

# SCIENTIFIC DATA

**OPEN**

**SUBJECT CATEGORIES**

- » Structural biology
- » Imaging
- » Single-molecule biophysics
- » Biological physics

## Data Descriptor: Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source

Anna Munke *et al.*<sup>#</sup>

Single particle diffractive imaging data from Rice Dwarf Virus (RDV) were recorded using the Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). RDV was chosen as it is a well-characterized model system, useful for proof-of-principle experiments, system optimization and algorithm development. RDV, an icosahedral virus of about 70 nm in diameter, was aerosolized and injected into the approximately 0.1 μm diameter focused hard X-ray beam at the CXI instrument of LCLS. Diffraction patterns from RDV with signal to 5.9 Ångström were recorded. The diffraction data are available through the Coherent X-ray Imaging Data Bank (CXIDB) as a resource for algorithm development, the contents of which are described here.

Received: 18 February 2016

Accepted: 22 June 2016

Published: 1 August 2016

<b>Design Type(s)</b>	virus particle imaging objective
<b>Measurement Type(s)</b>	physicochemical characterization • X-ray diffraction data
<b>Technology Type(s)</b>	X-ray free electron laser • dynamic light scattering • electrospray ionisation gas-phase electrophoretic mobility molecular analysis • particle-tracking velocimetry
<b>Factor Type(s)</b>	
<b>Sample Characteristic(s)</b>	Rice dwarf virus (isolate O)

Correspondence and requests for materials should be addressed to A.A. (email: aquila@slac.stanford.edu) or to M.M.S. (email: marvin.seibert@icm.uu.se).

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

## Background & Summary

For several decades, X-ray crystallography has been the dominant technique to solve the three-dimensional (3D) structure of biological macromolecules at atomic resolution. Structures of proteins, protein complexes and the machinery of entire biological reaction pathways have been elucidated, leading to numerous breakthroughs in our understanding of molecular architecture and function. However, not every protein complex crystallizes, a necessary condition for investigation using these methods. Radiation damage additionally limits the resolution for biological objects, both crystalline and non-crystalline, which for high exposures leads to the determination of structures representative of a photodamaged state<sup>1–3</sup>. The ultrashort and extremely bright pulses from X-ray free electron lasers (XFELs) were predicted to outrun radiation damage processes and allow the recording of diffraction data from samples prior to any significant motion of the nuclei occurring<sup>2</sup>. This has been experimentally demonstrated at nanometer resolution in isolated objects<sup>4–8</sup> and Ångström (Å) resolution in micro/nanocrystals<sup>9,10</sup>. Yet the goal of near-atomic resolution single particle imaging, using X-rays, remains elusive.

The single particle imaging (SPI) initiative is a large collaborative team of researchers from several institutions formed to identify and solve the challenges required for high resolution imaging with XFELs. The aim of the SPI initiative, as laid out in a published roadmap<sup>11</sup>, is to establish a community-wide approach to take up the scientific and technical challenges of single-molecule imaging with X-rays. In addition to developing solutions to the technical challenges, a critical part of this project was selecting a well-characterized model system needed for demonstration experiments. After considering homogeneity, uniform size distribution, particle concentration, having a known structure, and the ability to be aerosolized for injection into the XFEL beam, Rice Dwarf Virus (RDV) was selected for the first experiment (see methods) in this initiative.

RDV is an icosahedral RNA virus of about 70 nm in diameter, and is the causative agent of rice dwarf disease. This disease creates severe economic damage in China, Japan and other Asian countries due to speck formation and its destructive effects on plant growth for rice, wheat, barley, and other gramineae plants. Leafhopper insects are the primary host in which the virus particles replicate and from which they are then transmitted to leaves. The virus particles consist of two shells, an inner and an outer capsid, enclosing a double stranded RNA genome. The genome encodes 12 products, seven of which are considered structural proteins. A thin layer of P3 capsid proteins<sup>12</sup> makes up the inner capsid and three proteins of mainly P8 (ref. 13), but also P2 (ref. 14) and P9 (ref. 15), form the outer capsid. Found in the core, together with the genome, are P1 (putative RNA polymerase<sup>16</sup>), P5 (putative guanylyltransferase<sup>17</sup>) and P7 (a nonspecific nucleic acid binding protein<sup>18</sup>). A 3D structure of the capsid was previously solved by X-ray crystallography at 3.5 Å resolution (PDB 1UF2)<sup>19</sup>.

RDV was aerosolized and delivered into the hard X-ray Coherent X-ray Imaging (CXI) nanofocus instrument at Linac Coherent Light Source (LCLS)<sup>20,21</sup> using an aerodynamic lens injector<sup>22</sup>. Diffraction patterns were recorded at a rate of 120 Hz using two Cornell-SLAC Pixel Array Detectors (CSPAD), a large 2.3 Mpix detector located close to the sample for wide angle scattering and another smaller 0.14 Mpix '2 × 2' detector located further downstream to detect small angle scattering<sup>23,24</sup> (pictorially shown in center panel of Fig. 1). The photon energy was 7 keV, the pulse duration was < 50 fs, and the average pulse energy immediately downstream of the undulator was 4 mJ. (see Methods for a detailed description).

In the data deposited in the Coherent X-ray Imaging Data Bank (CXIDB)<sup>25</sup>, we record diffraction from RDV particles hit by the LCLS X-ray pulse on the back detector. Also recorded is elevated scattering on the front detector for 'hits', that are identified based on the data on the back detector. This indicates that measurable photons are recorded from the sample up to a scattering angle commensurate with 5.9 Å resolution.

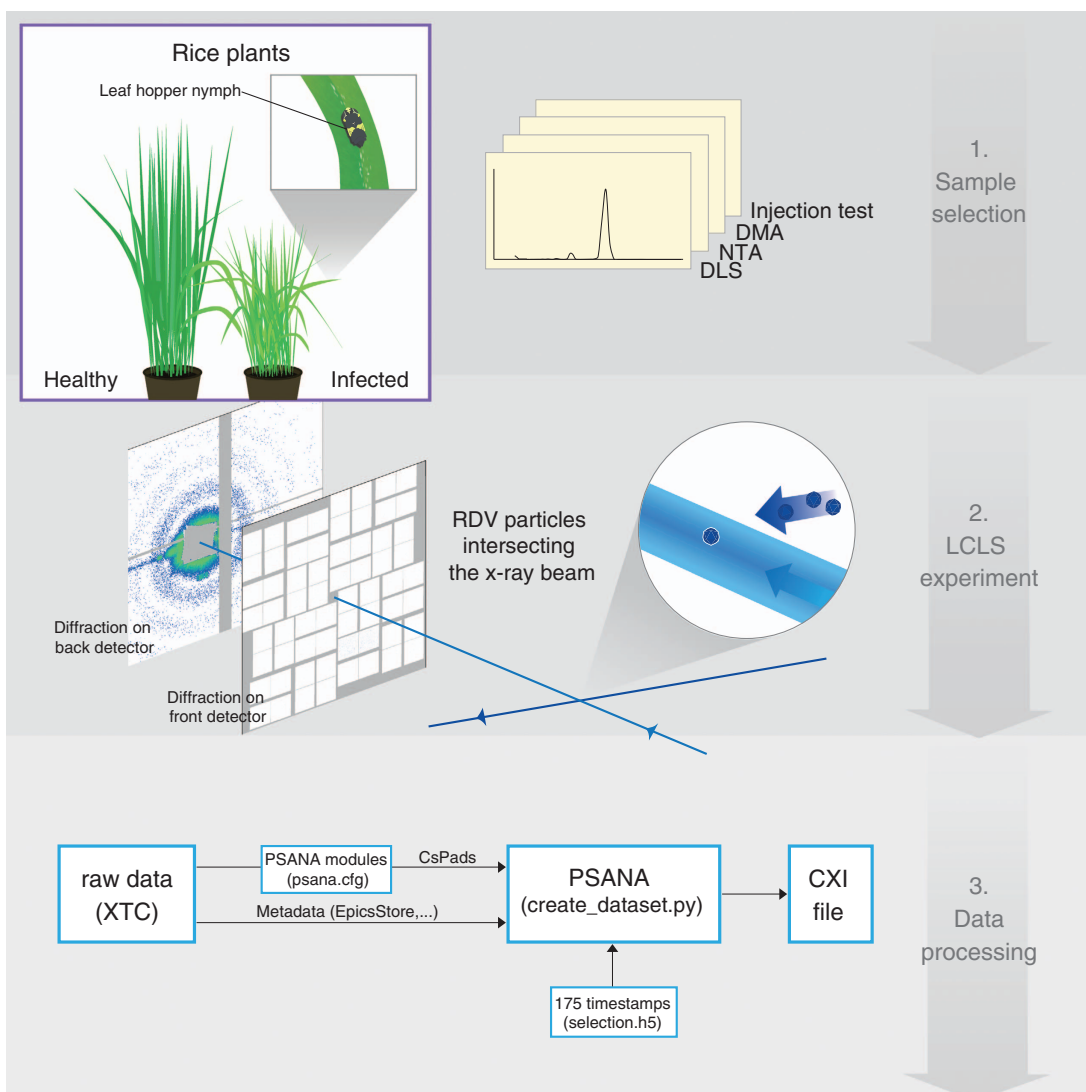
## Methods

### Sample preparation

Nymphs of *Nephotettix nigropictus* were fed on diseased rice plants. The purification procedure of the RDV O strain<sup>26</sup> followed the procedure of Omura *et al.*<sup>27</sup> with modified sucrose concentrations as follows. A meat chopper was used to grind the infected rice leaves and the resultant slurry was treated with CCl<sub>4</sub> and subjected to repeated precipitations and consecutive density gradient centrifugations in 40 to 60% and 40 to 70% sucrose. The pellet from the final centrifugation of the viral particle band was resuspended in a 0.1 M solution of histidine that contained 0.01 M MgCl<sub>2</sub> (pH 6.2). The sample contained all viral components except the P2 protein, which was removed by the CCl<sub>4</sub> treatment. This removal prevents infection through oral intake by the insect and direct injection is instead required for vector infection.

### Pre-characterization experiments

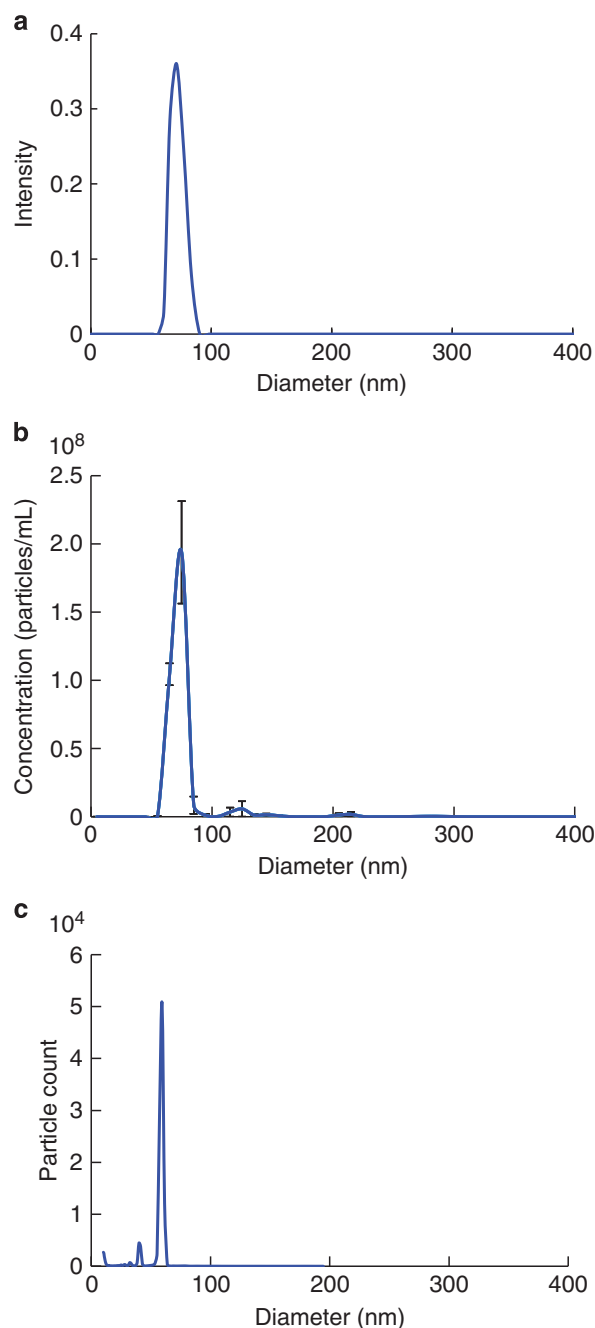
Reference samples need to have a known structure, be available at high concentration, exhibit monodisperse size distribution and be compatible with available sample delivery techniques. Pre-characterization was performed using Dynamic Light Scattering (DLS), Nanoparticle Tracking Analysis (NTA), Differential Mobility Analysis (DMA) and aerosol injection testing as shown in Fig. 2 (Data Citation 1). RDV was determined to satisfy these requirements and was selected as the test sample for the LCLS experiment.



**Figure 1.** Experimental design. (1) As the first step in the experiment, an analysis of candidate samples was carried out and a primary target (Rice Dwarf Virus, RDV) was selected. RDV was purified from grasshopper nymphs, which were fed infected rice plants as described in the text. (2) Purified virus particles were then injected into the X-ray beam of the LCLS and diffraction patterns were recorded on the front and back detectors of the CXI instrument<sup>20</sup>. (3) The raw data were pre-processed using *psana*<sup>30</sup> and converted to XTC files. 175 frames of strong hits were selected and converted into the CXI file form.

**Injection testing.** The sample was injected using an injection setup identical to that used in the LCLS experiment (see *Sample injection at the LCLS*) to investigate their ability to aerosolize and their resistance to the injection procedure. By placing a microscope glass slide covered by a gel piece (Gel-Pak) beneath the outlet of the aerodynamic lens (at the same position as the interaction region with the X-ray beam in the subsequent LCLS experiments), a particle dust could be observed through an objective lens mounted below. In a second set of experiments a formvar/carbon grid (#01754-F, F/C 400 mesh Cu, Ted Pella Inc.) was substituted for the glass slide, to capture RDV particles that had traversed the injector. These samples were examined further using an environmental scanning electron microscope (ESEM) (Quanta FEG 650, FEI). The pressure in the vacuum chamber was kept at approximately  $10^{-5}$  mBar.

**Sample size and monodispersity in the liquid phase.** The size distribution of the RDV sample in solution (250 mM ammonium acetate buffer, pH 7.5) was measured using both DLS (w130i, AvidNano Ltd. and Spectrolight 600, Molecular Dimensions) and NTA techniques (NanoSight, model LM10,



**Figure 2.** The purity of RDV was analyzed with DLS, NTA and DMA. (a), Size distribution determined with DLS. The average size of the RDV particles was approximately 76 nm. (b), Size distribution determined with NTA. The mean and mode of the RDV particles were approximately 76 and 73 nm, respectively. (c), DMA measurement of RDV gave an apparent particle diameter of approximately 60 nm. Here the actual gas flow was lower than the set value, which resulted in an under estimation of the particle size by about 10%. The main purpose of the DMA measurement was to assess sample purity and the results show that.

Malvern Instruments Ltd.). For these DLS and NTA measurements, the sample was diluted to  $10^9$  particles  $\text{ml}^{-1}$  and  $10^8$  particles  $\text{ml}^{-1}$ , respectively. The measured size distribution is shown in Fig. 2a,b.

**Sample size and monodispersity in the gas phase.** The size distribution of the RDV sample in the gas phase were measured by means of Electrophoretic DMA. RDV was aerosolized with a nano-Electrospray ionization (ESI) source (TSI model 3480) and passed through an electrostatic classifier (TSI model 3480) whose size selection window was continuously scanned. Transmitted particles were counted with a condensation particle counter (CPC, TSI model 3786). The size distribution is shown in Fig. 2c.

### Sample injection at LCLS

The experiment was carried out at the CXI instrument at the LCLS<sup>19,20</sup>. An aerosol injector (described in ref. 8) was used to introduce the particles into the X-ray beam. Purified RDV were transferred to a volatile buffer (250 mM ammonium acetate, pH 7.5) at a concentration of  $10^{12}$  particles  $\text{ml}^{-1}$  and introduced to the injector via a gas dynamic virtual nozzle (GDVN)<sup>28</sup> at a flow rate of  $1\text{--}2\ \mu\text{l min}^{-1}$ . The aerosol passed through a skimmer and a relaxation chamber and it was focused into a narrow particle beam by an aerodynamic lens. By regulating gas, liquid flow and skimmer pressure, the quality of the particle beam could be optimized. Injected particles intersected the X-ray beam in random orientations.

### Experimental setup and data collection

Data were collected at the CXI instrument at LCLS. LCLS was tuned to a photon energy of 7 keV and produced pulses with approximately 4 mJ pulse energy and  $< 50$  fs duration. The selection of the photon energy, within the CXI operation range of 5 to 11 keV, was driven by the competition between the sample's scattering cross section, which tends to be high at lower energies, and the ability to discriminate single-photon events in the detector, which increases at the high end of the range. Additionally, photon energies below the Iron K-alpha excitation edge at 7.1 keV avoid an isotropic fluorescence signal from the steel walls of the vacuum chamber, which in turn would complicate the identification of photons scattered by the sample.

X-rays were focused using a pair of Kirkpatrick-Baez (KB) mirrors to a nominal size of  $0.1 \times 0.1\ \mu\text{m}$ . The focused beam passed through a set of beam-defining apertures to reduce the X-ray scattering imperfections in the optical system as the FEL beam overfills the KB entrance. Additional cleanup slits and apertures, including a post sample aperture, are used to limit background scatter. The post sample aperture limited the collection angle to commensurate with  $5.9\ \text{\AA}$  resolution. The post sample aperture is a small 3 mm circular aperture positioned just downstream of the interaction region and is used to prevent scattered X-rays coming from locations other than the sample position reaching the sensitive surface of the detector. Small angle diffraction patterns were recorded with a CSPAD 0.14 Mpix detector (also referred to as the back detector), located 2.4 m downstream of the interaction and high-angle scattering was captured on a 2.3 Mpix CSPAD detector (also referred to as the front detector), located 217.4 mm downstream from the interaction point, in a tandem arrangement as shown in the center panel of Fig. 1 (refs 23,24). All data events were recorded and synchronized with the LCLS repetition rate of 120 Hz. The back detector was offset with respect to the optical axis of the focusing optics, and extended to a maximal resolution of c. 15.2 nm and c. 11.6 nm on the edge and in the corner, respectively. A semitransparent beam-stop, consisting of  $25\ \mu\text{m}$  Ti and  $100\ \mu\text{m}$  Zn, was utilized so that very low-q scattering could be collected, as well as provide a monitor for the direct beam. Data were analyzed onsite using *Hummingbird*, a fast online analysis tool developed for single particle imaging<sup>29</sup> and *Cheetah*, a software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data<sup>30</sup>.

### Data processing steps used

To provide interpretable data in addition to the raw XTC files, the native format of the LCLS data stream, we selected a small subset of diffraction events (175 frames) and converted them into a CXI file using *psana*<sup>31</sup>. For both back and front detectors, data were 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. Pixel gains were calculated by generating per-pixel histograms from a flat-field run and fitting a bimodal distribution with respect to the noise peak and the single photon peak. Using *psana's CSPadPixCoords.CSPadImageProducer* and *CSPadPixCoords.CSPad2x2ImageProducer*, the detector panels were assembled in order to form real images. For a list of provided files and data entries, see section Data Records. The 175 patterns were selected manually by means of identifying strong diffraction signal showing similarities to simulations. See Technical Validation for more details.

### Data Records

#### Data citation 1—Sample size and monodispersity

The data are available at Figshare (Data Citation 1) and contains an excel file with raw data from the DLS, NTA and DMA measurements.

#### Data citation 2—Coherent diffractive imaging data

The data are deposited in the CXIDB<sup>25</sup> (Data Citation 2) and stored in the CXIDB data format, which is based on the HDF5 format. HDF5 files are readable in many computing environments, including Python using the *h5py* module and MATLAB using e.g., the *h5read* function. Convenient functions for accessing the CXIDB data file exist in the *libspimage* package for C and Python<sup>32</sup>. For visualizing data in the CXIDB format, the *Owl* software is convenient (<https://github.com/FilipeMaia/owl/>). 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*, *psana* and *Cheetah* are provided for completeness of describing processing performed on the deposited data. Table 1 describes the files deposited in the CXIDB.

Experimental data		
Data type	Example filename	File format
Raw data (all exposures)	e625-r0002-s00-c00.xtc	XTC
Diffraction Pattern (only selected hits)	cxidb-36.cxi	CXI
Metadata		
Data type	Filename	File format
Cheetah initialization file (preprocessing)	cheetah.ini	Text
Cheetah configuration file (preprocessing)	psana.cfg	Text
Bad pixel mask ( <i>front</i> detector)	mask_cspad.h5	Hdf5
Bad pixel mask ( <i>back</i> detector)	mask_cspad2x2.h5	Hdf5
Conversion script	create_dataset.py	Python
File containing the timestamps of the selected hits	selection.h5	Hdf5
Cheetah preprocessing		
Data type	Folder name	File format
Calibration data	calib	various
Cheetah GUI related files	gui	various
Hdf5 files of dark runs and hits	hdf5	Hdf5
Indexing files	indexing	various
Cheetah preprocessing related scripts	process	various

**Table 1.** Deposited data and configuration files. This table describes the files deposited in the CXIDB under accession number ID 36. ID 36 consists of Experimental data, *Cheetah* preprocessing files, and metadata.

## Technical Validation

### Background scattering and direct beam scatter

A background scattering pattern was derived by averaging 1000 frames that did not include any hits or dark frames. This background, as well as suggested masks for non-responsive pixels and beamstops, are shown in Fig. 3.

Additionally, for the back detector data, manifold-embedding methods were used to detect and identify the nature and origin of stochastic changes, and quantify the necessary corrections to the background. The manifold of raw RDV single-particle snapshots is shown in Fig. 4, where each point represents a diffraction pattern<sup>33–35</sup>. The parabolic nature of this manifold reveals that a single parameter dominates the changes from snapshot to snapshot, namely fluctuations in pulse intensity, consistent with the self-amplified spontaneous emission process of the FEL. This can be corrected by appropriate normalization procedures<sup>36</sup>. In addition to a monotonic intensity change along the parabola, a prominent deviation is evident. This is caused by a shift of about one pixel, of the center of intensity along the lateral direction of the beam. The cause of this shift is a drift in pointing of the offset mirrors on the beam-defining aperture of the KB mirrors for the CXI beamline.

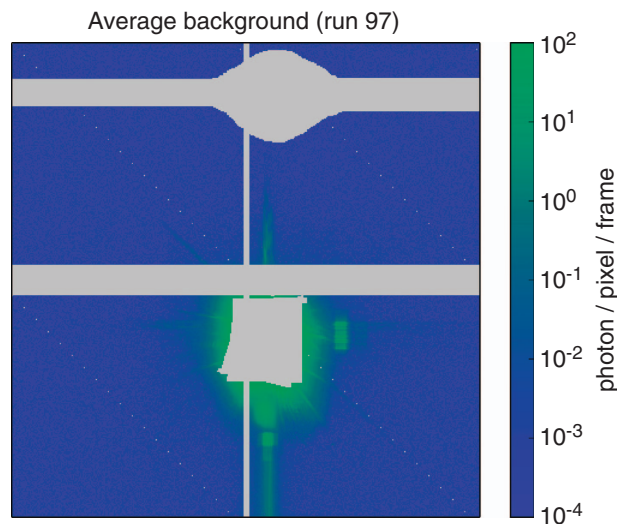
### Simulated diffraction data of expected size

In Fig. 5, two diffraction patterns from different single particle hits are shown in comparison to simulated diffraction from homogeneous spheres of size 71 nm. In the simulation, a photon energy of 7 keV and an assumed mass density for RDV of  $1.381 \text{ g cm}^{-3}$  was used. The back detector was simulated using a detector distance of 2.4 m, a pixel size of 110 microns and a signal conversion rate of 33 ADUs per photon.

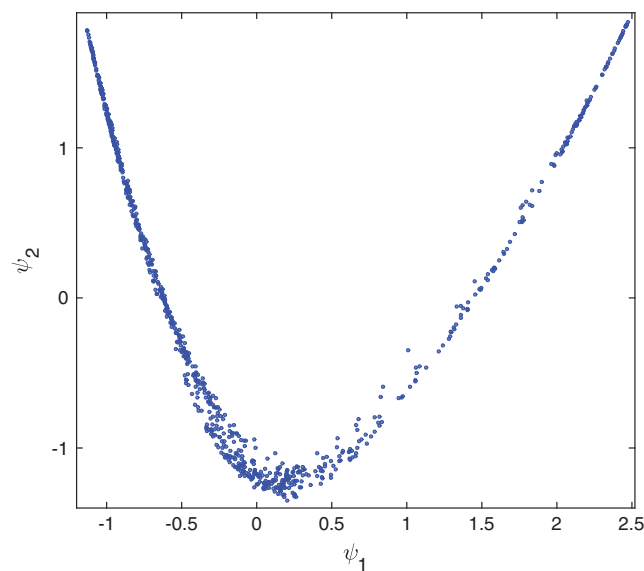
### Signal above background on front detector

The front detector was located 217.4 mm downstream of the sample interaction region and collected diffraction data extending to a resolution of 5.9 Å. Although hits are immediately apparent on the back detector, and this signal is used for hit finding, determining whether there is useful signal from the sample on the front detector above background levels is not immediately apparent from any individual image. A radial average of the sum of frames determined by *Cheetah* to be hits shows that there is indeed consistently elevated signal above background when sample is detected in the beam based on the back detector (Fig. 6). This intensity distribution falls off with the expected q-dependence, and stops at a resolution of 5.9 Å. This resolution limit is set by the angular acceptance of the post-sample aperture. Beyond this resolution both radial sums are identical, further supporting the notion that signal up





**Figure 3.** Background image of the back detector, averaged from 1000 non-hits and non-dark frames. The grey areas are masked out. They correspond to the beamstop, the gap between sensors, and a shadow mask of an additional beamstop from the beamstop holder. Unbonded pixels that do not read out signal, are likewise masked out.



**Figure 4.** Manifold of raw back detector data in two dimensions. Each point represents a diffraction pattern. The axes are orthogonal coordinates provided by the manifold embedding algorithm<sup>37</sup>.

to 5.9 Å resolution comes from individual particles. This validates the potential usefulness of signal on the front detector for image analysis. *Cheetah* processing scripts are included in the archive.

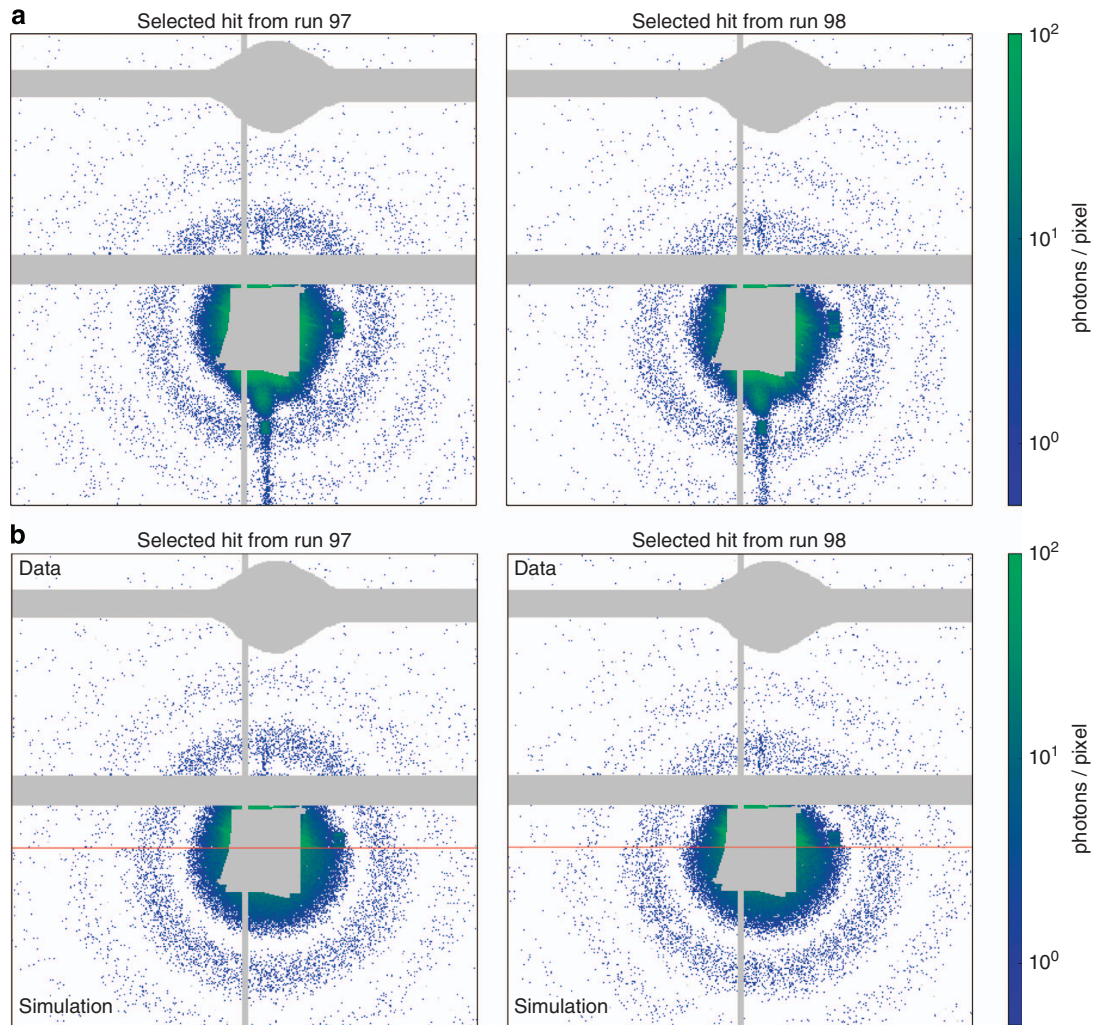
#### Validation that scattering comes from RDV

The method of least surprise (described below) was used to determine whether signal on the front detector corresponded to the expected signal from RDV.

Data were converted from Analog-to-Digital Units (ADU) measured by the detector into photon counts using the relation

$$k_i = \text{ceil} [(A_i - 0.5\gamma)/\gamma]$$

where  $k_i$  is the photon count at pixel  $i$ ,  $A_i$  is the dark, common-mode, gain-corrected ADU measured at pixel  $i$ , and  $\gamma$  is the average ADUs per photon for the detector which was calculated from a flat-field run.



**Figure 5.** Measured and simulated diffraction patterns of single RDV particles. **(a)**, Two hits on the back detector, selected on the expected size for RDV and high diffraction intensity. **(b)**, Measured and simulated data combined. The top half of each of the two patterns shows the measured signal and is identical to the top halves of the patterns in **(a)**. The bottom half of each pattern shows simulated diffraction data from a homogenous sphere of size 71 nm and with a mass density of  $1.381 \text{ g cm}^{-3}$ . The simulation assumes a photon energy of 7 keV, a detector distance of 2.4 m, a pixel size of 110 microns and a conversion of 33 ADUs per photon. Regions of beam-stops and gaps between the detector panels are masked in grey.

In this analysis, we use front detector data up to  $6.67 \text{ \AA}$  resolution, which corresponds to a radius of 265 pixels.

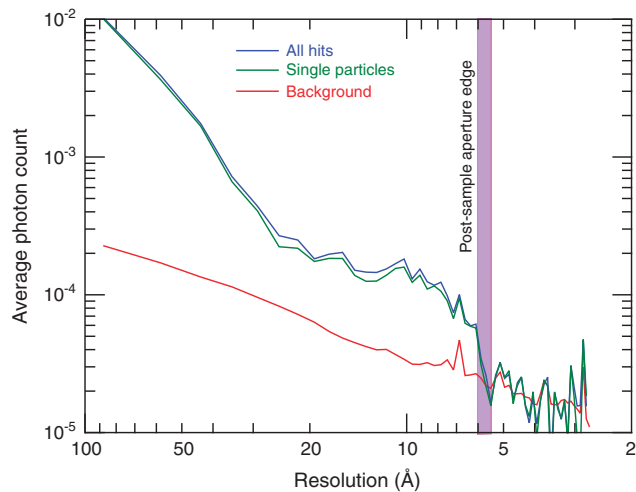
Assuming Poisson statistics, we define the surprise function as the negative log-likelihood

$$S(K; \Phi, \Omega_j) = - \sum_{i=1}^N \log \left( \frac{n_i^{k_i} e^{-n_i}}{k_i!} \right) \equiv - \sum_{i=1}^N \log P(n_i, k_i)$$

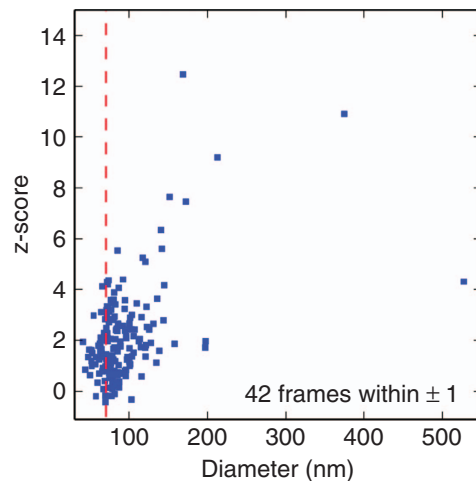
where  $K$  denotes the dependence on data, with  $k_i$  being the measured photon count at pixel  $i$ ,  $n_i$  is the average photon number at pixel  $i$  when the fluence is  $\Phi$  and the RDV particle has orientation  $\Omega_j$ , and the summation runs over all the pixels. Minimizing the surprise function, or maximizing the log-likelihood, across different orientations and fluence values, we assign each data frame with the orientation and fluence at which it was most likely recorded. To help us assess the quality of these assignments, we further ‘normalize’ the surprise function. Given the RDV model (PDB 1UF2)<sup>19</sup> with an estimate of the particle orientation and fluence, we calculate the mean

$$\langle S(\Phi, \Omega_j) \rangle = - \sum_{i=1}^N \langle \log P(n_i, k) \rangle$$





**Figure 6.** Radial average of signal on the front detector from blank frames compared to radial average from frames determined to be hits. Elevated photon counts from the sample are visible up to an angle commensurate with 5.9 Å resolution, this being the resolution limit set by the angular acceptance of the post-sample aperture. ‘All hits’ is the average of all hits and includes all particles independent of size (including clusters of particles), while ‘single particle’ is the average of hits that are of the appropriate size to be isolated single particles.



**Figure 7.** Front detector normalized surprise (z-score) versus back detector particle size fits. The dashed red line indicates the diameter (70.8 nm) of the RDV model. The normalized surprise function, or its z-score, measures the agreement of the data with a known model: The data are inconsistent with the model when the absolute value of the z-score is much greater than unity: a z-score much greater than unity is consistent with the data being ‘surprising’ given the assumed model.

and the standard deviation

$$\sigma_S(\Phi, \Omega_j) = \left[ \sum_{i=1}^N \langle (\log P(n_i, k))^2 \rangle - \langle \log P(n_i, k) \rangle^2 \right]^{1/2}$$

of the surprise function, where  $\langle \dots \rangle$  denotes the expectation value under the Poisson distribution. Note that  $\langle S(\Phi, \Omega_j) \rangle$  and  $\sigma_S(\Phi, \Omega_j)$  are independent of the data  $K$ . The normalized surprise function, or its z-score,

$$z(K; \Phi, \Omega_j) \equiv \frac{S(K; \Phi, \Omega_j) - \langle S(\Phi, \Omega_j) \rangle}{\sigma_S(\Phi, \Omega_j)}$$

measures the agreement of the data with a known model: The data are inconsistent with the model when

the absolute value of the z-score is much greater than unity: a z-score much greater than unity is consistent with the data being ‘surprising’ given the assumed model.

The z-scores of all the selected frames versus particle size are shown in Fig. 7. The particle sizes were determined by fitting back detector data to a homogeneous sphere model with adjustable size. Frames with particle size close to the diameter (70.8 nm) of the RDV model generally have smaller z-scores, though some still manifest inconsistency with the model. The source of this could be the presence of a water layer on the particle surface. This model-based surprise function calculation may potentially be useful for hit-finding, especially when the signal is as weak as the front detector data.

### Usage Notes

The dataset (CXIDB ID 36) contains the full data stream recorded during the experiment in .xtc format. The dataset also contains a set of pre-selected hits as a CXI file (as described above) from both CSPAD detectors plus instrument metadata. XTC files are the native format of LCLS and can be read using analysis frameworks provided by the LCLS (see <https://confluence.slac.stanford.edu/display/PSDM/LCLS+Data+Analysis>).

### References

- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
- 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).
- Howells, M. R. *et al.* An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy. *J. Electron Spectros. Relat. Phenomena* **170**, 4–12 (2009).
- Chapman, H. N. *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Phys* **2**, 839–843 (2006).
- Seibert, M. M. *et al.* Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* **470** 78–81 (2011).
- 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).
- van der Schot, G. *et al.* Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nat. Commun* **6**, 5704 (2015).
- Hantke, M. F. *et al.* High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nat. Photonics* **8**, 943–949 (2014).
- Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 (2011).
- Boutet, S. *et al.* High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
- Aquila, A. *et al.* The linac coherent light source single particle imaging road map. *Struct. Dyn* **2**, 041701 (2015).
- Kano, H. *et al.* Nucleotide sequence of rice dwarf virus (RDV) genome segment S3 coding for 114 K major core protein. *Nucleic Acids Res.* **18**, 6700 (1990).
- Omura, T. *et al.* The outer capsid protein of rice dwarf virus is encoded by genome segment S8. *J Gen Virol.* **70**, 2759–2764 (1989).
- Yan, J. *et al.* P2 protein encoded by genome segment S2 of rice dwarf phytoevirus is essential for virus infection. *Virology* **224**, 539–541 (1996).
- Zhong, B. *et al.* A minor outer capsid protein, P9, of Rice dwarf virus. *Arch. Virol.* **148**, 2275–2280 (2003).
- Suzuki, N. *et al.* Molecular analysis of rice dwarf phytoevirus segment S1: intervirion homology of the putative RNA-dependent RNA polymerase between plant- and animal-infecting reoviruses. *Virology* **190**, 240–247 (1992).
- Suzuki, N., Kusano, T., Matsuura, Y. & Omura, T. Novel NTP binding property of rice dwarf phytoevirus minor core protein P5. *Virology* **219**, 471–474 (1996).
- Ueda, S. & Uyeda, I. The rice dwarf phytoevirus structural protein P7 possesses non-specific nucleic acids binding activity in vitro. *Mol. Plant Pathol. Online* <http://www.bspp.org.uk/mppol/1997/0123ueda/> (1997).
- Nakagawa, A. *et al.* The atomic structure of Rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* **11**, 1227–1238 (2003).
- Boutet, S. & Williams, G. J. The Coherent X-ray Imaging (CXI) instrument at the Linac Coherent Light Source (LCLS). *New J. Phys.* **12**, 035024 (2010).
- Liang, M. *et al.* The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 514–519 (2015).
- Bogan, M. J. *et al.* Single particle X-ray diffractive imaging. *NANO Lett* **8**, 310–316 (2008).
- Hart, P. *et al.* The CSPAD megapixel X-ray camera at LCLS. *Proc. SPIE* **8504**, 8504 (2012).
- Herrmann, S. *et al.* CPSAD upgrades and CSPAD V1.5 at LCLS. *J. Phys. Conf. Ser* **493**, 012013 (2014).
- Maia, F. R. N. C. The coherent X-ray imaging data bank. *Nat. Methods* **9**, 854–855 (2012).
- Kimura, I., Minobe, Y. & Omura, T. Changes in a nucleic acid and a protein component of Rice dwarf virus particles associated with an increase in symptom severity. *J. gen. Virol.* **68**, 3211–3225 (1987).
- Omura, T., Morinaka, T., Inoue, H. & Saito, Y. Purification and some properties of rice gall dwarf virus, a new Phytoevirus. *Phytopath* **72**, 1246–1249 (1982).
- DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D. Appl. Phys* **41**, 195505 (2008).
- 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).
- Barty, A. *et al.* Cheetah: Software for High-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr* **47**, 1118–1131 (2014).
- Damiani, D. *et al.* Linac Coherent Light Source data analysis using psana. *J. Appl. Cryst.* **49**, 672–679 (2016).
- Maia, F. R. N. C., Ekeberg, T., van der Spoel, D. & Hajdu, J. Hawk: the image reconstruction package for coherent X-ray diffractive imaging. *J. Appl. Crystallogr* **43**, 1535–1539 (2010).
- Giannakis, D., Schwander, P. & Ourmazd, A. The symmetries of image formation by scattering. I. Theoretical framework. *Opt. Express* **20**, 12799–12826 (2012).
- Schwander, P., Giannakis, D., Yoon, C. H. & Ourmazd, A. The symmetries of image formation by scattering. II. Applications. *Opt. Express* **20**, 12827–12849 (2012).
- Hosseinzadeh, A. *et al.* High-resolution structure of viruses from diffraction snapshots. *Philos. Trans. R. Soc. Lond. B Biol. Sci* **369**, 20130326 (2014).

36. Hosseinizadeh, A., Dashti, A., Schwander, P., Fung, R. & Ourmazd, A. Single-particle structure determination by X-ray free electron lasers: possibilities and challenges. *Struct. Dyn* **2**, 041601 (2015).
37. Coifman, R. R. & Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006).

## Data Citations

1. Munke, A. *Figshare*. <http://dx.doi.org/10.6084/m9.figshare.c.2342581> (2016).
2. Munke, A. *Coherent X-ray Imaging Data Bank*. <http://dx.doi.org/10.11577/1252456> (2016).

## Acknowledgements

‘Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.’

ANL: 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.

ARC: Australian Research Council Centre of Excellence in Advanced Molecular Imaging (CE140100011) [www.imagingcoe.org](http://www.imagingcoe.org).

ASU: U.S. National Science Foundation (NSF) Science and Technology Center BioXFEL Award 1231306 and the National Institute of Health (NIH) U54 GM094599, R01GM095583 and GM097463.

BNL: Brookhaven National Laboratory is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-SC0012704.

CBIS: The Lee Kuan Yew Endowment Fund.

CFEL: Helmholtz Association through project oriented funds, Bundesministerium für Building und Forschung (BMBF) project 05K2012.

CSRC: Natural Science Foundation of China (project id: 11575021).

DESY: 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.

ELI: The Ministry of Education, Youth and Sports of the Czech Republic (ELI-Beamlines Registered No. CZ.1.05/1.1.00/02.0061) and the Academy of Sciences of the Czech Republic (M100101210).

EXFEL: The Helmholtz Virtual Institute: Dynamic Pathways in Multidimensional Landscapes.

GIST & POSTECH: NRF through the SRC (Grant No. NRF- 2015R1A5A1009962), MSIP and PAL, Korea.

LASSP: U.S. Department of Energy (DOE) Grant No. DE-SC0005827.

OSAKA: Japan Society for the Promotion of Science, The Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT); Grants-in-Aid for Scientific Research and Bilateral Joint Research Projects by Japan Society for the Promotion of Science; X-ray Free Electron Laser Priority Strategy Program (The Ministry of Education, Culture, Sports, Science and Technology of Japan, MEXT).

SLAC: The LCLS is supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (OBES), under contract DEAC0276SF00515. 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 ChEMH and SLAC National Accelerator Laboratory seed grant program.

UU: The Swedish Research Council (VR), the Röntgen-Ångström Cluster, the Swedish Foundation for Strategic Research (SSF), the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), the Knut and Alice Wallenberg Foundation, European Research Council (ERC).

UW: US Dept. of Energy, Office of Energy Sciences under award DE-FG02-09ER16114, and the US National Science Foundation under awards STC 1231306 and 1551489.

UWM: US Department of Energy, Office of Science, Basic Energy Sciences under award DE-SC0002164 (algorithm design and development, and data analysis), and by the US National Science Foundation under awards STC 1231306 (numerical trial models) and 1551489 (underlying analytical models).

Paulraj Lourdu Xavier: European Research Council, ‘Frontiers in Attosecond X-ray Imaging and Spectroscopy (AXSIS)’, ERC-2013-SyG 609920 (2014-2018)

Salah Awel: The Hamburg Center for Ultrafast Imaging—Structure, Dynamics and Control of Matter at the Atomic Scale’ of the Deutsche Forschungsgemeinschaft (CUI, DFG-EXC1074).

## Author Contributions

A.Hi., A.M., A.N., B.G.H., D.S.D.L., H.D., H.K.N.R., J.H., J.Z., M.S., P.F., P.L.X. and R.G.S. and prepared and characterized RDV for the study. A.A., A.B., A.Ho., A.P.M., A.M., A.O., B.G.H., C.H.Y., D.S.D.L., G.J.W., G.V.D.S., H.D., H.K.N.R., J.B., J.H., J.S., K.M., K.O., M.F.H., M.M.S., M.S., N.T., P.B., P.F., P.L.X., P.S., R.G.S., R.J.B., S.B., T.Y.L. and V.E. participated in the sample selection procedure. J.B., J.H., M.F.H., K.M., M.S., M.M.S. and D.W. developed the sample injector used in the experiment. A.A., A.B., A.M., B.D., B.G.H., C.H.Y., C.N., D.W., G.J.W., G.N., H.D., H.L., J.B., J.Z., K.A., M.B., M.F.H., M.R., N.T., P.B., P.L.X., P.S., R.A.K., R.G.S., R.J.B. and S.B. participated at the LCLS experiment. A.B., B.D., C.H.Y., C.N., F.R.N.C.M., H.L., K.A., M.F.H., M.R., N.D.L., T.Y.L. and V.E. processed and validated the data. A.O., A.Ho. and P.S. performed the manifold analysis. T.Y.L. and V.E. performed the surprise validation. All authors contributed to writing and editing the manuscript.

## Additional Information

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

**How to cite this article:** Munke, A. *et al.* Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. *Sci. Data* 3:160064 doi: 10.1038/sdata.2016.64 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.

© The Author(s) 2016

Anna Munke<sup>1</sup>, Jakob Andreasson<sup>1,2</sup>, Andrew Aquila<sup>3</sup>, Salah Awel<sup>4</sup>, Kartik Ayyer<sup>4</sup>, Anton Barty<sup>4</sup>, Richard J. Bean<sup>5</sup>, Peter Berntsen<sup>6</sup>, Johan Bielecki<sup>1</sup>, Sébastien Boutet<sup>3</sup>, Maximilian Bucher<sup>3,7,8</sup>, Henry N. Chapman<sup>4,9</sup>, Benedikt J. Daurer<sup>1</sup>, Hasan DeMirci<sup>3,10</sup>, Veit Elser<sup>11</sup>, Petra Fromme<sup>12,13</sup>, Janos Hajdu<sup>1</sup>, Max F. Hantke<sup>1</sup>, Akifumi Higashiura<sup>14</sup>, Brenda G. Hogue<sup>13,15,16</sup>, Ahmad Hosseinizadeh<sup>17</sup>, Yoonhee Kim<sup>18</sup>, Richard A. Kirian<sup>13,19</sup>, Hemanth K.N. Reddy<sup>1</sup>, Ti-Yen Lan<sup>11</sup>, Daniel S.D. Larsson<sup>1</sup>, Haiguang Liu<sup>20</sup>, N. Duane Loh<sup>21</sup>, Filipe R.N.C. Maia<sup>1</sup>, Adrian P. Mancuso<sup>5</sup>, Kerstin Mühlig<sup>1</sup>, Atsushi Nakagawa<sup>14</sup>, Daewoong Nam<sup>22</sup>, Garrett Nelson<sup>19</sup>, Carl Nettelblad<sup>1,23</sup>, Kenta Okamoto<sup>1</sup>, Abbas Ourmazd<sup>17</sup>, Max Rose<sup>24</sup>, Gijs van der Schot<sup>1</sup>, Peter Schwander<sup>17</sup>, M. Marvin Seibert<sup>1</sup>, Jonas A. Sellberg<sup>1,25</sup>, Raymond G. Sierra<sup>3,10</sup>, Changyong Song<sup>22</sup>, Martin Svenda<sup>1</sup>, Nicusor Timneanu<sup>1,26</sup>, Ivan A. Vartanyants<sup>24,27</sup>, Daniel Westphal<sup>1</sup>, Max O. Wiedorn<sup>4,9</sup>, Garth J. Williams<sup>28</sup>, Paulraj Lourdu Xavier<sup>4,9,29</sup>, Chun Hong Yoon<sup>3</sup> & James Zook<sup>12,13</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>Institute of Physics ASCR, v.v.i. (FZU), ELI-Beamlines Project, Prague 182 21, Czech Republic. <sup>3</sup>SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>4</sup>Center for Free Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Hamburg 22607, Germany. <sup>5</sup>European XFEL GmbH, Holzkoppel 4, Schenefeld 22869, Germany. <sup>6</sup>Australian Research Council Centre of Excellence in Advanced Molecular Imaging, La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086, Australia. <sup>7</sup>Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA. <sup>8</sup>Institut für Optik und Atomare Physik, Technische Universität Berlin, Hardenbergstraße 36, Berlin 10623, Germany. <sup>9</sup>Department of Physics, University of Hamburg, Hamburg 22761, Germany. <sup>10</sup>Stanford PULSE Institute, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>11</sup>Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA. <sup>12</sup>Arizona State University, School of Molecular Sciences (SMS), Tempe, Arizona 85287-1604, USA. <sup>13</sup>Biodesign Center for Applied Structural Discovery, Biodesign Institute at Arizona State University, Tempe 85287, USA. <sup>14</sup>Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan. <sup>15</sup>Arizona State University, School of Life Sciences (SOLS), Tempe, Arizona 85287-5401, USA. <sup>16</sup>Biodesign Center for Infectious Diseases and Vaccinology, Biodesign Institute at Arizona State University, Tempe 85287, USA. <sup>17</sup>Department of Physics, University of Wisconsin Milwaukee, 3135 North Maryland Ave., Milwaukee, Wisconsin 53211, USA. <sup>18</sup>School of Materials Science and Engineering, Gwangju Institute of Science and Technology, Gwangju 61005, Korea. <sup>19</sup>Arizona State University, Department of Physics, Tempe, Arizona 85287, USA. <sup>20</sup>Beijing Computational Science Research Center, 8 W Dongbeiwang Rd, Haidian, Beijing 100193, China. <sup>21</sup>Centre for Bio-imaging Sciences, National University of Singapore, 14 Science Drive 4, BLK S1A, Singapore 117543, Singapore. <sup>22</sup>Department of Physics, Pohang University of Science and Technology, Pohang 37673, Korea. <sup>23</sup>Department of Information Technology, Science for Life Laboratory, Uppsala University, Lägerhyddsvägen 2 (Box 337), Uppsala SE-75105, Sweden. <sup>24</sup>Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, Hamburg D-22607, Germany. <sup>25</sup>Biomedical and X-Ray Physics, Department of Applied Physics, AlbaNova University Center, KTH Royal Institute of Technology, Stockholm SE-106 91, Sweden. <sup>26</sup>Department of Physics and Astronomy, Uppsala University, Lägerhyddsvägen 1 (Box 516), Uppsala SE-75120, Sweden. <sup>27</sup>National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashirskoe shosse 31, Moscow 115409, Russia. <sup>28</sup>Brookhaven National Laboratory, NSLS-II, Upton, New York 11973, USA. <sup>29</sup>Max-Planck Institute for the Structure and Dynamics of Matter, CFEL, Hamburg 22607, Germany.