Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser

Tomas Ekeberg,[1,*] Martin Svenda,[1] Chantal Abergel,[2] Filipe R. N. C. Maia,[1,3] Virginie Seltzer,[2] Jean-Michel Claverie,[2] Max Hantke,[1] Olof Jönsson,[1] Carl Nettelblad,[1] Gijs van der Schot,[1] Mengning Liang,[4] Daniel P. DePonte,[4] Anton Barty,[4] M. Marvin Seibert,[1,5] Bianca Iwan,[1,6] Inger Andersson,[1] N. Duane Loh,[7] Andrew V. Martin,[8] Henry Chapman,[4,9] Christoph Bostedt,[5] John D. Bozek,[5] Ken R. Ferguson,[5] Jacek Krzywinski,[5] Sascha W. Epp,[10] Daniel Rolles,[10,11] Artem Rudenko,[11] Robert Hartmann,[12] Nils Kimmel,[13,14] and Janos Hajdu[1,15]

[1]*Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden*
[2]*Génomique & Structurale - IGS - UMR 7256, CNRS, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée, Parc Scientifique de Luminy, Case 934, 13288 Marseille Cedex 9, France*
[3]*NERSC, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[4]*Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany*
[5]*LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*
[6]*Attophysics Group, CEA-Saclay, 91191 Gif sur Yvette Cedex, France*
[7]*Centre for BioImaging Sciences, National University of Singapore, 14 Science Drive 4 Blk S1 A, Singapore 117546, Singapore*
[8]*The University of Melbourne, Parkville, 3010 Victoria, Australia*
[9]*University of Hamburg, Notkestrasse 85, 22607 Hamburg, Germany*
[10]*Max Planck Advanced Study Group, Center for Free Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany*
[11]*J. R. Macdonald Laboratory, Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, Kansas 66506, USA*
[12]*PNSensor GmbH, Römmerstrasse 28, 80803 München, Germany*
[13]*Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany*
[14]*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany*
[15]*European XFEL, Albert-Einstein-Ring 19, 22761 Hamburg, Germany*

(Received 3 October 2014; published 2 March 2015)

We present a proof-of-concept three-dimensional reconstruction of the giant mimivirus particle from experimentally measured diffraction patterns from an x-ray free-electron laser. Three-dimensional imaging requires the assembly of many two-dimensional patterns into an internally consistent Fourier volume. Since each particle is randomly oriented when exposed to the x-ray pulse, relative orientations have to be retrieved from the diffraction data alone. We achieve this with a modified version of the expand, maximize and compress algorithm and validate our result using new methods.

DOI: 10.1103/PhysRevLett.114.098102 PACS numbers: 42.30.-d, 87.64.-t, 87.64.Bx

Introduction.—Free-electron lasers provide femtosecond x-ray pulses with a peak brilliance ten billion times higher than any previously available x-ray source. Such a large jump in one physical quantity is very rare and can have far-reaching implications for several areas of science. It has been suggested that such pulses could outrun key damage processes and allow structure determination without the need for crystallization [1]. In 2006 came the first verification of this "diffraction before destruction" principle with the reconstruction of a silicon nitride nanostructure created with a focused ion beam (FIB) and exposed to the FLASH free-electron laser in Hamburg [2].

So far, imaging applications at free-electron lasers (FELs) have mainly been limited to nanocrystallography and to two-dimensional projections of single particles, while 3D reconstructions from single particles have remained elusive.

Nanocrystallography [1,3] is an extension to protein crystallography where the high intensity and short pulse duration of a FEL allow for the use of very small crystals. Some proteins only produce small crystals. However, the fundamental problem that some samples are hard or impossible to crystallize is still valid. For this reason, single-particle imaging was a key part of the scientific case for building x-ray free-electron lasers.

Two-dimensional imaging with FELs such as the imaging of live cells [4,5], organelles [6] and viruses [7] is a promising method for imaging irreproducible samples. Resolutions down to 21 nm have been achieved on carboxysomes in a recent study [6]. There is also one application where a 2D image from a single-shot FEL experiment was compared to regular x-ray diffraction tomography performed at synchrotrons [8]. In a recent paper the structure of simple gold nanostructures was recovered in 3D from one single diffraction pattern [9]. This technique is, however, restricted to structurally simple and strongly scattering structures with a high degree of symmetry.

Several fundamental challenges exist for a general method of 3D single-particle imaging. First, 3D imaging

 © 2015 American Physical Society

requires the assembly of diffraction patterns from many identical copies of a reproducible object. Many of the applications of 2D imaging so far have been dealing with cells or other particles where each sample is structurally unique. Second, there is no way to directly measure the orientation of the sample when it is hit by the x-ray pulse. Instead, the orientation of each sample particle has to be recovered from the noisy signal of the diffraction patterns.

Solving these problems not only gives more information about the sample by presenting the structure in 3D; it is also a necessity for extending the signal from weakly scattering samples such as proteins and small viruses. For these samples the scattering from a single particle may be too weak for reconstructing a 2D projection image, and increasing the signal-to-noise ratio (SNR) by merging many patterns could allow for phasing even in this case.

A solution to the orientation problem was proposed by Duane Loh and Veit Elser with the expand, maximize and compress (EMC) algorithm [10] which was verified for simulated diffraction patterns in the original publication. Later the algorithm was also tested for an artificial sample [11]. This paper presents the first application of the algorithm to a biological sample.

We used mimivirus (*Acanthamoeba* polyphaga mimivirus) particles [12–14] in this study. The mimivirus is one of the largest known viruses. The viral capsid is about 450 nanometers in diameter and is covered by a layer of thin fibers. A 3D structure of the viral capsid exists [14], but the 3D structure of the inside is currently unknown.

Experimental Setup and Data Preprocessing.—Mimivirus particles were aerosolized and then focused to a narrow particle stream using an aerodynamic lens. The beam of particles was intersected with the pulse train of the Linac Coherent Light Source (LCLS). Diffracted light was collected on a detector placed 0.7 m downstream of the interaction region. At this distance the diffraction signal obeys the Fraunhofer approximation. Mimivirus particles that were not hit by the FEL were shown to remain infectious after the injection process, suggesting that they were not harmed by the injection process. A detailed description of the setup can be found in the Supplemental Material [15]. A total of 198 diffraction patterns were selected and preprocessed, and a subset of 25 of these is shown in Fig. 1. Pattern selection and preprocessing is explained in the Supplemental Material [15].

Orientation Recovery.—Three-dimensional structure determination requires the assembly of many 2D diffraction patterns into an internally consistent 3D Fourier volume. A diffraction pattern represents an Ewald-sphere slice through the 3D Fourier transform of the electron density. Since each particle is randomly orientated when exposed to the x-ray pulse, the relative orientations of the particles have to be retrieved from the diffraction data alone. This was done using a modified version of the EMC algorithm [10]. This algorithm has been verified for simulated data [10] and has been experimentally tested using artificial "nanorice" particles at a resolution that is too low to permit phase retrieval [11].

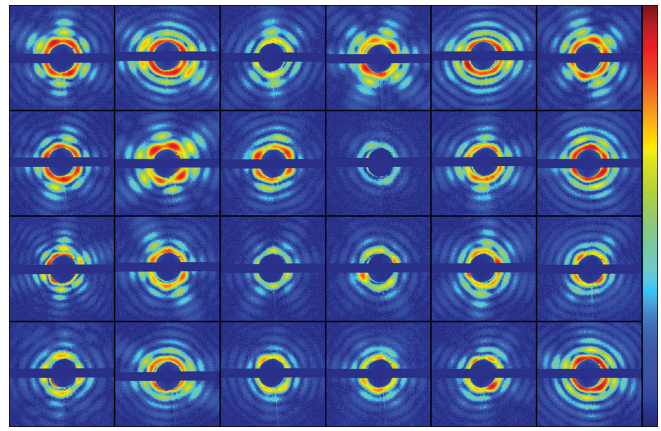

FIG. 1 (color). Shown are 24 of the 198 diffraction patterns that were included in the analysis. The central region is missing due to a hole in the detector that lets the beam through. Patterns were selected having signals beyond 83 nm^{-1} but not saturating the detector.

In the EMC algorithm a 3D diffraction space is iteratively updated to comply with the experimental data in the three steps: expand, maximize and compress. In the expand step the current diffraction space is expanded into tomograms by taking slices through the diffraction space at a discrete sampling of all rotations. In the maximize step all tomograms are compared to all experimental diffraction patterns by calculating the probability of detecting the experimental pattern while treating the tomograms as expectation values. New tomograms are then created by summing together all diffraction patterns weighted by the respective calculated probability. In the compress step a new 3D diffraction space is assembled from the new tomograms.

For this study we introduce a new similarity function in the maximize step that is based on a Gaussian model:

$$L(K,M) = \prod_i e^{-\frac{(M_i-K_i)^2}{2\sigma_i^2}} \tag{1}$$

where K is the diffraction pattern, M is the slice through the 3D diffraction space, i is the pixel index and σ_i is the standard deviation of the Gaussian. We set $\sigma_i = A\sqrt{M_i}$ where A is a constant. This similarity function balances well the contribution from the few high-intensity central pixels and the numerous low-intensity outer pixels.

The photon fluence at the particle is unknown in this type of experiment since neither the exact profile of the x-ray pulse nor the exact position of the particle in the beam is known [6]. The fluence therefore needs to be recovered from the diffraction pattern in the EMC process just like the orientation of the particles.

We used a variation of the method described in [11] with the following two key differences: (i) A new fluence is calculated for each comparison between a diffraction

pattern and a slice through Fourier space instead of using one fluence per pattern; (ii) the calculation of the fluence maximizes the likelihood function under the new distance metric given in Eq. (1). The fluence ϕ is thus given as

$$\phi(K, M) = \frac{\sum_i K_i^2 / M_i}{\sum_i K_i}. \tag{2}$$

Regions that lacked data, such as the beam stop area, had to be masked out for the analysis. We used a common mask for all diffraction patterns since the size and shape of the mask would otherwise bias the distance metric. The mask used was the union of the masks of the individual patterns.

Figure 2 shows the three-dimensional assembly of the diffraction patterns in the orientations recovered from the data and intensities properly scaled by the recovered fluence. The probability of achieving a full coverage of fourier space from 198 diffraction patterns is calculated in the Supplemental Material [15] to be 99.999991%. To verify this full coverage all slices were also assembled, giving each a thickness of one Shannon pixel. The assembled space contained no uncovered regions, meaning that the number of diffraction patterns was enough.

Phase Retrieval.—Noncrystalline objects produce oversampled diffraction patterns from which phases can be directly recovered in an iterative process [16] where two constraints are sequentially enforced. The first constraint is that the Fourier amplitudes have to be consistent with the collected data. The second constraint is to enforce a known upper size limit of the sample.

We use an advanced version of the above algorithm called the hybrid input output (HIO) algorithm [17] implemented in the HAWK software package [18] and enhanced by a positivity constraint [19]. The support was handled by a Shrinkwrap algorithm [19] with the constraint to have a specific area. The result was refined with 1000 iterations of the error reduction (ER) algorithm [16].

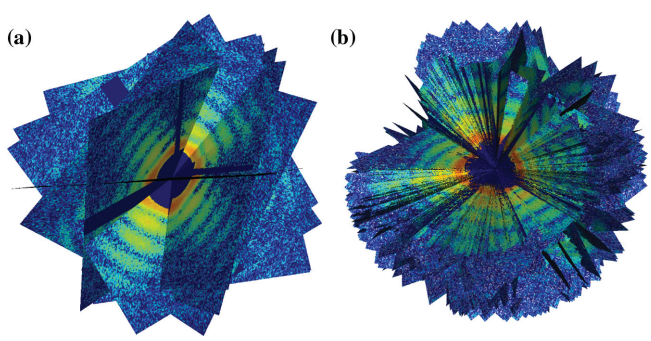

FIG. 2 (color). The assembled three-dimensional diffraction space. (a) The first ten patterns are shown in their best recovered orientations. Each diffraction pattern represents a slice through the squared modulus of the 3D Fourier transform of the electron density. (b) All 198 diffraction patterns plotted with a section cut out to show the central part of diffraction space. Diffraction symmetry and object symmetry can be directly recovered from the measured diffraction data in the EMC process.

The average Fourier error [20] was 0.019, and the average real-space error [20] was 0.0048. The reconstruction did not suffer from weakly constrained modes [21], meaning that the missing information in the center of the diffraction patterns could be completely recovered. This conclusion is based on an analysis method described in [7]. The iterative phase retrieval was repeated 200 times with independent random starting phases. Real-space error, Fourier-space error and unweighted pairgroup method with arithmetic mean (UPGMA) clustering show only one outlier. The average of the 199 successful and similar 3D reconstructions is shown in Fig. 3. No symmetry was imposed during the assembly of the 3D data set. Object symmetry was instead recovered from the measured diffraction data in the EMC process. The map reveals an asymmetric internal structure with a shift of density to one side of the particle along a pseudo-fivefold axis.

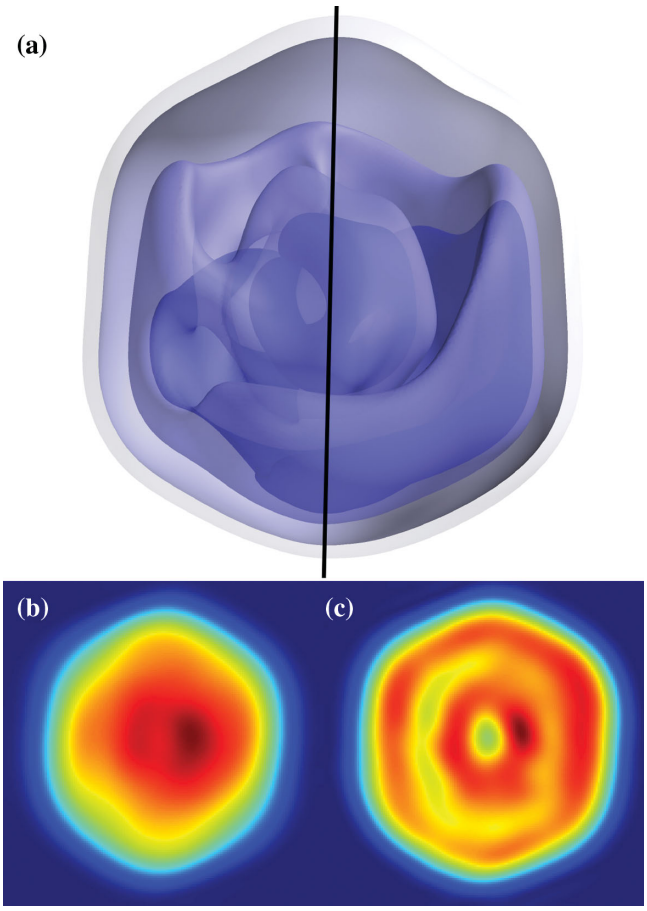

FIG. 3 (color). Reconstructed electron density. (a) The electron density of the mimivirus is recovered to a full-period resolution of 125 nm. The figure shows a series of isosurfaces, where blue represents denser regions and white represents lower density. The reconstruction shows a nonuniform internal structure, and the line indicates the pseudo-fivefold axis. (b) A projection image of the recovered electron density. (c) A slice through the center of the recovered electron density.

The resolution is estimated from the phase-retrieval transfer-function (PRTF) [21,22] which gives a full-period resolution of 125 nm [Fig. 4(a)]. As expected, the PRTF drops where the signal is low. This behavior also explains the oscillating nature of the PRTF that is common for nearly spherical objects.

Validation of Orientation Recovery.—Since this is the first 3D reconstruction from experimental data using the EMC algorithm, there were no validation methods available to assess the quality of our 3D orientation recovery. We have therefore developed two independent validation methods inspired by cryo-EM and x-ray crystallography.

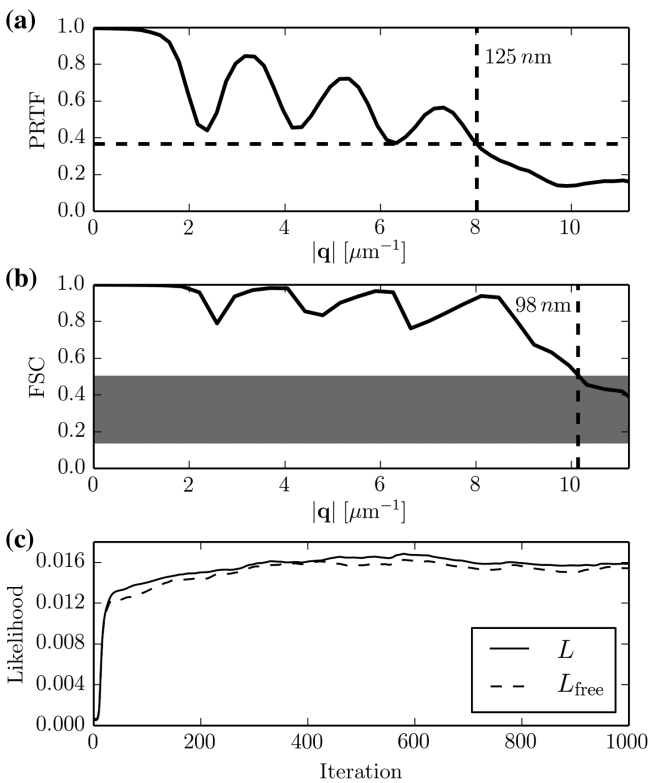

FIG. 4. (a) Phase-retrieval transfer function. From 200 independent phase retrievals the reproducibility of the phases was calculated as a PRTF. The resolution is estimated by the convention of applying a threshold to the PRTF at e^{-1} [2,7]. The full-period resolution is estimated as the inverse of the length of the scattering vector $\mathbf{q}$. By this measure we achieve a full-period resolution of 125 nm. (b) Fourier shell correlation validation. The data set was randomly split into two sets of equal size. Assembly and phase retrieval was performed independently on each set. A Fourier-shell correlation (FSC) was then calculated based on the two resulting density maps. Established thresholds for acceptable FSC values range between 0.14 and 0.5 (the gray area in the plot). Within our claimed resolution, the FSC stays well above all the commonly used thresholds. (c) Validation through the L_{free} similarity function. We excluded 10% of the patterns from the analysis. The best fit between these excluded patterns and the recovered model is plotted as a function of iteration (high values indicate a good fit). As a reference the same fit is plotted for the patterns included. The fact that L and L_{free} closely follow each other in (c) means that the data are not overfitted.

A standard validation method in cryo-EM is to randomly split the data and analyze each set independently [23]. A Fourier-shell correlation (FSC) is then calculated to quantify the differences between the two sets. We repeated the orientation and phase retrieval independently on two disjoint sets of 99 randomly selected diffraction patterns. It can be shown that the 99 diffraction patterns cover reciprocal space with more than 99% probability. After assembly and phase retrieval, we compared the two electron density maps and plotted the FSC [Fig. 4(b)]. Common thresholds for an acceptable FSC value range between 0.14 and 0.5 in cryo-EM literature [24,25]. The FSC for our two reconstructions stays well above these values, even beyond the resolution according to the PRTF. This shows that the assembly of the 3D data was correctly performed and that the recovery of the phases was accurate. The results also indicate that the 198 mimivirus particles used in this experiment were identical to the resolution of the reconstruction.

In x-ray crystallography experiments, a subset of the recorded Bragg peaks is often excluded from the analysis and only used for validating the result. If the recovered structure matches the excluded data to a similar degree as to which it matches the included data, one can conclude that the structure is not overfitted to the data [26]. Our second validation method is inspired by this analysis, but it differs from it in two important ways: (i) Entire diffraction patterns are excluded from the analysis instead of single Bragg peaks; (ii) the comparison is made between the omitted pattern and the recovered intensity distribution, based on the similarity function in Eq. (1). We excluded 10% of the diffraction patterns from the analysis and calculated the similarity function given in Eq. (1) for both the data included in the assembly, L, and for the data excluded from the assembly, L_{free}:

$$L = \frac{1}{N_{\mathrm{inc}}} \sum_{\{K_i;\mathrm{inc}\}} \max_{\{M_j\}} L(K_i, M_j), \tag{3}$$

$$L_{\mathrm{free}} = \frac{1}{N_{\mathrm{exc}}} \sum_{\{K_i;\mathrm{exc}\}} \max_{\{M_j\}} L(K_i, M_j). \tag{4}$$

Here, $\{K_i;\mathrm{inc}\}$ and $\{K_i;\mathrm{exc}\}$ are the sets of the included and excluded diffraction patterns, respectively, and N_{inc} and N_{exc} are the sizes of these sets. $\{M_j\}$ is the set of all model slices of the expanded model, and the function L is given in Eq. (1).

We calculate the average L and L_{free} values for 20 different random sets of excluded diffraction patterns and plot it in Fig. 4(c) as a function of iteration numbers in the EMC process. In contrast to the crystallographic measures R_{free} and R_{cryst}, high values of L and L_{free} indicate a good fit. The fact that L and L_{free} closely follow each other in Fig. 4(c) indicates that the data are not overfitted.

Conclusion.—In this article we have shown experimentally that 3D imaging of reproducible noncrystalline biological particles can indeed be performed at FELs. This was possible through an adapted version of the EMC algorithm. In this demonstration experiment we selected diffraction patterns with a high signal. We developed two validation methods for use in this type of analysis. The validity of our reconstruction is supported by both new methods as well as by validation through the phase retrieval transfer function.

The main factor limiting the resolution is the small number of diffraction patterns that were available for this study. New experiments already provide much higher hit rates thanks to an improved sample injector [6]. This suggests that the resolution can be significantly improved in future applications. Also, further development of the EMC algorithm could allow for the use of individual masks for each diffraction pattern. This would make it possible to use patterns with large saturated regions that are currently thrown away, thus using more of the collected data.

There are many important reproducible biological objects with sizes of 30–300 nm. Three important pathogenic viruses—HIV, influenza and herpes—are all in the 100–200 nm range. Furthermore, the EMC algorithm has been shown in simulation to be able to handle the much weaker signal strengths expected from single macromolecules or small viruses. It has been claimed that it is within the potential of free-electron lasers to image such objects at high resolution. This paper takes us one step closer to realizing this potential.

This work was supported by the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the European Research Council, the Röntgen-Ångström Cluster, and Stiftelsen Olle Engkvist Byggmästare. Portions of this research 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 thank Professor Ilme Schlichting for her help with data collection. We are grateful to the scientific and technical staff of the LCLS for support. We are indebted to the CAMP Collaboration for giving us access to their experimental setup and for supporting the experiment at the LCLS. We also acknowledge the Max Planck Society for funding the development and operation of the CAMP instrument, which benefitted many users at the LCLS. We thank the Institut de Biologie Structurale in Grenoble for electron microscopy.

*Corresponding author.
ekeberg@xray.bmc.uu.se

[1] R. Neutze, R. Wouts, and D. van der Spoel, Potential for biomolecular imaging with femtosecond X-ray pulses, Nature (London) **406**, 752 (2000).

[2] H. N. Chapman *et al.*, Femtosecond diffractive imaging with a soft-x-ray free-electron laser, Nat. Phys. **2**, 839 (2006).

[3] H. N. Chapman, P. Fromme, A. Barty *et al.*, Femtosecond X-ray protein nanocrystallography, Nature (London) **470**, 73 (2011).

[4] G. van der Schot *et al.*, Imaging single cells in a beam of live cyanobacteria with an x-ray laser, Nat. Commun. **6**, 5704 (2014).

[5] T. Kimura, Y. Joti, A. Shibuya *et al.*, Imaging live cell in micro-liquid enclosure by X-ray laser diffraction, Nat. Commun. **5**, 3052 (2014).

[6] M. F. Hantke, D. Hasse, F. R. N. C. Maia *et al.*, High-throughput imaging of heterogeneous cell organelles with an X-ray laser, Nat. Photonics **8**, 943 (2014).

[7] M. Marvin Seibert, T. Ekeberg, F. R. N. C. Maia *et al.*, Single mimivirus particles intercepted and imaged with an X-ray laser, Nature (London) **470**, 78 (2011).

[8] M. Gallagher-Jones, Y. Bessho, S. Kim *et al.*, Macromolecular structures probed by combining single-shot free-electron laser diffraction with synchrotron coherent X-ray imaging, Nat. Commun. **5**, 3798 (2014).

[9] R. Xu, H. Jiang, C. Song *et al.*, Single-shot three-dimensional structure determination of nanocrystals with femtosecond X-ray free-electron laser pulses, Nat. Commun. **5**, 4061 (2014).

[10] N.-T. Duane Loh and V. Elser, Reconstruction algorithm for single-particle diffraction imaging experiments, Phys. Rev. E **80**, 026705 (2009).

[11] N. D. Loh, M. J. Bogan, V. Elser *et al.*, Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns, Phys. Rev. Lett. **104**, 225501 (2010).

[12] N. Zauberman, Y. Mutsafi, D. Ben Halevy, E. Shimoni, E. Klein, C. Xiao, S. Sun, and A. Minsky, Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus, PLoS Biol. **6**, e114 (2008).

[13] D. Raoult, S. Audic, C. Robert *et al.*, The 1.2-megabase genome sequence of Mimivirus, Science **306**, 1344 (2004).

[14] C. Xiao, Y. G. Kuznetsov, S. Sun *et al.*, Structural studies of the giant mimivirus, PLoS Biol. **7**, e92 (2009).

[15] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.114.098102 for a description of the experimental setup in more detail and the preprocessing steps that were applied to the data. Furthermore the Supplemental Material presents a prediction of the number of diffraction patterns required for three-dimensional imaging.

[16] J. R. Fienup, Reconstruction of an object from the modulus of its Fourier transform, Opt. Lett. **3**, 27 (1978).

[17] J. R. Fienup, Phase retrieval algorithms: A comparison, Appl. Opt. **21**, 2758 (1982).

[18] F. R. N. C. Maia, T. Ekeberg, D. van der Spoel, and J. Hajdu, Hawk: The image reconstruction package for coherent X-ray diffractive imaging, J. Appl. Crystallogr. **43**, 1535 (2010).

[19] 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**, 140101 (2003).

[20] S. Marchesini, Invited article: A unified evaluation of iterative projection algorithms for phase retrieval, Rev. Sci. Instrum. **78**, 011301 (2007).

[21] D. Shapiro, P. Thibault, T. Beetz *et al.*, Biological imaging by soft x-ray diffraction microscopy, Proc. Natl. Acad. Sci. U.S.A. **102**, 15343 (2005).

[22] J. Steinbrener, J. Nelson, X. Huang, S. Marchesini, D. Shapiro, J. J. Turner, and C. Jacobsen, Data preparation and evaluation techniques for x-ray diffraction microscopy, Opt. Express **18**, 18598 (2010).

[23] G. Harauz and M. van Heel, Exact filters for general geometry three dimensional reconstruction, Optik (Stuttgart) **73**, 146 (1986).

[24] P. B. Rosenthal and R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy, J. Mol. Biol. **333**, 721 (2003).

[25] B. Böttcher, S. A. Wynne, and R. A. Crowther, Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy, Nature (London) **386**, 88 (1997).

[26] A. T. Brünger, Free R value: A novel statistical quantity for assessing the accuracy of crystal structures, Nature (London) **355**, 472 1992.