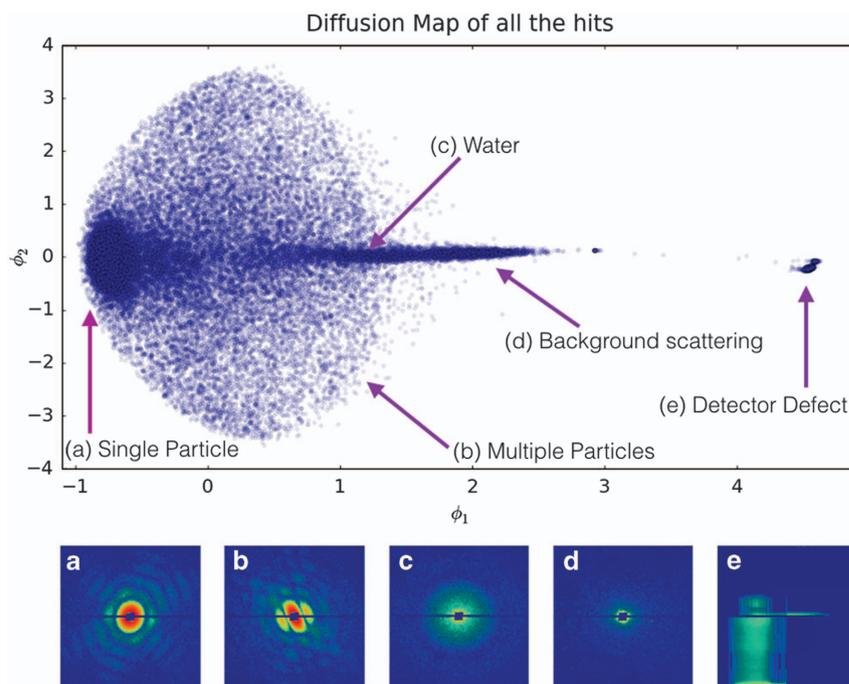
**Liquid phase.** The size and monodispersity of the PR772 in ammonium acetate buffer were measured using Nanoparticle Tracking Analysis (NTA) (NanoSight LM10, Malvern Instruments Ltd.) (Figure 2) and Dynamic Light Scattering (DLS) (w130i, AvidNano Ltd.) (Figure 2). The sample was diluted to the required concentrations for measurement. For NTA  $10^8$  particles per ml were used to limit the number of tracks to 200. For DLS about  $10^{10}$  particles per ml was used to reach a recommended counts per s.

**Gaseous phase.** The size distribution in the gas phase was measured using Electrophoretic DMA (Figure 2). PR772 was aerosolized with a nano-electrospray ionization (ESI) source (TSI model 3480) and passed through an electrostatic classifier (TSI model 3480) to continuously classify particles from 10 to 500 nm. Classified particles were counted with a condensation particle counter (CPC, TSI model 3786).

**Injection testing.** Injection tests on PR772 were performed with a setup similar to the one subsequently used at the LCLS experiment to investigate the aerosolization characteristics of the samples and to study sample behavior during the injection procedure<sup>8</sup>. A glass microscope slide covered by a transparent sticky gel piece (GelPak) was positioned beneath the exit point of the aerodynamic lens (at a position similar to the interaction region with the X-ray beam at LCLS). Particle dusting spots were observed through an objective lens mounted below the glass slide. The focused particles from the injector were also collected on formvar/carbon grids (#01754F, F/C 400 mesh Cu, Ted Pella Inc.). These samples were examined by EM without any further modification.

#### Sample injection at LCLS

Samples were delivered into the X-ray beam using the aerodynamic lens stack system used in previous experiments<sup>9</sup>. Purified PR772 was transferred to a volatile ammonium acetate buffer at a concentration of  $10^{12}$  particles ml<sup>-1</sup> and introduced to the injector via a GDVN<sup>3</sup> at a flow rate of 1–2  $\mu$ l min<sup>-1</sup>. The aerosol continued through a gas skimmer and a relaxation chamber, forming a fine beam of focused particles by the aerodynamic lens stack<sup>10</sup>. The focus of the particle beam and multiplicity



**Figure 3.** Diffusion map analysis of all the hits reduced to two dominant dimensions  $\Phi_1$  and  $\Phi_2$  as described in the text. Clustering helps identify the single particle diffraction patterns in the dataset. In addition to single hits (a) and multiple hits (b), the spherical particles (c) classified as ‘water’ may also contain contaminant residue from the sample and buffer. Additionally slightly larger than average background scattering (d) and frames with a defective readout (e) are also shown in the diffusion map.

of the particles could be optimized by regulating the flow of sample and gas along with the skimmer pressure.

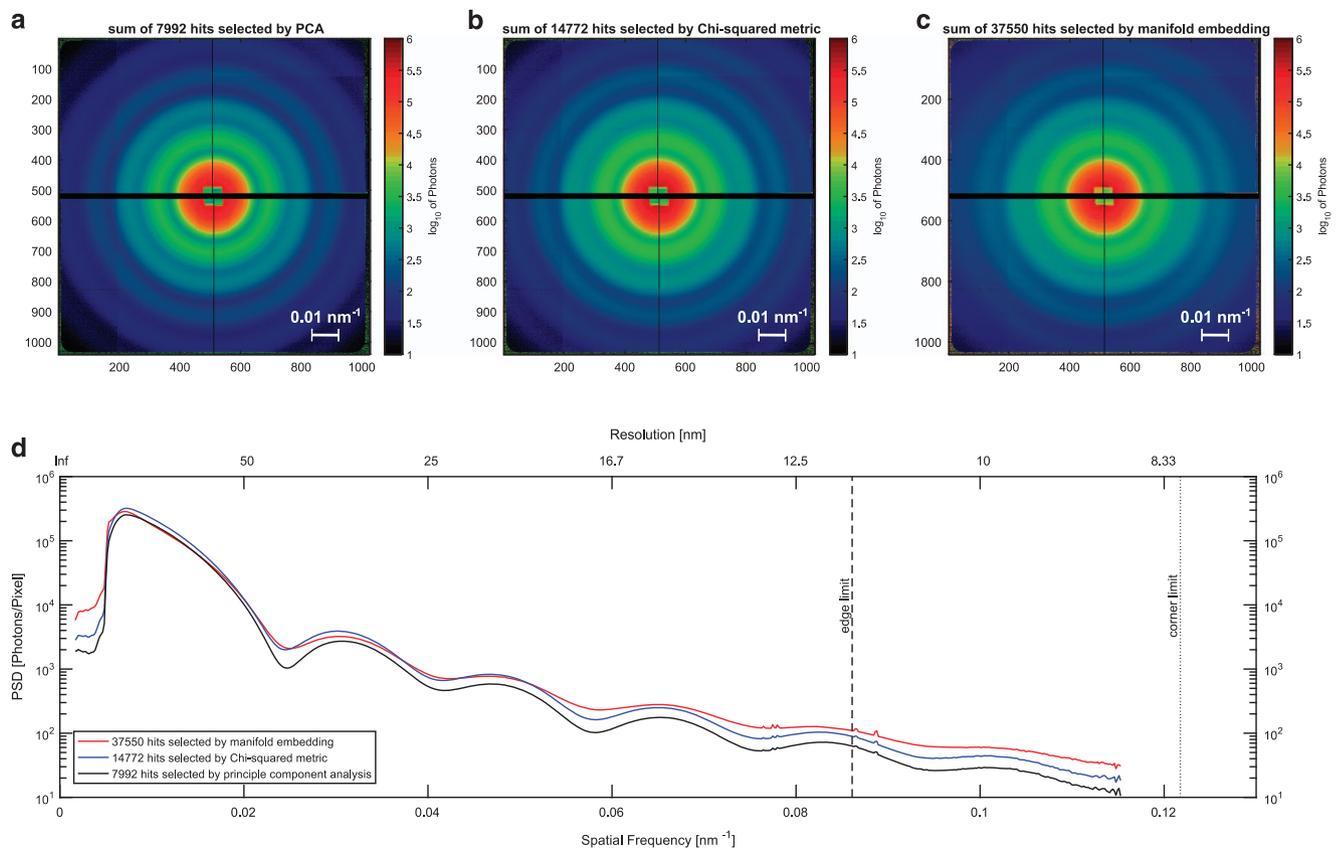
### Experimental setup and data collection

Coherent diffraction snapshots using single soft X-ray pulses were recorded at the AMO beamline of the LCLS XFEL using the LAMP endstation<sup>4</sup>. The configuration was similar to that used in previous coherent diffraction experiments<sup>11</sup>, but with a shorter distance from the sample interaction region to the detector due to the gate valve between detector and sample chambers being removed for the present experiment.

Measurements were performed using LCLS tuned to a photon energy of 1.6 keV delivering 4 mJ into a 70 fs duration pulse at the end of the undulators. This was focused into a nominal  $1.5 \mu\text{m}^2$  FWHM region using a pair of Kirkpatrick-Baez (KB) mirrors, giving a nominal power density of  $\sim 3.8 \cdot 10^{18} \text{ W cm}^{-2}$  or  $10^{13}$  photons per pulse assuming no beamline losses. Comparison of measured and calculated from known scattering objects indicate that the actual power density may differ from this estimate by a factor of 10. This is due to three combined effect of overfilling of the focusing optics, carbon contamination on the KB mirrors increasing mirror roughness (reducing reflectivity), and a less-than-perfect focal spot size due to carbon contamination distorting the KB mirrors (increasing effective focal spot size).

The photon energy was selected by considering the highest achievable resolution, while remaining below the silicon absorption edge at 1.8 keV. At photon energies above the silicon K-edge of 1.8 keV, the silicon beam conditioning apertures’ fluorescence increases the background. A series of silicon apertures, including a post-sample aperture, were used to limit the amount of background scatter incident on the detectors from the beamline. Three  $1 \text{ mm} \times 1 \text{ mm}$   $\text{Si}_3\text{N}_4$  apertures were used to reduce low  $q$  scatter. The first two apertures were positioned laterally to the beam to form a small rectangle, with the third aperture used to clean up diffraction from the first two. The first two apertures were separated by 5 cm, with the third aperture separated by 10 cm. Additionally the third aperture was 10 cm away from the focus. A large round aluminum post sample aperture was used to block high  $q$  scatter. It was located 2.4 cm downstream of the sample with a diameter of 4 cm.

Diffraction snapshots were recorded using two pairs of pnCCD detectors<sup>12</sup> at the LCLS pulse repetition rate of 120 Hz. One pair of pnCCD detectors was located 10 cm downstream of the interaction region, and a second pair of detectors further downstream at 58.1 cm from the interaction region. The pixel size are  $75 \mu\text{m}$  and each detector in the pair has  $512 \times 1,024$  pixels giving the full set-up a pixel count of two planes each with  $1,024 \times 1,024$  pixels. The experimental configuration was similar to that



**Figure 4. Comparison of single hit selections by principal component analysis (PCA), Chi-squared metric and manifold embedding.** Summed diffraction patterns of (a) PCA, (b) Chi-squared metric and (c) manifold embedding. (d) Averaged signal over angle (power spectral density, PSD) as a function of spatial frequency and resolution. Five diffraction maxima are fully captured within the detector edges. The sixth maximum is captured within the detector corners.

described in ref. 13, albeit with a higher photon energy (1.6 keV instead of 1.2 keV), and the use of a different beam conditioning configuration and the front pnCCD. The majority of the data processing and evaluation of the quality of the data was conducted on the downstream detector. The resolution of the downstream detector is 11.6 nm at the edge or 8.3 nm in the corner giving an oversampling of the diffraction patterns of >10.

The data were analyzed online using *Hummingbird*<sup>14</sup> and *psana*<sup>15</sup>.

### Data processing

Data saved in the native XTC format used at the LCLS were analyzed and converted to the CXI file format using the LCLS data analysis framework *psana*<sup>15</sup>. The raw pnCCD pixels contain analog to digital units (ADUs), to which various corrections must be made in order to obtain photon counts. As a first step, dark calibration and row-by-row common mode correction were performed on the pnCCD detector images by the LCLS software environment, *psana*. Data was calibrated using *psana*'s *ImgAlgos.NDArrCalib* module, with pedestal subtraction (*do\_peds*), common-mode correction (*do\_cmod*), statistical correction (*do\_stat*) and gain corrections (*do\_gain*) turned on. A flat-field data set using silicon K edge fluorescence (1.7 keV) was used to calibrate the gain on a pixel basis. These gain-corrected ADU values were then thresholded based on an average of 128 ADUs/photon in order to obtain photon counts.

Not all detector frames (or events) contained diffraction from single particles. Since the average number of particles per X-ray focus volume is much less than one, most of the pulses do not hit any particles. Hits (events in which a particle was intercepted by the X-ray beam) were identified using a chi squared metric adapted from ref. 16.

$$\chi_j^2 = \frac{\overline{(I_j - B_j)^2}}{\text{Var}(B_j)}$$

Chi squared for the *j*th image  $I_j$  was calculated by subtracting a running median background  $B$  and normalizing by the variance of  $B$ . The chi squared value was calculated within an annular area of radii

Run Number	Number of frames in XTC file	Number of hits by diffusion map embedding	Number of hits by principal component analysis	Number of hits by manifold embedding
182	97,733	15	2	0
183	169,227	3	0	0
184	136,624	1	0	0
185	77,079	7	4	0
186	453,285	2,048	1,121	4,759
188	322,097	1,171	599	2,885
190	265,585	1,498	838	4,175
191	293,284	1,718	944	4,427
192	224,583	1,279	656	3,209
193	329,049	2,525	1,370	6,781
194	214,891	1,661	897	4,555
196	204,236	1,436	818	3,792
197	188,895	1,410	743	2,967
Total	2,976,568	14,772	7,992	37,550

**Table 1. Single particle hits data table.** This table describes the estimated single particle PR772 hits, per raw XTC file, using three different classification methods. Additionally it provided the number of data frames per run, raw XTC file, deposited in the CXIDB.

150 and 400 pixels from beam center and hits were defined to frames with Chi-squared above 10. Hits were saved to file with detector corrections applied (i.e., pedestal, common mode and gain corrected) and then down-sampled by a factor of 4.

### Data Records

The data are deposited in the Coherent X-ray Imaging Data Bank (CXIDB)<sup>20</sup> in the CXIDB data format, which is based on the HDF5 format (Data Citation 1). Convenient functions for accessing the CXIDB data file exist in the libspimage package for C and Python (Maia 2010), as well as many computing environments, including Python using the *h5py* module and MATLAB using e.g., the *h5read* function. The Owl software is convenient for visualizing data in the CXIDB format. In addition to the CXI file, the conversion script (*create\_dataset.py*) and additional metadata files (*selection.h5*, *psana.cfg*) are provided along with usage instructions. Detector panel calibration files mapping data to real space are also provided. Configuration files for *Hummingbird* and *psana* are provided for completeness of describing processing performed on the deposited data.

### Technical Validation

Single-particle diffraction patterns were identified by analyzing all the hits in a reduced set of dimensions using diffusion map embedding similar to the method described in ref. 17 shown in Figure 3. The normalized graph Laplacian is used to calculate the likelihood of diffusion from the center of the single particle cluster at  $\Phi_1, \Phi_2, \Phi_3 = -0.75, 0, 0$ . A likelihood value of 0.725 or higher is considered inside the single particle cluster. A total of 12,678 out of 14,772 images were identified as single particle hits. An alternative manifold-based data analytical approach yielded a larger dataset consisting of 37,550 single-particle snapshots, whose indices are provided. This approach reveals and corrects intensity variations and a range of other imaging artifacts<sup>18</sup>. Another method for data sorting and classification is based on principal component analysis (PCA)<sup>19</sup>. With this approach we have filtered 21,733 images as hits by intensity thresholding (all images above 200k photons). A total of 7,992 images were identified by the PCA technique as single hits and removed outliers by radial intensity filtering (also provided). A comparison of the summation of the single-particle hits, forming a pseudo small angle X-ray scattering patterns, is shown in Figure 4 and is shown in Table 1 per XTC run number.

Manual selection was not employed due to the large number of hits and the potential for user bias. Because identification of single-particle hits is dependent on the input parameters for clustering, we have deposited all data frames and identified potential single particle hits as a list of events. This enables the testing and comparison of different hit sorting algorithms.

Raw XTC files are included in the data deposition for anyone wishing to repeat the analysis from scratch.

### References

- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
- Aquila, A. *et al.* The linac coherent light source single particle imaging road map. *Struct. Dyn* **2**, 041701 (2015).

3. DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D: Appl. Phys* **41**, 195505 (2008).
4. Ferguson, K. R. *et al.* The Atomic, Molecular and Optical Science instrument at the Linac Coherent Light Source. *J. Synchrotron Rad* **22**, 492–497 (2015).
5. Emma, P. *et al.* First lasing and operation of an ångström-wavelength free-electron laser. *Nature Photonics* **4**, 641–647 (2010).
6. White, W. E., Robert, A. & Dunne, M. The Linac Coherent Light Source. *J. Synchrotron Rad* **22**, 472–476 (2015).
7. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of bacteriophages using the small drop plaque assay system. *Methods in molecular biology* **501**, 81–85 (2009).
8. Hantke, M. F. *et al.* High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nature Photonics* **8**, 943–949 (2014).
9. Munke, A. *et al.* Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. *Scientific Data* **3**, 160064 (2016).
10. Bogan, M. J. *et al.* Single particle X-ray diffractive imaging. *Nano Lett* **8**, 310–316 (2008).
11. Seibert, M. *et al.* Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* **470**, 78–81 (2011).
12. Strüder, L. *et al.* Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multi-purpose chamber for experiments at 4th generation light sources. *Nuclear Instruments and Methods in Physics Research Section A* **614**, 483–496 (2010).
13. Ekeberg, T. *et al.* Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. *Phys. Rev. Lett* **114**, 098102 (2015).
14. Daurer, B. J., Hantke, M. F., Nettelblad, C. & Maia, F. R. N. C. Hummingbird: monitoring and analyzing flash X-ray imaging experiments in real time. *J. Appl. Cryst* **49**, 1042–1047 (2016).
15. Damiani, D. *et al.* Linac Coherent Light Source data analysis using psana. *J. Appl. Cryst* **49**, 672–679 (2016).
16. Loh, N. D. Effects of extraneous noise in Cryptomography. *Proc. SPIE* Vol. 8500, 85000K (Image Reconstruction from Incomplete Data VII, 2012).
17. Yoon, C. Novel algorithms in coherent diffraction imaging using X-ray free-electron lasers. *Proc. SPIE* Vol. 8500, 85000H (Image Reconstruction from Incomplete Data VII, 2012).
18. Hosseinzadeh, A., Dashti, A., Schwander, P., Fung, R. & Ourmazd, A. Single-particle structure determination by X-ray free-electron lasers: possibilities and challenges. *Structural Dynamics* **2**, 041601 (2015).
19. Bobkov, S. A. *et al.* Sorting algorithms for single-particle imaging experiments at X-ray free-electron lasers. *J. Synchrotron Rad* **22**, 1345 (2015).
20. Maia, F. R. N. C. The Coherent X-ray Imaging Data Bank. *Nature Methods* **9**, 854–855 (2012).

## Data Citation

1. Yoon, C. H. *Coherent X-ray Imaging Data Bank* <http://dx.doi.org/10.11577/1349664> (2016).

## Acknowledgements

Portions of this research were carried out at the LCLS at the SLAC National Accelerator Laboratory. The LCLS is supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (OBES), under contract DE-AC02-76SF00515. This work is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Division of Chemical, Geological, and Biological Sciences, under Contract No. DE-AC02-06CH11357. This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-SC0012704. This work was supported in part by the U.S. National Science Foundation Award No. STC 1231306 (data sorting), the U.S. Department of Energy, Office of Science, Basic Energy Sciences Award No. DE-SC0002164 (algorithm development), and the U.S. National Science Foundation Award No. MCB-1240590 (artifact correction methods). Part of the work was supported by the National Science Foundation Science and Technology Center (NSF-STC) BioXFEL award: NSF-1231306. This work was supported in part by the NIH grant GM095583. Furthermore support is acknowledged from the Biodesign Center for Applied Structural Discovery at Arizona State University. This work is supported by the OBES through the AMOS program within the CSGB and of the DOE through the SLAC Laboratory Directed Research and Development Program and also support from the joint Stanford ChEM-H and SLAC National Accelerator Laboratory seed grant program. This work was supported by the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the European Research Council, the Röntgen Ångström Cluster and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT). R.A.K. acknowledges support from NSF STC Award (1231306). This work is supported by National Research Foundation of Korea (grant NRF-2015R1A5A1009962, NRF-2016R1A2B3010980, NRF-2016R1A6B2A02005471). Helmholtz Association through program-oriented funds The DFG Gottfried Wilhelm Leibniz Program This work was supported by Russian Ministry of Education and Science under Grant No.14.616.21.0003 (ID: RFMEFI61614X0003). European Research Council, ‘Frontiers in Attosecond X-ray Science: Imaging and Spectroscopy (AXSIS)’, ERC-2013-SyG 609920 (2014–2018). This work was supported by DESY. This work was supported by the Partnership of Innovation, Education and Research (PIER) Hamburg, a collaboration between DESY and Universität Hamburg. Financial Support came from the Joachim Herz Foundation Hamburg. National University of Singapore internal grant, and the Lee Kuan Yew fellowship. Higher Education and the Helge Ax:son Johnsons Stiftelse, ELI Extreme Light Infrastructure and ELIBIO (CZ.02.1.01/0.0/0.0/15\_003/0000447) from the European Regional Development Fund. This work was supported by the Australian Research Council Centre of Excellence in Advanced Molecular Imaging (CE140100011) [www.imagingcoe.org](http://www.imagingcoe.org).

## Author Contributions

A.A., A.B., A.Ma., A.Mu., B.H., C.S., D.N., G.C., G.V.D.S., H.D., J.A.S., K.A., M.B., M.F.H., M.Se., M.Sv., N.D.L., P.B., R.A.K., S.A., and Y.K.: Instrument alignment and data collection. A.B., A.O., C.H.Y., C.N., G.V.D.S., I.A.V., J.A.S., K.A., M.F.H., M.R., N.D.L., P.S., R.K., and S.B.: Data analysis. J.B., J.C.H.S., J.H., K.M., M.Sv., and P.B.: Injector Development. A.Mu., B.H., K.M., M.Sv., P.F., and P.L.X.: Sample characterization & Preparation. D.L., H.C., and M.B.: Instrumentation Development. C.H.Y., G.C., P.H., and S.C.: Detector Operations and development. A.B., A.O., B.D., C.H.Y., C.N., F.R.N.C.M., G.V.D.S., K.A., M.F.H., N.D.L., P.S.: Software development.

## Additional Information

**Competing interests:** The authors declare no competing financial interests.

**How to cite this article:** Reddy, H. K. N. *et al.* Coherent soft X-ray diffraction imaging of coliphage PR772 at the Linac coherent light source. *Sci. Data* 4:170079 doi: 10.1038/sdata.2017.79 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2017

Hemanth K.N. Reddy<sup>1</sup>, Chun Hong Yoon<sup>2</sup>, Andrew Aquila<sup>2</sup>, Salah Awel<sup>3,4</sup>, Kartik Ayer<sup>3</sup>, Anton Barty<sup>3</sup>, Peter Berntsen<sup>5</sup>, Johan Bielecki<sup>1</sup>, Sergey Bobkov<sup>6</sup>, Maximilian Bucher<sup>2,7,8</sup>, Gabriella A. Carini<sup>2,10</sup>, Sebastian Carron<sup>2,†</sup>, Henry Chapman<sup>3,4,9</sup>, Benedikt Daurer<sup>1</sup>, Hasan DeMirci<sup>2,10</sup>, Tomas Ekeberg<sup>1</sup>, Petra Fromme<sup>11,12</sup>, Janos Hajdu<sup>1,13</sup>, Max Felix Hanke<sup>1</sup>, Philip Hart<sup>2</sup>, Brenda G. Hogue<sup>11,14,15</sup>, Ahmad Hosseinizadeh<sup>16</sup>, Yoonhee Kim<sup>17</sup>, Richard A. Kirian<sup>18</sup>, Ruslan P. Kurta<sup>19</sup>, Daniel S.D. Larsson<sup>1</sup>, N. Duane Loh<sup>20,21</sup>, Filipe R.N.C. Maia<sup>1</sup>, Adrian P. Mancuso<sup>19</sup>, Kerstin Mühlig<sup>1</sup>, Anna Munke<sup>1</sup>, Daewoong Nam<sup>17</sup>, Carl Nettelblad<sup>1,22</sup>, Abbas Ourmazd<sup>16</sup>, Max Rose<sup>23</sup>, Peter Schwander<sup>16</sup>, Marvin Seibert<sup>1</sup>, Jonas A. Sellberg<sup>1,24</sup>, Changyong Song<sup>17</sup>, John C.H. Spence<sup>18</sup>, Martin Svenda<sup>1</sup>, Gijs Van der Schot<sup>1</sup>, Ivan A. Vartanyants<sup>23,25</sup>, Garth J. Williams<sup>26</sup> & P. Lourdu Xavier<sup>3,9,27</sup>

<sup>1</sup>Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), Uppsala SE-75124, Sweden. <sup>2</sup>SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>3</sup>Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. <sup>4</sup>Centre for Ultrafast Imaging, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. <sup>5</sup>Australian Research Council Centre of Excellence in Advanced Molecular Imaging, La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086, Australia. <sup>6</sup>National Research Centre 'Kurchatov Institute', Akademika Kurchatova pl. 1, 123182 Moscow, Russia. <sup>7</sup>Argonne National Laboratory, 9700 S. Cass Ave, Lemont, Illinois 60439, USA. <sup>8</sup>Institut für Optik und Atomare Physik, Technische Universität Berlin, Hardenbergstraße 36, 10623 Berlin, Germany. <sup>9</sup>Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. <sup>10</sup>Stanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>11</sup>BioDesign Center for Applied Structural Discovery, BioDesign Institute at Arizona State University, Tempe, Arizona 85287-5001, USA. <sup>12</sup>School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287-1604, USA. <sup>13</sup>Institute of Physics AS CR, v.v.i., Na Slovance 2, 18221 Prague 8, Czech Republic. <sup>14</sup>BioDesign Center for Immunotherapy, Vaccines, and Virotherapy, BioDesign Institute at Arizona State University, Tempe, Arizona 85287, USA. <sup>15</sup>Arizona State University, School of Life Sciences (SOLS), Tempe, Arizona 85287-5401, USA. <sup>16</sup>Department of Physics, University of Wisconsin Milwaukee, KIRC, 3135 N. Maryland Ave, Milwaukee, Wisconsin 53211, USA. <sup>17</sup>Department of Physics, Pohang University of Science and Technology, Pohang 37673,

Korea. <sup>18</sup>Department of Physics, Arizona State University, Tempe, Arizona 85287, USA. <sup>19</sup>European X-ray Free Electron Laser GmbH, Holzkoppel 4, 22869 Schenefeld, Germany. <sup>20</sup>Department of Physics, National University of Singapore Blk S12, 2 Science Drive 3, Singapore 117551. <sup>21</sup>Centre for Bio-imaging Sciences, National University of Singapore Blk S1A, 14 Science Drive 4, Singapore 117557. <sup>22</sup>Division of Scientific Computing, Department of Information Technology, Science for Life Laboratory, Uppsala University, Husargatan 3 (Box 337), Uppsala SE-75105, Sweden. <sup>23</sup>Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, D-22607 Hamburg, Germany. <sup>24</sup>Biomedical and X-Ray Physics, Department of Applied Physics, AlbaNova University Center, KTH Royal Institute of Technology, Stockholm SE-106 91, Sweden. <sup>25</sup>National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe shosse 31, 115409 Moscow, Russia. <sup>26</sup>NSLS-II, Brookhaven National Laboratory, Upton, New York 11873, USA. <sup>27</sup>Max-Planck Institute for the Structure and Dynamics of Matter, 22607 Hamburg, Germany. †Present address: California Lutheran University, 60 Olsen Road, Thousand Oaks, California 9360, USA.