Femtosecond free-electron laser x-ray diffraction data sets for algorithm development

Stephan Kassemeyer,[1,2] Jan Steinbrener,[1,2] Lukas Lomb,[1,2] Elisabeth Hartmann,[1] Andrew Aquila,[3] Anton Barty,[3] Andrew V. Martin,[3] Christina Y. Hampton,[4] Saša Bajt,[5] Miriam Barthelmess,[5] Thomas R.M. Barends,[1,2] Christoph Bostedt,[6] Mario Bott,[1,2] John D. Bozek,[6] Nicola Coppola,[3,16] Max Cryle,[1] Daniel P. DePonte,[3] R. Bruce Doak,[7] Sascha W. Epp,[2,8] Benjamin Erk,[2,8] Holger Fleckenstein,[3] Lutz Foucar,[1,2] Heinz Graafsma,[5] Lars Gumprecht,[3] Andreas Hartmann,[9] Robert Hartmann,[9] Günter Hauser,[10,11] Helmut Hirsemann,[5] André Hömke,[2,8] Peter Holl,[9] Olof Jönsson,[12] Nils Kimmel,[10, 11] Faton Krasniqi,[1,2] Mengning Liang,[3] Filipe R.N.C. Maia,[13] Stefano Marchesini,[13] Karol Nass,[3] Christian Reich,[9] Daniel Rolles,[1,2] Benedikt Rudek,[2,8] Artem Rudenko,[2,8] Carlo Schmidt,[2,8] Joachim Schulz,[3] Robert L. Shoeman,[1,2] Raymond G. Sierra,[4] Heike Soltau,[9] John C. H. Spence,[7] Dmitri Starodub,[4] Francesco Stellato,[3] Stephan Stern,[3] Gunter Stier,[1] Martin Svenda,[12] Georg Weidenspointner,[10,11] Uwe Weierstall,[7] Thomas A. White,[3] Cornelia Wunderer,[5] Matthias Frank,[14] Henry N. Chapman,[3,15] Joachim Ullrich,[2,8] Lothar Strüder,[2,10] Michael J. Bogan,[4] and Ilme Schlichting[1,2,*]

[1]Max-Planck-Institut für medizinische Forschung, Jahnstr. 29, 69120 Heidelberg, Germany
[2]Max Planck Advanced Study Group, Center for Free Electron Laser Science (CFEL), Notkestrasse 85, 22607 Hamburg, Germany
[3]Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany
[4]Stanford PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road. Menlo Park, California 94025, USA
[5]Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany
[6]LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road. Menlo Park, California 94025, USA
[7]Department of Physics, Arizona State University, Tempe, Arizona 85287 USA
[8]Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany
[9]PNSensor GmbH, Otto-Hahn-Ring 6, 81739 München, Germany
[10]Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany
[11]Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany
[12]Uppsala University, Department of Cell and Molecular Biology, Husargatan 3, 75124 Uppsala, Sweden
[13]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[14]Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA
[15]University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany
[16]Present address: European XFEL GmbH, Albert-Einstein-Ring 19, 22761 Hamburg, Germany
[*]ilme.schlichting@mpimf-heidelberg.mpg.de

Abstract: We describe femtosecond X-ray diffraction data sets of viruses and nanoparticles collected at the Linac Coherent Light Source. The data establish the first large benchmark data sets for coherent diffraction methods freely available to the public, to bolster the development of algorithms that are essential for developing this novel approach as a useful imaging technique. Applications are 2D reconstructions, orientation classification and finally 3D imaging by assembling 2D patterns into a 3D diffraction volume.

OCIS codes: (170.7160) Ultrafast technology; (140.2600) Free-electron lasers (FELs); (110.7440) X-ray imaging; (290.5850) Scattering, particles; (320.7100) Ultrafast measurements.

References and links

1. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, "Potential for biomolecular imaging with femtosecond X-ray pulses," Nature **406**(6797), 752–757 (2000).
2. P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, Y. Ding, D. Dowell, S. Edstrom, A. Fisher, J. Frisch, S. Gilevich, J. Hastings, G. Hays, Ph. Hering, Z. Huang, R. Iverson, H. Loos, M. Messerschmidt, A. Miahnahri, S. Moeller, H.-D. Nuhn, G. Pile, D. Ratner, J. Rzepiela, D. Schultz, T. Smith, P. Stefan, H. Tompkins, J. Turner, J. Welch, W. White, J. Wu, G. Yocky, and J.

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

Galayda, "First lasing and operation of an Ångström-wavelength free-electron laser," Nat. Photonics **4**(9), 641–647 (2010).
3. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, "Femtosecond X-ray protein nanocrystallography," Nature **470**(7332), 73–77 (2011).
4. M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, M. Hantke, D. P. DePonte, A. Barty, J. Schulz, L. Gumprecht, N. Coppola, A. Aquila, M. Liang, T. A. White, A. Martin, C. Caleman, S. Stern, C. Abergel, V. Seltzer, J. M. Claverie, C. Bostedt, J. D. Bozek, S. Boutet, A. A. Miahnahri, M. Messerschmidt, J. Krzywinski, G. Williams, K. O. Hodgson, M. J. Bogan, C. Y. Hampton, R. G. Sierra, D. Starodub, I. Andersson, S. Bajt, M. Barthelmess, J. C. Spence, P. Fromme, U. Weierstall, R. Kirian, M. Hunter, R. B. Doak, S. Marchesini, S. P. Hau-Riege, M. Frank, R. L. Shoeman, L. Lomb, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, C. Schmidt, L. Foucar, N. Kimmel, P. Holl, B. Rudek, B. Erk, A. Hömke, C. Reich, D. Pietschner, G. Weidenspointner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, I. Schlichting, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kühnel, R. Andritschke, C. D. Schröter, F. Krasniqi, M. Bott, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, H. N. Chapman, and J. Hajdu, "Single mimivirus particles intercepted and imaged with an X-ray laser," Nature **470**(7332), 78–81 (2011).
5. M. J. Bogan, W. H. Benner, S. Boutet, U. Rohner, M. Frank, A. Barty, M. M. Seibert, F. Maia, S. Marchesini, S. Bajt, B. Woods, V. Riot, S. P. Hau-Riege, M. Svenda, E. Marklund, E. Spiller, J. Hajdu, and H. N. Chapman, "Single particle X-ray diffractive imaging," Nano Lett. **8**(1), 310–316 (2008).
6. D. P. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. C. H. Spence, and R. B. Doak, "Gas dynamic virtual nozzle for generation of microscopic droplet streams," J. Phys. D Appl. Phys. **41**(19), 195505 (2008).
7. L. Strüder, S. Epp, D. Rolles, R. Hartmann, P. Holl, G. Lutz, H. Soltau, R. Eckart, C. Reich, K. Heinzinger, C. Thamm, A. Rudenko, F. Krasniqi, K. Kühnel, C. Bauer, C. Schröter, R. Moshammer, S. Techert, D. Miessner, M. Porro, O. Hälker, N. Meidinger, N. Kimmel, R. Andritschke, F. Schopper, G. Weidenspointner, A. Ziegler, D. Pietschner, S. Herrmann, U. Pietsch, A. Walenta, W. Leitenberger, C. Bostedt, T. Möller, D. Rupp, M. Adolph, H. Graafsma, H. Hirsemann, K. Gärtner, R. Richter, L. Foucar, R. L. Shoeman, I. Schlichting, and J. Ullrich, "Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources," Nucl. Instrum. Methods Phys. Res. A **614**(3), 483–496 (2010).
8. J. D. Bozek, "AMO instrumentation for the LCLS X-ray FEL," Eur. Phys. J. Spec. Top. **169**(1), 129–132 (2009).
9. J. L. Van Etten, D. E. Burbank, Y. Xia, and R. H. Meints, "Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae," Virology **126**(1), 117–125 (1983).
10. W. H. Benner, M. J. Bogan, U. Rohner, S. Boutet, B. Woods, and M. Frank, "Nondestructive characterization and alignment of aerodynamically focused particle beams using single particle charge detection," J. Aerosol Sci. **39**(11), 917–928 (2008).
11. M. Altarelli, R. Kurta, and I. A. Vartanyants, "X-ray cross-correlation analysis and local symmetries of disordered systems: General theory," Phys. Rev. B **82**(10), 104207 (2010).
12. D. K. Saldin, H. C. Poon, M. J. Bogan, S. Marchesini, D. A. Shapiro, R. A. Kirian, U. Weierstall, and J. C. Spence, "New light on disordered ensembles: ab initio structure determination of one particle from scattering fluctuations of many copies," Phys. Rev. Lett. **106**(11), 115501 (2011).
13. C. H. Yoon, P. Schwander, C. Abergel, I. Andersson, J. Andreasson, A. Aquila, S. Bajt, M. Barthelmess, A. Barty, M. J. Bogan, C. Bostedt, J. Bozek, H. N. Chapman, J. M. Claverie, N. Coppola, D. P. DePonte, T. Ekeberg, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, H. Graafsma, L. Gumprecht, J. Hajdu, C. Y. Hampton, A. Hartmann, E. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Kiskinova, M. Liang, N. T. Loh, L. Lomb, F. R. Maia, A. V. Martin, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, I. Schlichting, J. Schulz, M. Seibert, V. Seltzer, R. L. Shoeman, R. G. Sierra, H. Soltau, D. Starodub, J. Steinbrener, G. Stier, L. Strüder, M. Svenda, J. Ullrich, G. Weidenspointner, T. A. White, C. Wunderer, and A. Ourmazd, "Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering," Opt. Express **19**(17), 16542–16549 (2011).
14. A. V. Martin, J. Andreasson, A. Aquila, S. Bajt, T. R. M. Barends, M. Barthelmess, A. Barty, W. H. Benner, C. Bostedt, J. D. Bozek, P. Bucksbaum, C. Caleman, N. Coppola, D. P. DePonte, T. Ekeberg, S. W. Epp, B. Erk, G. R. Farquar, H. Fleckenstein, L. Foucar, M. Frank, L. Gumprecht, C. Y. Hampton, M. Hantke, A. Hartmann, E. Hartmann, R. Hartmann, S. P. Hau-Riege, G. Hauser, P. Holl, A. Hoemke, O. Jönsson, S. Kassemeyer, N. Kimmel, M. Kiskinova, F. Krasniqi, J. Krzywinski, M. Liang, N.-T. D. Loh, L. Lomb, F. R. N. C. Maia, S. Marchesini, M. Messerschmidt, K. Nass, D. Odic, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, C. Schmidt, J. Schultz, M. M. Seibert, R. L. Shoeman, R. G. Sierra, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, M. Svenda, H. Tobias, J. Ullrich, G. Weidenspointner, D. Westphal, T. A. White, G.

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

Williams, J. Hajdu, I. Schlichting, M. J. Bogan, and H. N. Chapman, "Single particle imaging with soft X-rays at the Linac Coherent Light Source," Proc. SPIE **8078**(807809), 807809 (2011).
15. A. Belabbaci, A. Razzouk, I. Mokbel, J. Jose, and L. Negadi, "Isothermal Vapor-Liquid Equilibria of (Monoethanolamine + Water) and (4-Methylmorpholine + Water) Binary Systems at Several Temperatures," J. Chem. Eng. Data **54**(8), 2312–2316 (2009).
16. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Probl. **21**(1), 37–50 (2005).
17. S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," Phys. Rev. B **68**(14), 140101 (2003).
18. M. J. Bogan, S. Boutet, A. Barty, W. H. Benner, M. Frank, L. Lomb, R. L. Shoeman, D. Starodub, M. M. Seibert, S. P. Hau-Riege, B. Woods, P. Decorwin-Martin, S. Bajt, J. Schulz, U. Rohner, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, J. Hajdu, and H. Chapman, "Single-shot femtosecond X-ray diffraction from ellipsoidal nanoparticles in random orientations," Phys. Rev. Special Topics - Accelerators and Beams **13**, 94791 (2010).
19. N. D. Loh, M. J. Bogan, V. Elser, A. Barty, S. Boutet, S. Bajt, J. Hajdu, T. Ekeberg, F. R. N. C. Maia, J. Schulz, M. M. Seibert, B. Iwan, N. Timneanu, S. Marchesini, I. Schlichting, R. L. Shoeman, L. Lomb, M. Frank, M. Liang, and H. N. Chapman, "Cryptotomography: reconstructing 3D Fourier intensities from randomly oriented single-shot diffraction patterns," Phys. Rev. Lett. **104**(22), 225501 (2010).
20. N. T. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **80**(2), 026705 (2009).

1. Introduction

X-ray free-electron lasers (XFELs) exceed the peak brilliance of conventional synchrotrons by almost a factor of 10 billion. It has been proposed that radiation damage, which limits the high resolution imaging of soft condensed matter, can be "outrun" by using ultrafast and extremely intense X-ray pulses that pass the sample before the onset of significant radiation damage [1]. Thus, one of the most promising scientific applications of XFELs is in sub-nanometer resolution imaging of biological objects, including cells, viruses, macromolecular assemblies, and nanocrystals. The concept of "diffraction-before-destruction" has been demonstrated recently at the Linac Coherent Light Source (LCLS) [2], the first operational hard X-ray FEL, for protein micro- and nanocrystals [3] and single mimivirus particles [4]. Since the enabling technologies, the XFEL itself [2], sample injection [5, 6], and fast-framing integrating X-ray detectors [7] are recent developments, it is of paramount importance to understand their capacity and limitations in delivering the data sets required for reliable sub-nanometer three-dimensional bio-imaging. Algorithms that assemble 2D diffraction data of randomly orientated molecules into a 3D volume require a highly homogeneous data set and a detailed understanding of measurement errors. Thus, sorting and orientation/classification algorithms need to be further developed to handle current experimental conditions and identify suitable data sets. Due to their mode of operation, experiments at XFELs are limited by scarcity of beam time, thus to date largely restricting the testing of algorithms to simulated data.

Here, we report on femtosecond coherent diffraction imaging experiments on model systems that differ in size, symmetry, and complexity. Data were collected from *Paramecium bursarium Chlorella* virus (PBCV-1), bacteriophage T4, and nanorice, an ellipsoidal (~250 x 50 nm) iron oxide nanoparticle which serves as a strongly scattering morphological analogue of the T4 tail. The experiments were carried out at the LCLS at the Atomic, Molecular Optical Science (AMO) beamline [8] in the CFEL-ASG Multi-purpose (CAMP) instrument [7], in a similar manner as described previously [4] and detailed below. Aerosolized particles were injected into the FEL interaction region using an aerodynamic lens stack [5] and diffraction patterns were collected on two pairs of pnCCD detectors [7] (Fig. 1A).

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

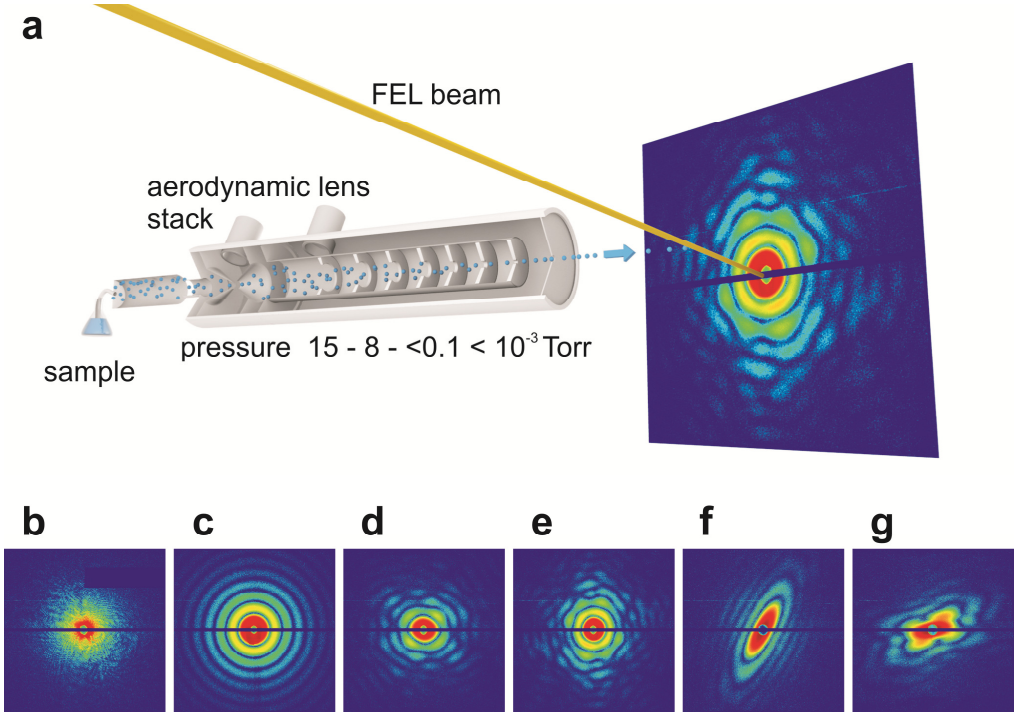

Fig. 1. (a) Experimental setup. Aerosolized particles are injected into the FEL beam in random, unknown orientations using an aerodynamic lens stack [5]. The diffraction patterns are captured with a set of pnCCD detectors [7]. (b-g) Diffraction patterns of different samples and byproducts of the injection. In addition to single particles, solvent droplets, sample aggregates or multiple particles are also recorded. Shown are diffraction patterns of a large aggregate (b), a water droplet (c), single T4 phage particles (d, e), a nanorice grain (f), and two nanogrrains (g).

The diffraction data were deposited in the Coherent X-ray Imaging Data Bank (www.CXIDB.org). By making the data publicly available, we provide a rich resource for testing the performance of existing algorithms on real data from known samples, which should help drive the development of improved or new algorithms and identify the next steps toward fulfilling the potential of 3D bio-imaging at XFELs.

2. Materials and methods

2.1 Samples

Nanorice (SiO_2 coated Fe_2O_3 ellipsoids) was purchased from Corpuscular, Inc. (Cold Spring Harbor, NY, USA). It was supplied as a 50% ethanol suspension at a concentration of 6.25 x 10^{12} particles per ml. Transmission electron microscopy (TEM) (Fig. 2) demonstrated some variability in the size and shape of the individual nanorice grains. Prior to injection into the FEL beam, the sample was sonicated for 5 minutes to break down larger clusters of particles. T4 and PBCV-1 [9] were purified as described previously. T4 was dialyzed against 50 mM ammonium acetate pH 7.2 long before the measurements whereas buffer exchange for PBCV-1 was done just prior to the experiments. Measurements of the virus preparations via NanoTracking Analysis with a Nanosight LM10-HS sizing system (Schaefer Technologie, Langen, Germany) demonstrated a largely monodisperse distribution of particles in solution with approximately 90% having the expected sizes for both T4 (mean 147 nm, FWHM of 42.2 nm) and PBCV-1 (mean 194 nm, FWHM of 82.2 nm); about 10% of the particles had a larger size corresponding to a clear dimer peak. Samples were analyzed by TEM (Fig. 2) using a Zeiss EM912 microscope running at 120 kV and equipped with a 1024 × 1024 pixel

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

GATAN CCD camera. PBCV-1 and T4 samples were negatively stained using 1-2% uranyl acetate.

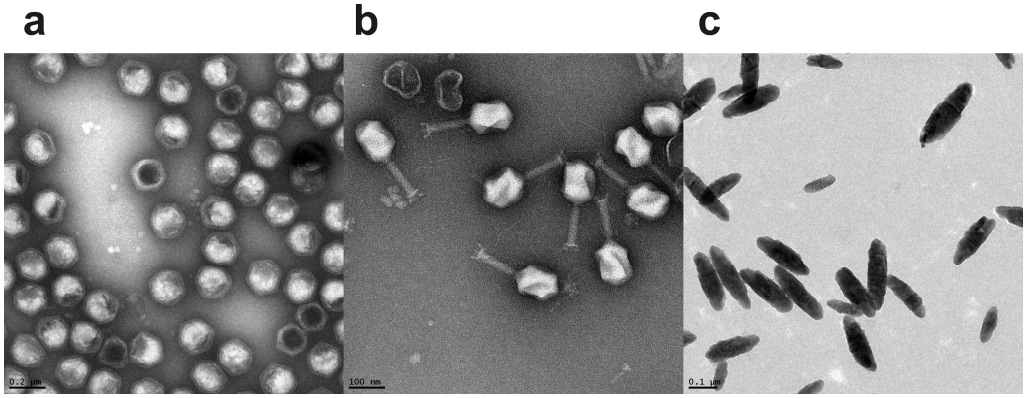

Fig. 2. TEM images of (a) PBCV-1, (b) bacteriophage T4, and (c) and nanorice.

2.2 Experimental setup/injection

The experiments were carried out at the LCLS at the AMO beamline [8] in the CAMP instrument [7] using 150 fs X-ray pulses with a photon energy of 1.2 keV and 2 keV, and with a pulse energy of 3.2 mJ and 2.7 mJ, respectively. Particles were injected as a continuous aerosol beam into the 10 µm2 X-ray focus using an aerodynamic lens stack [5]. Diffraction patterns were collected on two pairs of pnCCD detectors [7] that were read out with the FEL repetition rate of 60 Hz. The front detector, initially placed at a distance of ~150 mm and moved to the furthest possible downstream position, assumed to be ~250 mm [7] in the course of the beam time, was shielded by a 3 µm thick polyimide filter to prevent contamination. The back detector was placed 738 mm downstream of the interaction region. Front and back detector were operated such that a 1 eV photon corresponds to 0.01031 analog-to-digital units

Purified T4 and PBCV1 samples (~5 x 10^{11} particles per ml) were transferred into a volatile buffer (50 mM ammonium acetate pH 7.2) and the suspension was aerosolized with a gas dynamic virtual nozzle [6] or a commercial nebulizer (Burgener Mira Mist CE nebulizer, AHF Analysentechnik, Tübingen, Germany) using constant liquid flow rates between 10 – 20 µl/min (from a Shimadzu HPLC system) and nitrogen gas pressure of 5 to 7 bar. In the latter case, the droplet distribution was polydisperse with an estimated size range from several hundred nanometers to several micrometers. During the injection, much of the surrounding volatile buffer evaporates, although this seemed somewhat sample dependent.

The divergence of the particle beam exiting the aerodynamic lens stack was 0.57°, which results in a particle beam diameter of approximately 440 µm at the X-ray interaction region (22 mm from the lens stack exit). A "brim" of particles at a much lower concentration that formed a 2 mm diameter halo was also observed. Previous measurements of spherical particles (98-190 nm diameter) indicate that particle speed inside the vacuum is on the order of 100 m/s [10]. Neglecting the halo, and assuming that only single particles form the aerosol and that the transmission through the system is 100%, the probability to have a particle in the interaction region can be approximated. Having a concentration of 5 x 10^{11} particles per ml as for our virus samples and a typical flow rate of 15 µl/min, 1.25 x 10^{8} particles are transferred into vacuum per second. With a speed of 100 m/s, the particles fly across the 3 µm FEL focus in 30 ns. Due to the much larger diameter of the particle beam focus of 440 µm, only a fraction of the particles injected (0.7%) cross the FEL focus. This results in a maximal achievable hit rate of:

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

$$\frac{concentration \times flowrate \times Xray\_focus}{(aerodynamic\_lens\_focus / Xray\_focus) \times speed} \approx \frac{5 \cdot 10^{11} / ml \times \frac{15 \cdot 10^{-3}}{60} \frac{ml}{s} \times 3 \cdot 10^{-6} m}{\frac{440}{3} \times 100 m/s}$$

$\approx 2.6\%$

Big droplets quite often contain more than one single particle. During the evaporation process these particles agglomerate into clusters. Additionally, biological samples can already aggregate in solution, in particular at the high concentrations required for efficient data collection. Both of these effects reduce the probability of a good hit and lead to diffraction patterns of bigger clusters that have to be identified in the analysis process.

2.3 Pre-processing of diffraction images

Images were pre-processed to remove the following artifacts that are a property of the experimental system (Barty unpublished): (i) The pnCCD detectors have multiple readout channels in order to achieve in excess of 120 Hz readout speed, and the offsets on these readout channels drift slowly over time. Such fluctuations are corrected by periodically collecting data with no photons ('dark frames') and subtracting this from the raw data. (ii) The parallel readout results in gain fluctuations between different portions of the CCD. These are corrected for by using flat-field measurements obtained using an ^{55}Fe-source. (iii) Incoherent scattering from beamline optics and carrier gas in the aerodynamic lens is estimated from non-hit images that occurred before and after each hit. Each time a non-hit is identified, it is written to a buffer, replacing the oldest frame stored in the buffer, and the background recalculated. The recalculation of the estimated background accounts for fluctuations in background during the experiment. The median of the last 50 non-hits is subtracted from the signal in the hits as incoherent background. A median is used because it is less sensitive to accumulating signal from any extremely weak hits that fall below threshold. This approach additionally subtracts any remaining drifts in CCD readout offsets over time. (iv) Bad pixels are identified on the fly as those pixels that remain above a threshold of 140 ADU in more than 80% of sequential frames, and are set to zero. There are detector artifacts which are not corrected in the preprocessing: (i) charge spill to neighboring pixels for high intensities. (ii) The offset of each readout channel fluctuates independently on a shot-by-shot basis (referred to as common-mode noise).

3. Results and discussion

Particles entering the interaction volume (defined as focal area of the FEL beam times the longitudinal diameter of the particle beam) are intercepted randomly by the LCLS pulses. The hit rate depends on sample concentration and the overlap between the particle and the X-ray beams. Of ~5 million data frames collected during the experiment, 0.7% were classified as 'hits'. Hits were identified based on scattering strength by counting the number of pixels containing values above a predetermined analog to digital unit (ADU) threshold applied after background subtraction (>500 pixels above 170 ADU excluding bad pixels required to register as a 'hit'). This simple threshold approach is biased in favor of false positives, yet still yields over 99% rejection rate. "Hits" comprise diffraction patterns of single particles, of multiple particle clusters, of water drops, buffer aggregates, other false positives, detector glitches and hits too weak to be of use (Fig. 1B). Sample carry-over introduced at the aerosol source or from accumulated material within the aerodynamic lens stack commonly produces contamination from one run to the next. For most studies, only single particle hits of a single species are of interest, although correlation analysis can be used on multiple particle clusters [11, 12].

Particles were first sized based on dimensions of the particle autocorrelation, rejecting most of the outliers (Barty unpublished). Further distinction between samples such as nanorice, mimivirus and spherical water droplets was achieved by means of statistical learning methods operating on diffraction data with reduced dimensionality. Meaningful dimensions were selected by principal component analysis and an analysis of the diffraction

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

patterns' rotational symmetry and speckle size. Details will be published elsewhere. In addition, diffraction patterns have been classified in an unsupervised manner [13]. Lists of individual frames belonging to single classes of samples have also been included in the CXIDB deposition. These lists are not guaranteed to be perfect; however data for T4 and nanorice have been examined and edited manually. The pre-processed data (see Material and Methods for details) were deposited in the CXIDB for use with minimal post-processing, allowing development and testing of algorithms involved at different stages of the sorting and 3D-structure determination process. Raw data files are available on request.

Expected size distributions were not observed for all samples analyzed. The size of a particle can be determined from the inverse Fourier transform of its diffraction pattern (resulting in the autocorrelation of the object) if the geometry of the experiment is known. The maximum spatial frequency that is recorded on the detector is determined by the wavelength λ of the experiment and the maximum diffraction angle ϑ as $q_{\max} = \frac{\sin \vartheta}{\lambda}$, where $\sin \vartheta \approx \frac{N \Delta p}{2 Z_d}$ for small scattering angles, N is the number of pixels in the detector, Δp is the size of a pixel of the detector, and Z_d is the distance between detector and particle. According to the Nyquist theorem, the corresponding real space sampling is given as $\Delta x = \frac{1}{2 q_{\max}}$. A bounded object of diameter D will have an autocorrelation that is bounded to a diameter $2D$. Thus, one can determine the size of a spherical particle from the radius of its autocorrelation and the real space sampling interval Δx.

Particles can be intercepted at different positions along the beam propagation axis z in the interaction volume as well as at different locations in the approximately Gaussian-shaped intensity profile of the FEL focus. This will result in shot-to-shot variations of total diffracted intensity as well as a radial scaling of the diffracted intensities that is approximately linear with change in z-position. Due to the finite extent of the interaction volume along the direction of the X-ray beam, diffraction patterns will show a distribution of Z_d values leading to an apparent distribution in particle size whose center corresponds to the actual size and whose width is comparable to the width of the anticipated Z_d distribution. This effect explains the rather narrow size distribution observed *e.g.* in size selected polystyrene latex spheres (Loh unpublished) and mimivirus, which was identified as a member of a single class by classification algorithms [13] and showed the expected size of the strongly scattering capsid [4, 14]. We find this to be true for most samples except for small biological objects, like enterobacteria phage T4. Figure 3 shows the size distribution of a subset of diffraction patterns from T4 phages determined by applying a threshold to the autocorrelation. The size distribution is broad and the average size too large (mean 330 nm for T4 (head diameter ~90 nm, tail length ~100 nm). The width of the distribution far exceeds the anticipated spread of < 0.1% in Z_d values. (A particle hit 1cm away from the nominal interaction region would experience an apparent change in size of 14%.). This size increase may be caused by nonvolatile components such as protein fragments or salts or by a residual shell of ammonium acetate buffer that has not completely evaporated during aerosolization and injection and seems to have concentrated during the evaporation process, surrounding small particles. The reduced contrast at 1.2 keV between the partially evaporated, concentrated buffer and the virus leads to apparent particle sizes that are too large on average, with the size distribution

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

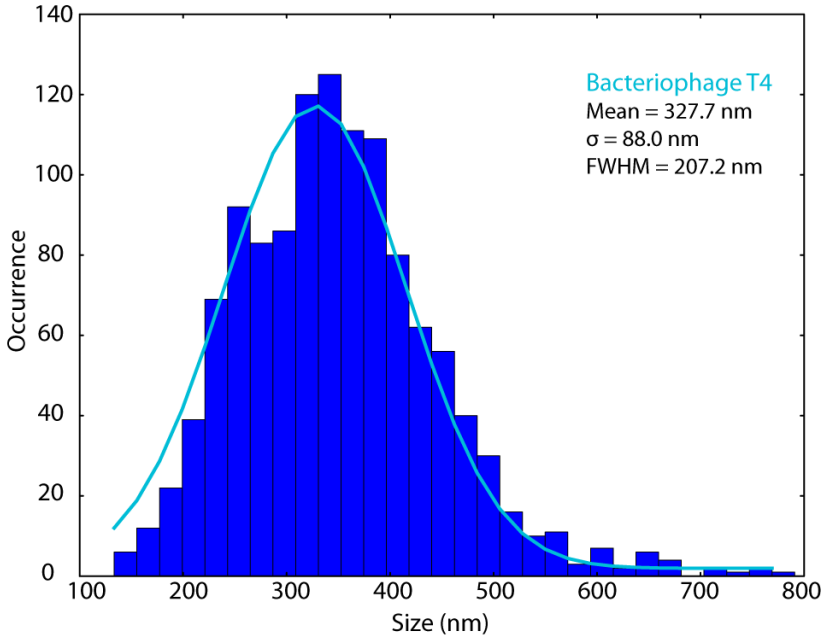

Fig. 3. Size distribution of a subset of recorded diffraction patterns of phage T4 determined by thresholding the autocorrelation. The apparent average size (~330 nm) is about 1.5-3 times larger than the actual particle size (~100 nm (strongly diffracting head) – ~200 nm (head and tail)). The width of the distribution far exceeds the anticipated spread in sample-to-detector distance values (< 0.1%).

reflecting that of the initial aerosol rather than actual particle size. Aerosol generation and surface tension may lead to a preferred aerosol droplet size coinciding with single large particles such as mimivirus [4, 14], a cluster of several smaller ones in the case of highly concentrated solutions, or a single smaller one in a solvent droplet. Thus, single smaller viruses are likely affected more by solvent shells and the consequential apparent change in size. In the future, the interplay of particle size, buffer composition and the aerosolizing process has to be adjusted carefully for each sample to work around this effect. Volatile buffers with a higher vapor pressure than ammonium acetate such as 4-methylmorpholine might be more suitable and are well tolerated by T4 and PBCV-1; 4-methylmorpholine also has the advantages of lowering the overall vapor pressure of the solution [15].

The inhomogeneity in the data collected on the T4 sample prevents a 3D reconstruction using the 2D patterns. Nevertheless, the quality of these single-shot diffraction patterns has proved sufficient for two dimensional imaging via phase retrieval. Numerous diffraction patterns from this experiment have been successfully phased [14]. Some examples of reconstructed images of the T4 bacteriophage are shown in Fig. 4. Reconstructions were performed with the Relaxed-Alternating-Averaged-Reflections algorithm [16]. The only constraint in the object plane was the support constraint. Each reconstruction in Fig. 4 is an average of 10 reconstructions from random-starting phase, each of 2000 iterations. In the region where data was not measured, the initial wave function for each reconstruction was set to the Fourier transform of the initial support, which is aimed to improve the rate of convergence. The Shrinkwrap algorithm [17] was used to refine the support during the reconstruction. The full-period resolution of these reconstructed images lies in the range 20-40 nm. Phasing is more reliable when the effective size of the central hole relative to the imaged object is such that the number of missing speckles is small, facilitating the unambiguous reconstruction of the missing low frequency data in the iterative process. Under our experimental conditions, this is true for objects up to 500 nm in diameter. The classification of diffraction patterns may identify those that can be phased with similar initial

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

parameters, opening up the possibility of automated phasing, an important development for processing large volumes of FEL data.

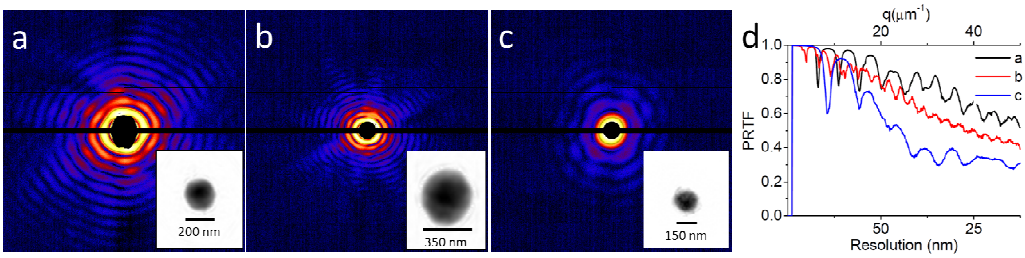

Fig. 4. (a-b: Diffraction patterns and reconstructed images of the T4 bacteriophage. (d): The phase retrieval transfer functions (PRTFs) for the reconstructions shown in (a-c). The full-period resolution is estimated where the PRTF falls to 0.5 and varies between 20 and 40 nm.

Nanorice provided the first three-dimensional reconstruction of a non-crystalline object from randomly oriented, continuous diffraction patterns captured with an X-ray laser [18, 19] using an expectation-maximization algorithm [20]. The study was based on 56 diffraction patterns obtained at the Free-electron Laser in Hamburg, FLASH, using a wavelength of 7 nm [18]. The reconstruction was challenging due to missing spatial frequencies caused by saturation effects of the detector and background scatter [19]. The nanorice data presented here do not suffer from these limitations (Fig. 5), sample the 3D diffraction space in a finer manner, and extend to higher resolution, which will allow different algorithms for 3D reconstructions to be tested. Around 1000 useful nanorice diffraction patterns have been found by combining hit identification, statistical learning and manual selection. A list is provided at CXIDB.

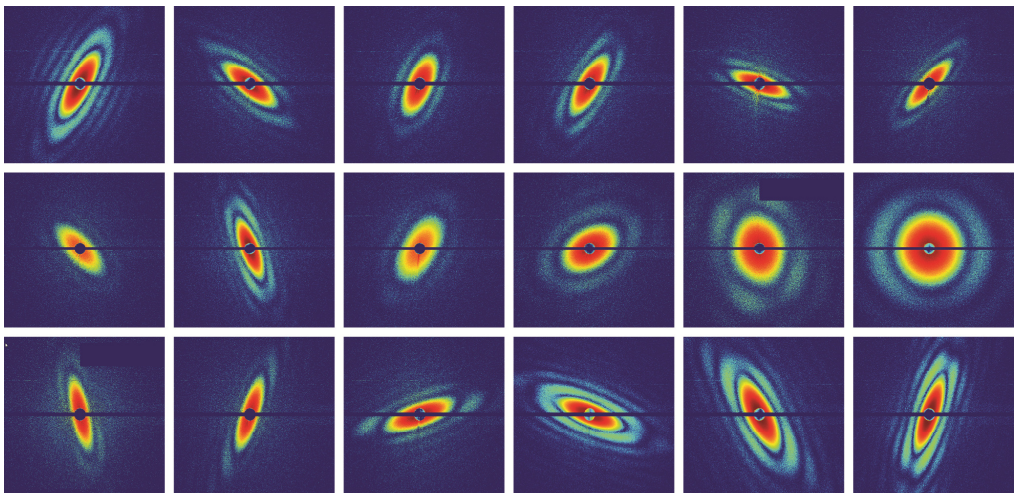

Fig. 5. A selection of nanorice diffraction patterns. They show individual nanorice particles in a variety of orientations (sideways at various angles and some more head-on), exposed to different X-ray fluences.

4. Conclusion

We present the experimental details, some caveats, and the recording of large volumes of high quality diffraction data from femtosecond XFEL pulses together with the successful sorting of data and the phasing of single-shot diffraction patterns obtained from single-particle-imaging experiments at the LCLS using model systems. The data is deposited in the CXIDB database where it is available for download and further analysis. The data has already been used to test a spectral clustering based unsupervised method able to sort experimental snapshots without

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012

recourse to templates, specific noise models, or user-directed learning [13] and we expect it to help drive further developments of improved or new algorithms. Publications resulting from this data should cite this paper and acknowledge the CXIDB.

Acknowledgments

Experiments were carried out at the Linac Coherent Light Source, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. We acknowledge support from the Max Planck Society for funding the development and operation of the CAMP instrument within the ASG at CFEL, the Helmholtz Association, the U.S. Department of Energy through the PULSE Institute at the SLAC National Accelerator Laboratory and Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, the US National Science Foundation (awards MCB 0919195 and 1120997), the Joachim Herz Stiftung, and the Petascale Initiative in Computational Science at NERSC, the Director, Office of Science, Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We are grateful to Christina Wege, Manuela Gebhardt, Gerhard Thiel, and James van Etten for help and advice concerning the viruses. We especially thank the staff of the LCLS for their outstanding facility and support in carrying out these experiments.

#159841 - $15.00 USD Received 12 Dec 2011; revised 15 Jan 2012; accepted 16 Jan 2012; published 3 Feb 2012