Serial electron diffraction for high-throughput protein nano-crystallography

Robert Bücker[1], Pascal Hogan-Lamarre[1,2], Pedram Mehrabi[1], Eike C. Schulz[1], Lindsey A. Bultema[1], Yaroslav Gevorkov[3,4], Wolfgang Brehm[3], Oleksandr Yefanov[3], Dominik Oberthür[3], Günther H. Kassier[1], and R. J. Dwayne Miller[1,2,*]

[1] Max Planck Institute for the Structure and Dynamics of Matter, CFEL, Luruper Chaussee 149, 22761 Hamburg, Germany

[2] Departments of Chemistry and Physics, University of Toronto, 80 St. George Street, Toronto, Ontario M5S 3H6, Canada

[3] Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany

[4] Institute of Vision Systems, Hamburg University of Technology, Harburger Schlossstrasse 20, 21079 Hamburg, Germany

* To whom correspondence should be addressed: dwayne.miller@mpsd.mpg.de

1 Sentence summary: Serial nano-beam electron diffraction (SerialNED) enables high-throughput, low dose protein crystallography from sub-micron crystals at conventional S/TEM microscopes.

Keywords: electron diffraction, serial crystallography, cryo-electron microscopy, protein structure

Abstract

We describe a new method for serial electron diffraction of protein nano-crystals using a conventional S/TEM microscope. Here, randomly dispersed crystals are mapped, and dose-efficient diffraction patterns measured at each identified position for structure determination. This fully automated workflow is suitable for high-throughput applications with acquisition rates of up to 1 kHz and a hit fraction approaching 100 percent. We demonstrate this method by solving the structure of lysozyme and crystalline granulovirus occlusion bodies to a resolution of 1.80 Å and 1.55 Å, respectively. This method promises to provide rapid high-quality structure determination for all classes of materials, with minimal sample consumption, using routinely available devices.

Introduction

An understanding of macromolecular structure is crucial for insight into complex biological systems. In spite of the advances in single-particle cryo-electron microscopy (cryo-EM), the vast majority of structures are determined by crystallographic methods (http://www.rcsb.org/stats/summary). This includes the majority of membrane proteins, which are often too small for computational alignment as required by single-particle analysis [1]. An important limitation lies in the difficulty to obtain large, well ordered crystals, which is particularly prevalent for membrane proteins and macromolecular complexes. Sub-micron crystals can be obtained more readily and are a common natural phenomenon, but often escape structure determination as the small diffracting volume and low tolerated dose of typically tens of MGy [2,3] prohibit the measurement of sufficient signal. However, during the past few years crystallographic techniques have emerged that are able to exploit nano-crystals for diffraction experiments. Notably, X-ray free-electron lasers (XFELs) have driven the development of *serial* crystallography [4]. Here, acquiring single snapshots from each crystal instead of a rotation series prevents dose accumulation, permitting higher fluences which concomitantly decreases the required diffracting volume. Ideally, radiation damage effects are entirely evaded either by imposing doses too low to cause significant structural damage of each crystal [5], or by a "diffract-before-destroy" mode using femtosecond XFEL pulses [4]. However, the scarcity of XFEL beamtime currently limits the use of protein nano-crystals for routine structure determination. Electron microscopes are a comparatively ubiquitous and cost-effective alternative for measuring diffraction from nano-crystals. While the low penetration depth of electrons renders them unsuitable for large three-dimensional crystals, their physical scattering properties are specifically advantageous for sub-micron crystals of radiation-sensitive materials. Compared to X-rays, the obtainable diffraction signal for a given crystal volume and tolerable radiation dose is up to three orders of magnitude larger, due to the higher ratio of elastic to inelastic electron scattering events, and a much smaller energy deposition per inelastic event [1]. While seminal experiments on 2D crystals [6] were restricted to a small class of suitable samples; diffraction from 3D nano-crystals has recently led to a resurgence of interest in biomolecular electron crystallography. Several research groups have succeeded in solving structures by merging electron diffraction data from as little as one, up to a few sub-micron sized vitrified protein crystals using rotation electron diffraction, analogous to conventional protein X-ray crystallography methods [7–9]. This technique, also known as micro-crystal electron diffraction (Micro-ED), is commonly used for structure determination of small organic or inorganic molecules [10–12]. However, only very recently Micro-ED enabled a *de-novo* protein structure determination [13]. Despite the high dose efficiency of electrons, damage accumulation throughout the rotation series remains a limiting factor, and acquisition as well as sample screening require careful operation at extremely low dose rates [14]. While some automated procedures are becoming available [15], acquisition of Micro-ED data still involves considerable manual effort for identifying crystals at low-dose conditions, and careful dose fractionation during exposure.

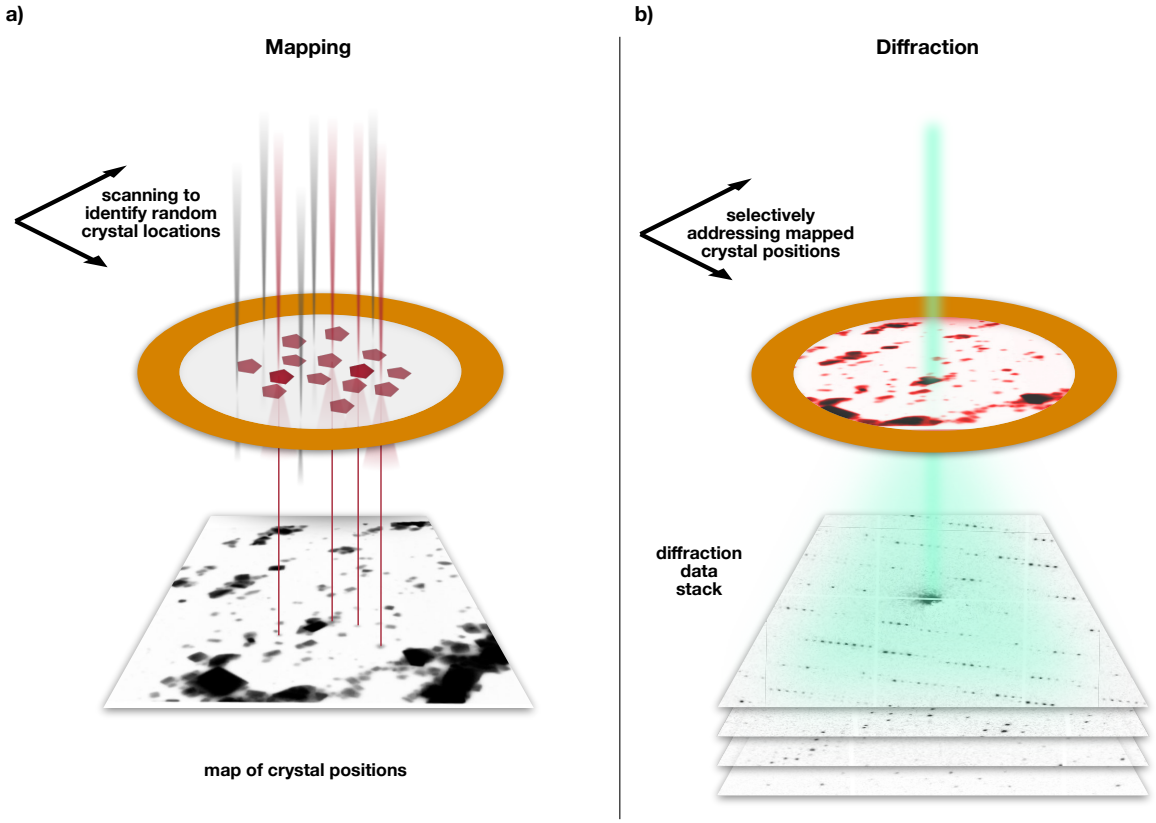


Figure 1: Serial nano-beam electron diffraction scheme. (a) The sample is first mapped in low-dose STEM mode over a large region, yielding a real-space map of a region of approximately 18 µm edge length. Crystals show up as clear features and can be identified automatically. (b) The beam of ≈ 100 nm diameter is sequentially steered to each found crystal position, and diffraction patterns are acquired at a rate of up to 1 kHz.

Results

Here we introduce a new method to perform *serial* electron crystallography using a parallel nano-beam in a scanning transmission electron microscope (S/TEM). Analogous to the approach of serial X-ray crystallography, with a high degree of automation and ease of use, we mitigate the problem of damage accumulation by exposing each crystal only once. A newly developed indexing algorithm (see Supplementary Material) allows to merge the obtained patterns into a full crystallographic data set. Our serial nano-beam electron diffraction (SerialNED) approach consists of two automated steps (see Online Methods for details): First, an overview image is recorded in scanning (STEM) mode at a negligible radiation dose, and the positions of the randomly oriented crystals are mapped (Figure 1a). Second, still electron diffraction patterns are recorded from each crystal at a fixed sample rotation angle, synchronizing the microscope's beam deflectors with a high frame rate camera (Figure 1b). Thereby, a hit fraction approaching 100% with a data collection rate of up to thousands of diffraction patterns per second can be achieved, limited only by source brightness and camera speed. Importantly, since the full workflow is conducted in a conventional S/TEM instrument no special sample delivery devices are

required. In contrast to wide-field TEM-based serial diffraction schemes [16], our nano-beam approach enables a collimated beam geometry and precise dose control, both of which are mandatory for application to protein crystals. The beam diameter of typically 100 nm can be matched to the crystal size, minimizing background scattering and diffraction from multiple lattices. Furthermore, the microscope remains in diffraction mode at all times, and both the position mapping and addressing of the crystals for diffraction are achieved using the same set of deflectors; hence, no mode switching or real-space calibrations are required.

We first determined the structure of the standard reference system hen egg white lysozyme (HEWL). 2 µl of suspension of HEWL crystals of typically 100-500 nm diameter was deposited on standard TEM-grids and vitrified. Two independently prepared samples were measured in separate acquisition runs over a total measurement duration of 5 h and a sample area of 0.015 mm^2. Diffraction patterns from ≈ 1900 nano-crystals were collected, achieving a hit fraction of 57% at an acquisition rate of 20 Hz and 50 Hz for the two samples, respectively. 90% of the obtained patterns could be successfully processed, resulting in a Coulomb-potential map of excellent quality to 1.8 Å resolution (Figure 2; Supplementary Table 1). The determined HEWL structure compares well to previously determined structures by X-ray and micro-electron diffraction techniques (see Supplementary Information). It is noteworthy that in comparison to rotation Micro-ED, the data exhibits a superior signal-to-noise ratio as well as multiplicity and correlation coefficients (see Supplementary Table 1).

As a second test system, we used vitrified native crystalline granulovirus occlusion bodies. These have previously been studied at LCLS [17], and are therefore ideally suited to compare the SerialNED approach to XFEL data. Granulovirus occlusion bodies display a higher morphological homogeneity due to their natively crystalline state compared to HEWL nano-crystals. Using the same methodology as for the HEWL sample, we obtained ≈ 32000 diffraction patterns within a 4 h net measurement duration from a total sample area of 0.036 mm^2, achieving a hit fraction of 69% at an acquisition rate of 50 Hz. The goniometer tilt was occasionally changed (up to 40°) between acquisition runs of different regions, to mitigate effects of preferred sample orientation. Up to 81% of these hits could subsequently be indexed and used for merging. Diffraction data yielded Coulomb potential maps and data statistics of excellent quality at 1.55 Å resolution (Figure 2, Supplementary Table 1). In comparison to published XFEL data, we obtained a 0.45 Å improvement in resolution while displaying higher map-quality.

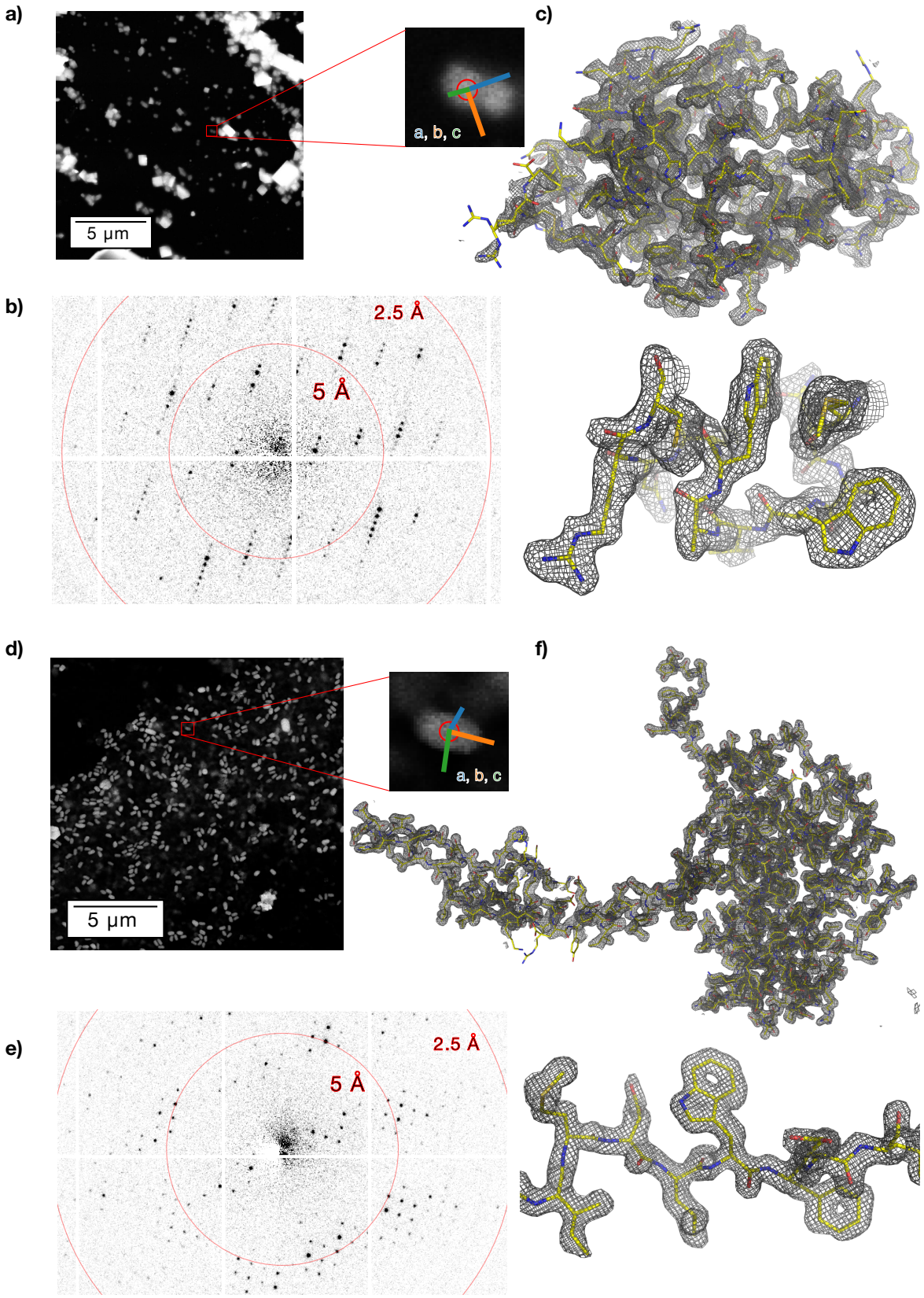


Figure 2: SerialNED results for lysozyme nanocrystals (a-c) and granulovirus occlusion bodies (d-f). (a,d) STEM mapping images. Bright features are visible on the dark background. Zoomed views of a single representative feature are shown, where the red circle corresponds to the diffraction nano-beam diameter of ≈110 nm. Coloured lines indicate the lattice vectors found after indexing of the diffraction pattern. (b,e) Background-subtracted diffraction pattern acquired from the shown feature, respectively. (c,f) Obtained structures of lysozyme (c) and granulin (f), and zoom into a randomly chosen region.

The high frame rate of the detector applied in our experiment allows recording a burst-series comprising several frames instead of a single snapshot for each crystal, yielding a dose-fractionated "*diffraction-during-destruction*" data stack. In the granulovirus data set, the total exposure time of 20 ms was fractionated into 10 shots of 2 ms each, recorded at a rate of 500 Hz. This stack of patterns enables one to find the optimal dose before onset of radiation damage via observing the decrease of Bragg peak intensities between consecutive frames (Supplementary Figure 4). A final set of diffraction images is generated *after* data acquisition has concluded by trading off between optimum resolution (short integration) and signal-to-noise ratio (long integration), typically discarding the first shot of each stack due to residual artefacts (see Supplementary Figure 4). This approach further allows to exploit the improved low-resolution signal-to-noise ratio at long integration for crystal orientation determination (indexing), while deriving reflection intensities for structure determination only from the damage-free portion of frames. For our data sets, we find that exposures exceeding about 1.6 $e^-/Å^2$ result in a detectable loss of the high-resolution shell intensity, in accordance with previous studies [14] (Supplementary Results). Optimal data quality was reached for an exposure time of 4 ms, corresponding to an integrated dose of ≈ 3.2 $e^-/Å^2$.

Discussion

In conclusion, we have demonstrated that serial nano-beam electron diffraction (SerialNED) allows the determination of protein structures at high resolution from extremely small and beam-sensitive crystals in a highly efficient and automated manner. No sample rotation during data acquisition is required, simplifying the measurement and allowing to use higher doses for each diffraction pattern. We have demonstrated a net acquisition rate of 35 Hz when factoring in the hit fraction, comparing favourably with that of ≈ 25 Hz so far reached at the European X-ray free-electron laser [18,19]. Note that a further increase of more than another order of magnitude can be achieved if dose-fractionation is omitted. The SerialNED-approach can potentially also be applied to heterogeneous systems such as cells containing *in-vivo* grown nano-crystals [2], or to map and exploit local lattice structures [20]. It is moreover not only limited to proteins but encompasses all nano-crystalline compounds. Similarly, mixtures of crystals within a single grid or contaminated samples can be studied without significant modifications by assigning each found lattice to one of the contained sample classes using multiple indexing runs [12]. Augmenting parallel-beam crystallography with coherent scanning diffraction techniques such as coherent nano-area diffraction, convergent-beam diffraction or low-dose ptychography might be a viable way to obtain Bragg reflection phase information [21,22]. Finally, integrating the serial acquisition approach with emerging methods of *in-situ* and dynamical electron microscopy [23] may open up avenues for time-resolved structural studies on beam-sensitive systems. All of this makes serial nano-beam electron diffraction a versatile, highly efficient and low-cost alternative to canonical structure determination approaches for proteins and beyond.

Acknowledgments

We thank Djordje Gitaric for mechanical design work, Fabian Westermeier for help with adapting the *Lambda* detector to our microscope, Anton Barty and Henry Chapman for numerous helpful discussions and critical reading of the manuscript, Thomas A. White for help with modifying *CrystFEL* for electron diffraction, and Michiel de Kock for help with image processing. We are indebted to Kay Grünewald and his research group for lending to us their cryo-transfer holder.

Author contributions

RB, GK and RJDM conceived the SerialNED scheme. RB implemented the experimental set-up and conducted the measurements. PM and GK prepared the lysozyme nano-crystals, DO prepared the granulovirus sample, and GK and LB prepared vitrified cryo-EM grids from each sample. RB, PH and PM performed the data processing. YG, OY and WB developed the *PinkIndexer* code. RB, ES and GK wrote the paper, with contributions from all authors.

Competing interests

The authors declare no competing interests.

Methods

Sample preparation

Hen Egg-White Lysozyme (HEWL) was purchased from Sigma-Aldrich as a lyophilized powder. It was dissolved in 20 mM NaAcetate pH 4.7 to a concentration of 80 mg/ml. HEWL crystals were grown via batch crystallization, whereby equal volumes of the protein solution and 80 mg/ml NaCl were added. Crystals ranging from 5-10 µm rapidly formed within 2-3 hours. The resulting crystal mixture was centrifuged down and 75% of the supernatant was removed creating a dense crystal slurry. Subsequent vortexing with steel ball bearings in a microfuge tube for 30 minutes resulted in a concentrated suspension of crystal fragments in the sub-500 nm size range.

The granulovirus sample mixture was prepared identically to the prescription described in [24]. In order to achieve a sufficiently high particle density, the thus prepared mixture was centrifuged down with subsequent removal of 90% of the supernatant.

For each of the above suspensions, 2 µl were applied to 400-mesh carbon grids (type S160-4 purchased from Plano GmbH), whereupon blotting and vitrification using a mixture of liquid ethane/propane was performed in a Vitrobot Mark IV (Thermo Fisher Scientific).

Diffraction data acquisition

All data has been acquired on a Philips Tecnai F20 TWIN S/TEM, equipped with a Gatan 626 cryo-transfer holder, a X-Spectrum Lambda 750k pixel array detector based on a 6×2 Medipix3 array [25], and a custom-built arbitrary pattern generator addressing the deflector coil drivers, based on National Instruments hardware. Initially, the grids were screened in low-magnification STEM mode for regions exhibiting a high crystal density without excessive overlap and aggregation. While software such as SerialEM [26] could be used in a straightforward manner to automate this screening step, this was not required for our test samples, as sufficiently homogeneous regions were readily found.

After screening, the microscope was set to nanoprobe STEM mode at the lowest possible magnification, corresponding to a $(18\mu m)^2$ field of view. To achieve a high-current (≈ 0.2 nA), small (≈ 110 nm), and collimated (≪ 0.5 mrad) nano-beam, the field-emission gun parameters were set to the weakest excitation of both gun lens and C1 condenser lens (Spot size), and a small (5 µm) condenser (C2) aperture was inserted.

At each of the identified sample regions, the two-step acquisition sequence as described in the main text and shown in Figure 1 was performed:

1. The beam was focused on the sample (see Supplementary Figure 1a), and an overview STEM image was taken across the entire field of view (Figure 1a) using the high-angle annular dark field (HAADF) detector. The acquisition parameters were set such that the exposure dose (fluence) remains small, well below 0.1 $e^-/Å^2$. From this image, crystals were automatically identified using standard feature extraction methods, and a list of scan points, corresponding to discrete values of the microscope's scan coil currents, was derived (see Supplementary Information).
2. The beam was defocused into a collimated nano-beam (Köhler illumination) of 110 nm diameter, yielding sharp diffraction patterns on the detector. The actual diffraction data acquisition was then performed by sequentially moving the beam to each of the crystal coordinates using the STEM deflectors and recording diffraction patterns in a synchronized fashion. Image sequences comprising at least ten frames at a single crystal position were recorded at 200 Hz or 500 Hz frame rate, yielding a dose-fractionated "diffraction-during-destruction" data stack at ≈ 1.6 $e^-/Å^2$ per frame.

Once data acquisition from the mapping region was complete, the beam was blanked, and the sample stage moved to the next previously identified sample region. This sequence was repeated until several thousand diffraction patterns had been recorded. All steps were automated and controlled using Jupyter notebooks based on Python 3.6 and a custom-written instrument control library (see Supplementary Information).

Data processing

The recorded diffraction patterns were processed in a pipeline comprising dead-pixel and flat-field detector response correction, partial summing of dose-fractionation stacks, as well as centring of each pattern using the position of the transmitted beam and position-matching of simultaneously excited Friedel-mate reflections. Diffraction spots were identified using the peakfinder8 algorithm contained in the CrystFEL package 27,28. The extracted spot positions were used to determine the orientation of each crystal, and to predict the position of the corresponding Bragg reflections using the indexing and refinement algorithm PinkIndexer (Gevorkov et al., in preparation). Intensities of the Bragg reflections were extracted using two-dimensional peak fitting, and a full reciprocal-space data set was obtained using the partialator merging code 29, without explicit modelling of reflection partiality. Refer to the Supplementary Information for additional details on data pre-processing, indexing, and merging. Phasing of the models was achieved by molecular replacement using Phaser 30 from the PHENIX software suite 31 using PDB-ID: 4ET8 and PDB-ID: 5G3X as template models, respectively. Upon obtaining phases, iterative cycles of model building were made using Coot 32. For correct refinement of the Coulomb potential maps, subsequent rounds of refinement were performed using phenix.refine, taking electron scattering factors into account 33. Illustrations of the electron density map and model were generated using PyMOL by Schrödinger. Crystallographic statistics are reported in Supplementary Table 1.

Supplementary Methods

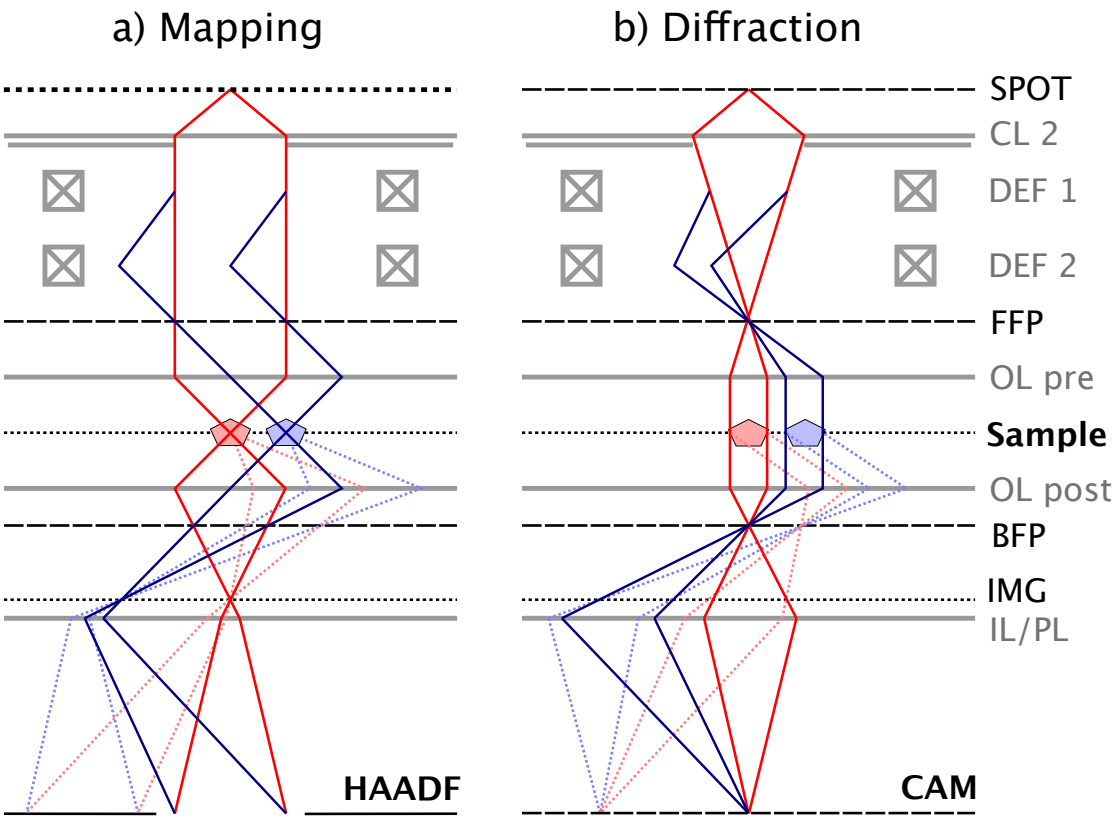
Instrumentation

The serial electron nano-beam diffraction scheme can in principle be performed in any S/TEM or dedicated STEM with pixelated detector and software scripting interface. We used a Philips Tecnai F20 S/TEM with a TWIN pole piece, a Schottky field-emission gun, and a Fischione Model 3000 HAADF-STEM detector. Two non-standard auxiliary hardware devices are required: a diffraction-capable, hardware-triggerable camera, and an arbitrary signal generator controlling the scanning deflectors, which can accept a list of discrete points to be addressed serially.

So that camera speed does not limit the throughput of the acquisition, the frame rate should be at least tens of Hertz if beam currents of tens of picoamperes are used. Implementation of the dose-fractionated movie mode requires a significantly higher frame rate, ideally hundreds of Hertz or greater. A hybrid pixel detector [25] meets these requirements optimally, as long as the count rates do not exceed the saturation threshold. We used a 6×2-panel Medipix3-based detector (X-Spectrum Lambda 750k) operating in 12-bit continuous-readout mode at up to 2 kHz and a resolution of 1536×512 pixels. Scintillator-coupled detectors based on latest-generation fast CMOS sensors are a viable alternative. Back-thinned monolithic pixel detectors, while offering ultimate resolution and high speed, may not provide sufficient dynamic range and radiation hardness to record the central region of diffraction patterns, such that a beam-block is required, and low-resolution peaks may become saturated due to the strong inelastic background signal.

To direct the sequential beam motion, the X/Y (line/frame) control voltages of the STEM deflector drivers were addressed from an off-the-shelf PC-based data acquisition board (National Instruments PCI-6251). The list of scan points derived as described below is directly written into the output buffer of the board's digital-analogue converters. While the data acquisition is running, synchronized trigger signals for the camera are provided. The same hardware is also used for acquiring data from the HAADF-STEM detector during the mapping step.

To control the microscope, detector, and scan generator, we use custom software based on Python 3.6 and National Instruments LabVIEW, implementing high-level functions for serial crystallography workflow automation. Instead of a dedicated graphical user interface, we use *Jupyter* notebooks to control an acquisition run and visualize its progress, which can be adjusted to the sample under study and annotated, providing a self-documenting protocol for each data acquisition.



Supplementary Figure 1: Ray path diagrams for a) mapping (focused) and b) diffraction (collimated) condenser configuration. Red and blue lines correspond to on-axis and one exemplary off-axis positions of the beam. Dotted lines correspond to a Bragg reflection. Optical planes and electron-optical elements are shown in black and grey, respectively. In the mapping configuration, the beam is collimated by the lower condenser lens (CL 2) and focused on the sample using the objective lens pre-field (OL pre). Scattered beams from the illuminated sample position are imaged on the high-angle annular dark field (HAADF) detector using the objective lens post-field (OL post) and the intermediate and projection lenses (IL/PL). A STEM mapping image is formed by scanning the beam over the sample and registering the integrated current on the detector. In the diffraction configuration, on the other hand, the condenser focuses the beam on the front-focal plane (FFP) of the objective; hence, the sample is exposed to a collimated pencil beam. Diffraction orders now appear as discrete spots on the diffraction detector (CAM), the positions of which are independent of the beam position on the sample. Note that switching between these configurations involves changing of the CL 2 excitation only, as the detectors always remain in a plane conjugate with the back and front focal planes of the objective lens (diffraction mode). (SPOT – first condenser lens (spot) crossover; DEF1/2 – upper and lower beam deflector pair; IMG – intermediate image plane).

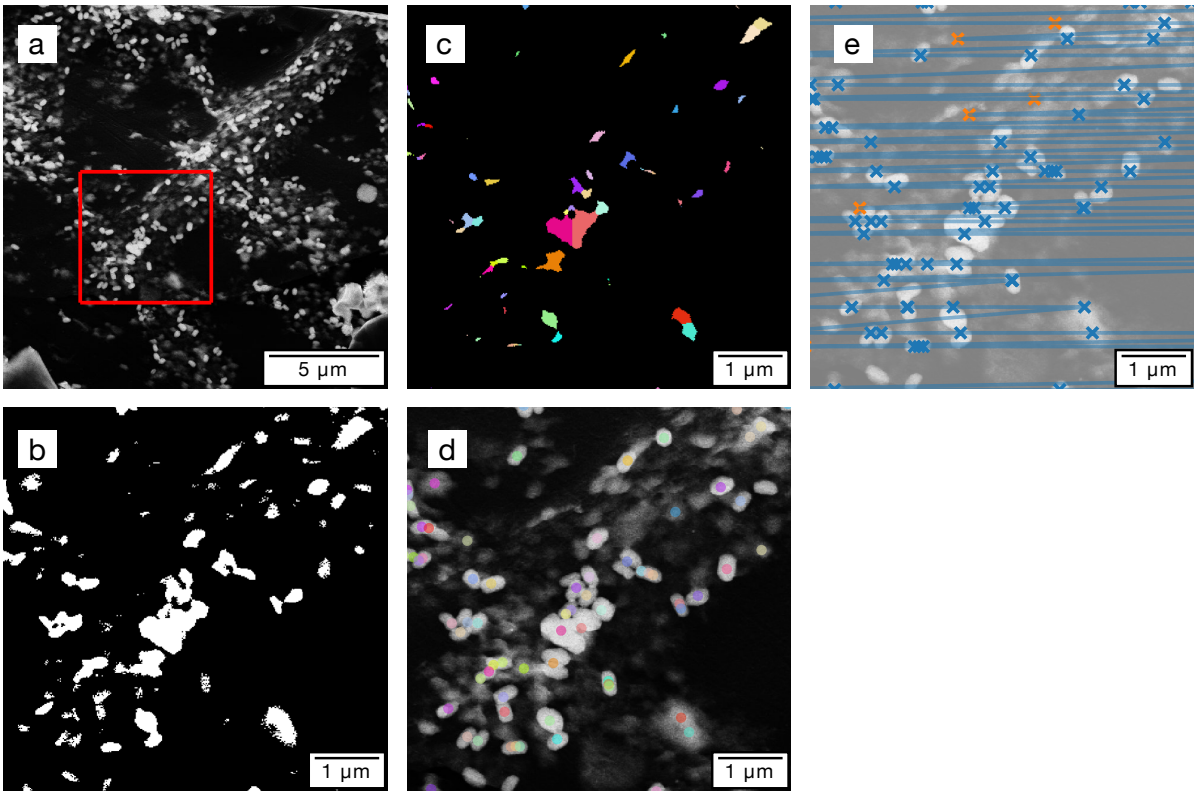
Instrument preparation

In the following, we lay out in more detail the steps required to acquire a serial crystallography data set as performed in our work. Before executing the procedure outlined in the online methods section, the following preparation steps have to be taken:

- Common parameters of the microscope should be properly aligned and characterized. Of specific importance are the STEM pivot points and distortion, condenser and diffraction stigmators, rotation center (beam tilt) and centering of the condenser (C2) aperture, which

defines the final nano-beam size when operating in collimated mode, and the convergence angle in focused mode. The camera length and any diffraction distortion should be carefully calibrated using a standard polycrystalline target such as thallous chloride (TlCl). While helpful for interpreting the results, a precise real-space calibration of the deflectors (STEM magnification, static beam shift) is not required.

- By changing the settings of the electron gun (spot size, gun lens, and extraction voltage on a FEG instrument), the beam current is optimized to match the beam diameter d, camera frame rate f, desired total dose (fluence) D, and number of dose-fractionation movie frames K as $I = eDf\pi(d/2)^2/K$, where e denotes the elementary charge.
- The setting of the condenser lens corresponding to a collimated beam is determined by focusing diffraction spots with the projection system in diffraction mode and the diffraction lens focused on the back focal plane of the objective lens (Diffraction setting as shown in Supplementary Figure 1a). Diffraction data is taken using this setting, as described in the main text.
- The sample is brought to the eucentric height by minimizing image motion when wobbling the stage tilt. The condenser setting (defocus in STEM mode) required to achieve a focused STEM image is stored in the automation software (Mapping setting in Supplementary Figure 1b). Now, the position of the condenser aperture (C2) can be precisely aligned by switching the microscope repeatedly between focused (mapping) and collimated (diffraction) condenser lens settings and observing the position of the beam in the sample plane. In our microscope we notice that the residual beam shift between both settings can be minimized by renormalizing the illumination system after each change between settings. Even without renormalization, a satisfactory repeatability can however be reached, as long as the condenser lens is not intermittently set to other values.
- Finally, the offset of the STEM mapping image along the fast-scanning axis x due to finite scan speed Δx_{scan} needs to be determined. This can be achieved by recording STEM data along a few y-lines only, but spanning the full x-range. This is repeated twice, once for the scan parameters as used for the mapping image (dwell time, pixel size, resolution, magnification, fly-back time, detector time constant, etc.), and once at a low scan speed (dwell time of typically 1 ms, no fly-back time), with the beam assuming a quasi-static position within each pixel dwell time. Δx_{scan} can be determined via a straightforward cross-correlation registration between the obtained intensity data. The offset calibration is fully automated and requires less than one minute. We find that the obtained value remains stable over measurement sessions spanning several days.



Supplementary Figure 2: Particle picking and scan path generation. (a) Typical mapping image, with granuloviruses and residual solvent contamination. The red square indicates the zoomed region of panels (b-e). (b) Zoomed view after binarization using automatically determined threshold. (c) Labelled regions after segmentation. (d) Mapping image with final beam coordinates for diffraction. (e) Scan path with acquisition points (blue) and auxiliary points (orange).

Crystal finding and acquisition programming

Once the mapping image has been acquired using STEM, the coordinates of the nano-beam for diffraction recording, corresponding to crystal features, are determined using the following automated image analysis procedure, which can be adapted in its parameters to the sample of interest:

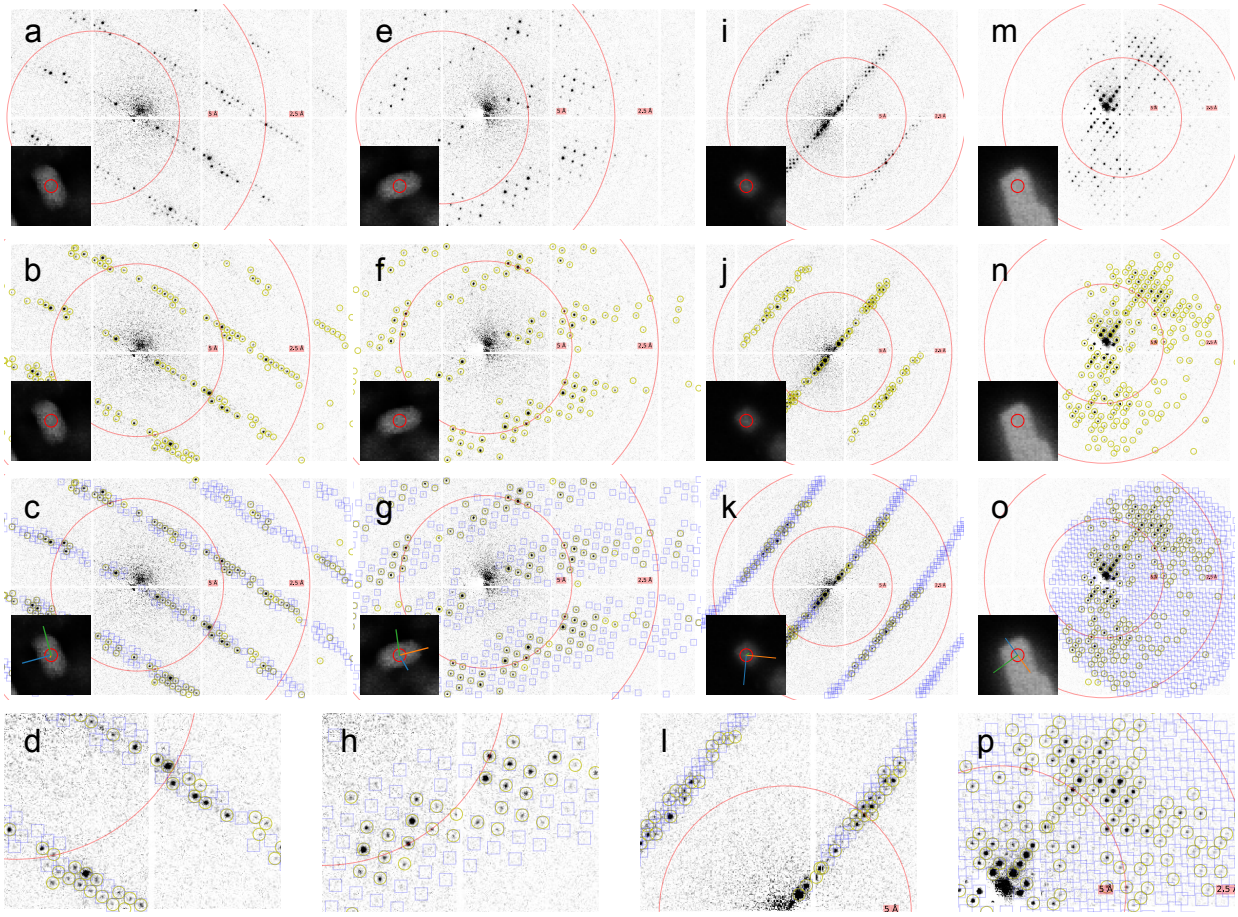
- The image is binarized using a fixed or automatically determined grey-value threshold, the latter derived using Otsu's or Li's method [34,35].
- A morphological closing operation with a structuring element of a typical minimum size of a single crystal is performed to exclude noise and features that are too small.
- Crystals are often found within aggregates or thicker sample regions, which are registered as a single bright segment. A second round of thresholding and binarization can now be performed on each individual bright segment to locate individual crystals.
- In order to further discern crystals in connected regions, a watershed segmentation starting from either local intensity or distance-transform maxima is performed. The resulting segments are assumed to belong to a single crystal and assigned a unique ID.
- For each segment, a diffraction beam position is selected by determining its center of mass. Alternatively, multiple beam positions spaced by a distance of approximately the nano-beam

diameter can be distributed over the segment, using a k-means clustering approach similar to [16]. While this has not been performed in the present work, it may prove beneficial for cases where clear boundaries of adjacent crystals are not readily discernible, or to study and exploit local lattice structures [20,36].

At this point, the coordinates of the desired beam positions for recording diffraction patterns are known. Note that these coordinates do not have to be calibrated to real space, but merely represent control voltages of the deflection coil drivers applied during the mapping image acquisition. Next, a list of scan points, by which we mean the nominal values of the scan generator outputs for the diffraction acquisition step, are derived. Due to effects such as beam hysteresis, these have to be corrected and modified with respect to the previously determined crystal coordinates. The derivation is conducted as follows:

- The crystal coordinates of all crystals along the y axis (vertical, slow-scanning) are clustered into a discrete set (scan rows at coordinates y') using a one-dimensional k-means algorithm. The number of scan rows is lower than the total number of crystals, but chosen such that the maximum deviation between desired (y) and discretized (y') coordinates remains below a given threshold, typically chosen as half the scanning beam radius. Coordinates along the x axis (horizontal, fast-scanning) remain unaffected. At this point, the list of scan points is initialized from the (x, y') coordinate tuples.
- Scan points are identified, where the distance to the previous one along any axis is either negative or exceeding an empirically determined threshold. Before each such point, an auxiliary scan point is inserted, at the same y'-position as the actual point, and an x-position reduced by a certain amount. The dwell time at the auxiliary points is typically shorter than on the actual recording points, and either no diffraction data is recorded for them, or it is discarded in later processing steps. This step ensures that each scan point is addressed from the same direction (from the top and/or from the left) from a distance that is not exceedingly large. While the former ensures position reproducibility despite lens hysteresis, the latter helps to avoid artefacts in the diffraction patterns arising from the finite beam scan speed.
- Finally, the offset Δx_{scan} obtained from calibration procedures as described above are applied.

The full algorithm to derive a list of scan points from a STEM mapping image is illustrated in Supplementary Figure 2. The obtained list of scan points is then written into the memory of the scan generator. Dose-fractionation movies are implemented as repeated points with identical coordinates, each one triggering a new camera acquisition.



Supplementary Figure 3: Data processing pipeline. (a) Diffraction pattern from a single granulovirus after dead-pixel and flatfield correction. Note that, for better visibility, we have subtracted the radially symmetric background signal in this figure, which is however not done in the actual data analysis. As the diffraction pattern is not centred of the detector, the resolution rings shown at 2.5 Å and 5 Å are not aligned with the pattern. The inset shows a close-up of the STEM mapping image on the corresponding granulovirus; the red circle corresponds to the beam diameter of 100 nm. (b) Same pattern after determination of beam centre and diffraction peaks, marked in yellow. (c) Same pattern including predictions of Bragg reflection positions after running indexing and refinement in *PinkIndexer*, shown as blue squares. The lines in the mapping image inset correspond to the derived real-space lattice vectors. For single granuloviruses it is typically found, that one of the lattice vectors is aligned with the long axis of the virus shell. (d) Zoom into a region of (c), highlighting the matching between predicted peaks (blue squares) and pixel intensity data. (e-h) As (a-d), for another virus. (i-p) As (a-h), for two lysozyme nano-crystals.

Data pre-processing

Detector data for each single acquisition run is streamed into a single HDF5 file according to the *NeXus* specification [37], which is commonly used in X-ray diffraction. The diffraction data is arranged in a three-dimensional image stack, with a height $Kn_{\mathrm{cryst}} + n_{\mathrm{aux}}$, with n_{cryst} the number of crystals in the sample region, K the number of dose-fractionation movie frames, and n_{aux} the number of auxiliary points inserted as described above. Furthermore, the scan position list generated as outlined before, the mapping STEM image with metadata for each found feature, and all accessible settings of the microscope, detector, and scanning unit, are stored within the *NeXus* file.

Starting with these raw input files, the steps outlined in the following are performed to pre-process the data set for use in standard diffraction data reduction software. Using the Python package *dask*, all operations are performed using chunked lazy evaluation in a single calculation step, and efficiently scale on multi-processor systems, with only modest memory requirements; metadata are handled using the *pandas* package. Custom accessor methods contained in our in-house *diffractem* software provide high-level functions for typical operations, as well as ensuring consistency between metadata and diffraction images.

- The recorded diffraction data are filtered such that all images corresponding to auxiliary scan points are removed. If dose-fractionation movies have been recorded, an effective integration time can now be set by summing a correspondingly large slice of the movie stack for each crystal.
- Dead-pixel correction is applied by either replacing all dead pixels with a given integer number (typically -1) or interpolating from adjacent pixels. Optionally, a flat-field correction can be applied by multiplying each pixel value with a previously determined normalization value. The pixels near the gaps of the 12 detector panels, which are three times more elongated in the direction facing the gap and hence have a different effective gain and saturation behaviour, can either be omitted from the analysis, or scaled to have their intensity matched with the other pixels.
- The centre of each diffraction pattern is determined in a multi-step process, and the images are correspondingly shifted (Supplementary Figure 1, second row). This is mandatory, as even for a good alignment of the STEM pivot point before data acquisition, a slight position-dependent beam tilt will remain. This manifests as displacement of the diffraction pattern, hampering the accuracy of the subsequent indexing step. First, the centre-of-mass of pixel intensities within the inner region of the image is found for each shot. Next, the obtained position is used as a starting value for least-squares fitting of a rotationally symmetric Lorentzian function over a small domain (30×30 pixels) around the centre-of-mass position. Finally, if peaks are found in the diffraction pattern, a refinement of the centre position can be performed by matching the position of Friedel-mate reflections, which are generally found at low resolutions. Further refinement of the centre of each diffraction pattern is done at the indexing-prediction step using *CrystFEL* (see section on indexing below).

The final result of this pipeline is a data stack containing the corrected, dose-integrated and centred diffraction data and corresponding metadata of all diffraction shots, contained in *NeXus*-compatible HDF5 files. We could successfully export the data to the *CrystFEL* [27,29], *DIALS* [38], and *nXDS* [39] packages.

Data reduction

To obtain a fully merged crystallographic dataset from the single-crystal snapshots, we use the tools provided in *CrystFEL* 0.8.0. Bragg reflections in the diffraction patterns are registered using the *peakfinder8* algorithm [28] (Supplementary Figure 3, second row). Because this algorithm internally estimates the radially symmetric background for each resolution shell, we have found no increase of accuracy when background subtraction is applied to the diffraction patterns before peak finding.

Indexing

One of the most difficult tasks when processing a single electron diffraction pattern is to find the orientation of the crystal that generated this pattern. Due to the very short de Broglie wavelength of electrons (0.025 Å at 200 kV, as compared to several Å in the case of X-rays), the measured part of the Ewald sphere is almost flat in the resolution range used for the measurements. Therefore, hardly any three-dimensional information can be extracted from a single pattern. To overcome this limitation, the *PinkIndexer* algorithm used in this paper employs prior unit-cell information as a constraint to compensate the lack of three-dimensional information (Gevorkov *et al.*, in preparation).

Given knowledge of the unit cell, the straight-forward approach of checking every possible orientation of the crystal for conformity with the diffraction pattern has been proven feasible [16]. However, this approach is limited by its computational complexity of $O(N^3)$, with N being the number of sampled angles per full rotation, and hence impractical for systems with large unit cells and a correspondingly densely structured reciprocal space. On the other hand, the *number of orientations predicting one given peak* is only proportional to N. *PinkIndexer* computes this manifold of consistent orientations for each of the peaks, and determines the most likely crystal orientation from the overlap of all manifolds. Hence reducing the complexity from $O(N^3)$ to $O(N)$ concomitantly allows for finer sampling of the angles. After finding the best match among the sampled orientations, a refinement step is performed to interpolate between the sampled angles, as well as to account for uncertainties in the unit cell size and the beam centre location. Varying these parameters, the average distance between observed and predicted peaks is minimized using a simple gradient descent approach with randomly selected starting points in proximity to the initial indexing solution (Supplementary Figure 3, third and fourth row). With this method, we could reach an indexing rate of better than 80% for all data.

PinkIndexer can be used as part of the *CrystFEL* package [27,29,40].

Merging

The dataset is then merged using *partialator* [29] without partiality modeling (assigning unit partiality to all recorded reflections), yielding a plain-text *hkl*-File containing the full reduced dataset, which can readily be converted to the *mtz* format and imported into phasing software.

Supplementary Results

Crystallographic data

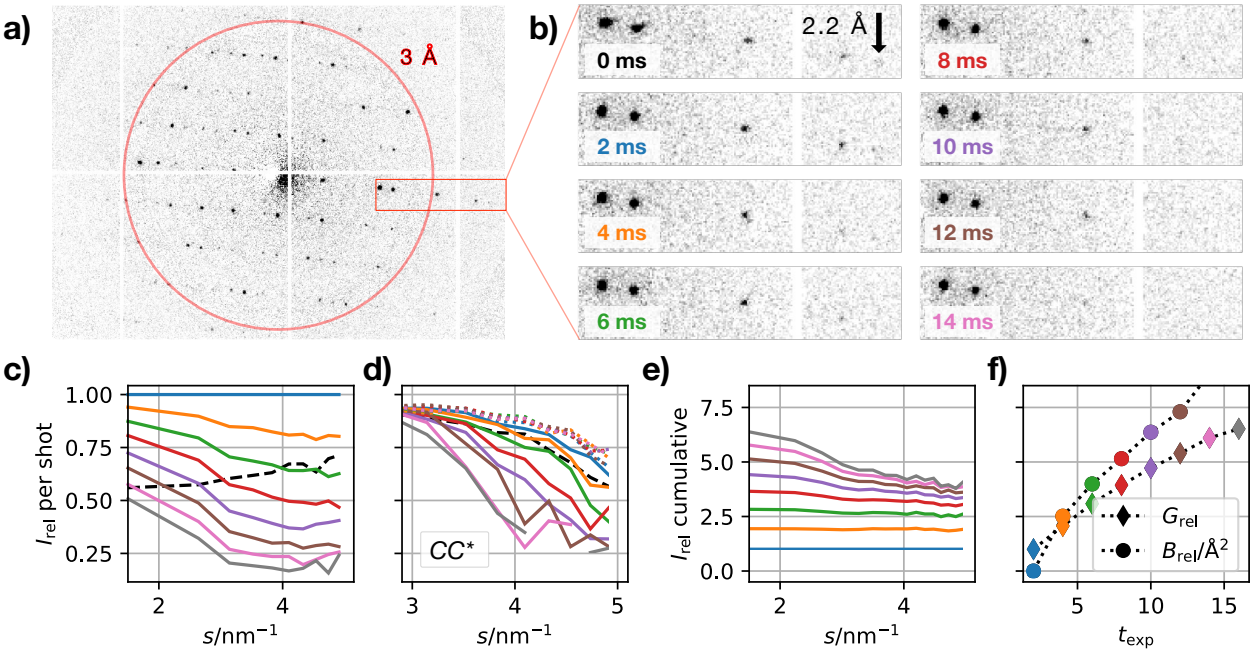
Supplementary Table 1: Data collection and refinement statistics for Lysozyme (Lys) and granulovirus occlusion bodies (GV). Statistics for the highest-resolution shell are shown in parentheses. Data were truncated at CC* ≈ 0.5 [41].

	Lys (PDB-ID: 6S2N)	**GV (PDB-ID: 6S2O)**
Data collection		
Wavelength (Å)	0.025	0.025
Dose (fluence) per crystal	3.2 $e^-/Å^2$	3.2 $e^-/Å^2$
Resolution range	55.93-1.80 (1.87-1.80)	72.12-1.55 (1.61-1.55)
Space group	P 43 21 2	I 2 3
Unit cell	79.1 79.1 38 90 90 90	103.2 103.2 103.2 90 90 90
Total reflections	16136 (762)	9700219 (90049)
Unique reflections	9444 (489)	26569 (2858)
Multiplicity	32 (4)	367 (32)
Completeness (%)	80.8 (43.3)	100 (99)
Mean I/sigma(I)	3.95 (2.59)	9.17 (1.90)
Wilson B-factor	9.98	8.98
R-split (%)	23.2 (47.0)	9.48 (50.9)
Refinement		
CC1/2	0.94 (0.20)	1.00 (0.18)
CC*	0.98 (0.57)	1.00 (0.56)
Reflections used in refinement	9418 (489)	26569 (2730)
Reflections used for R-free	458 (23)	1310 (134)
R-work	0.293 (0.394)	0.211 (0.370)
R-free	0.320 (0.397)	0.259 (0.414)
Number of non-hydrogen atoms	1049	2150
macromolecules	1001	2030
solvent	48	120
Protein residues	129	243
RMS (bonds)	0.003	0.004
RMS (angles)	0.52	0.66
Ramachandran favored (%)	93.7	97.1
Ramachandran allowed (%)	6.30	2.49
Ramachandran outliers (%)	0.00	0.41
Rotamer outliers (%)	0.00	0.00
Clashscore	5.61	2.75
Average B-factor	15.12	12.5
macromolecules	15.27	12.2
solvent	12.01	16.3

To validate the consistency of our structures with known data, we calculate the rms deviations (RMSD) of atom positions with respect to previously published structures. For Lys with respect to PDB-ID 5K7O (measured by MicroED) we find a value of RMSD = 0.487 Å; for Lys w.r.t. 5WR9 (measured by XFEL serial crystallography) RMSD = 0.353 Å. For GV with respect to PDB-ID 5G3X (synchrotron crystallography) we find RMSD = 0.206 Å; w.r.t. 5G0Z (XFEL) RMSD = 0.353 Å.

Radiation damage and dose fractionation

The dose-fractionated "diffraction-during-destruction" movie frame stacks acquired for each crystal as explained in the main text can be used to optimize the exposure times of the diffraction images *after* the actual measurement. This is achieved by summing movie frames up to a given frame number k, corresponding to a specific total integration time $k \cdot f^{-1}$ and dose $k \cdot I(Af)^{-1}$, where f, I, and A denote the camera frame rate, beam current, and beam area, respectively. Comparing figures of merit of



Supplementary Figure 4: Radiation damage during dose-fractionated acquisition. (a) Typical diffraction pattern from a lysozyme nano-crystal. The red box indicates the enlarged region in b). For improved clarity, the radially symmetric background has been subtracted in this figure. (b) Enlarged diffraction pattern section for single frames from the dose-fractionated movie stack. Note the fading of the diffraction spots especially at high resolutions. (c) Diffracted intensity as a function of the diffraction vector length, averaged over ≈ 1000 patterns for each time point after starting the diffraction acquisition, and normalized to data taken from second time point (2 ms). Line colors correspond to those of the labels in b), except for the grey line which corresponds to 16 ms. The dashed line indicates the data for the first time point, where the beam only partially hits the sample and artefacts are present. (d) Resolution-dependent correlation coefficients CC*, y axis as in c). (e) Same as in c), but for data corresponding to different integration times, derived from progressive partial summation of diffraction patterns from each movie stacks, starting from the second (2 ms) frame and ending with the time points corresponding to the line colors. The corresponding correlation coefficients are shown in d) as dotted lines. (f) Relative scaling (G_{rel}) and Debye-Waller (B_{rel}) factors of merged data sets derived from such partially summed movie stacks.

the resulting data sets, an optimal value of k can be determined. In our current implementation, the first frame of each stack is unfortunately often corrupted due to artefacts from residual beam motion, visible as peak broadening in the "0 ms" image in Supplementary Figure 4b. To obtain the final data set, we thus start summing the movie frames from the *second* frame, and label the frames correspondingly by the acquisition time passed since the end of the (discarded) first frame. In Supplementary Figure 4, results from dose fractionation are collected for the second sub-set of lysozyme nano-crystals as described in the main text. Note the fading of diffraction peaks in the single movie frames displayed in panel (b) during exposure. In panels (c) and (d), aggregated data for such *single* movie frames are shown. Both the diffracted intensity and the consistency of the resultant merged dataset as witnessed by the correlation coefficient CC^* [41] rapidly decline, especially at high resolution. Results from progressive partial summation of movie frames as discussed above are shown in panels (e) and (f). In panel e), for example, the 6 ms plot (green) corresponds to the sum of 2 ms (blue), 4ms (orange), and 6 ms (green) plots in panel c). Consistently with the single frames, the diffracted signal rises in a sub-linear fashion with exposure time, along with a decrease in apparent resolution. In panel (f) this is quantified by relative scaling and Debye-Waller factors G_{rel} and B_{rel}, derived from a least-squares fit of the relation $I_k(s) = G_{\mathrm{rel}} \cdot e^{-B_{\mathrm{rel}} s^2} I_1(s)$ to the merged reflection intensities $I_k(s)$ in a given resolution shell s, derived from summing k movie frames. In panel (d), the correlation coefficients CC^* are shown for the partially summed data as dotted lines. It becomes clear that the data quality for this data set is no longer significantly improved after summing two movie frames only, and any further increase of exposure time will only decrease the apparent resolution, as shown in panel (f). We hence infer that setting a total integration time of 4 ms in this data set yields an optimal overall result. Applying the same analysis to the granulovirus data also favours an integration time of 4 ms. According to the screen current readout of the TEM, this corresponds to a dose (fluence) of $\approx 3.2\ e^-/Å^2$. More detailed measurements of site-specific and global radiation-damage effects, as well as optimization of data acquisition and analysis strategies to further improve dose efficiency and resolution will be the subject of future work.

References

1. 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).

2. Schönherr, R., Rudolph, J. M. & Redecke, L. Protein crystallization in living cells. *Biol. Chem.* **399**, 751–772 (2018).

3. Owen, R. L., Rudino-Pinera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl. Acad. Sci.* **103**, 4912–4917 (2006).

4. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–7 (2011).

5. Stellato, F. *et al.* Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ* **1**, 204–212 (2014).

6. Henderson, R. & Unwin, P. N. T. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32 (1975).

7. Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *Elife* **2**, e01345 (2013).

8. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. Electron crystallography of ultrathin 3D protein crystals: Atomic model with charges. *Proc. Natl. Acad. Sci.* **112**, 3368–3373 (2015).

9. Clabbers, M. T. B. *et al.* Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 738–748 (2017).

10. Mugnaioli, E., Gorelik, T. & Kolb, U. 'Ab initio' structure solution from electron diffraction data obtained by a combination of automated diffraction tomography and precession technique. *Ultramicroscopy* **109**, 758–765 (2009).

11. Gruene, T. *et al.* Rapid Structure Determination of Microcrystalline Molecular Compounds Using Electron Diffraction. *Angew. Chemie Int. Ed.* **57**, 16313–16317 (2018).

12. Jones, C. G. *et al.* The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* **4**, 1587–1592 (2018).

13. Xu, H. *et al.* Solving the first novel protein structure by 3D micro-crystal electron diffraction. *bioRxiv* 600387 (2019). doi:10.1101/600387

14. Hattne, J. *et al.* Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* **26**, 759-766.e4 (2018).

15. Cichocka, M. O., Ångström, J., Wang, B., Zou, X. & Smeets, S. High-throughput continuous rotation electron diffraction data acquisition via software automation. *J. Appl. Crystallogr.* **51**, 1652–1661 (2018).

16. Smeets, S., Zou, X. & Wan, W. Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* **51**, 1–12 (2018).

17. Gati, C. *et al.* Atomic structure of granulin determined from native nanocrystalline granulovirus using an X-ray free-electron laser. *Acta Crystallogr. Sect. A Found. Adv.* **73**, a292–a293 (2017).

18. Wiedorn, M. *et al.* Megahertz serial crystallography. *Nat. Commun.* 1–11 (2018). doi:10.1038/s41467-018-06156-7

19. Grünbein, M. L. *et al.* Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9**, 1–9 (2018).

20. Gallagher-Jones, M. *et al.* Lattice nano-ripples revealed in peptide microcrystals by scanning electron nanodiffraction. (2018).

21. Pelz, P. M., Qiu, W. X., Bücker, R., Kassier, G. & Miller, R. J. D. Low-dose cryo electron ptychography via non-convex Bayesian optimization. *Sci. Rep.* **7**, 9883 (2017).

22. Ophus, C. Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond. *Microsc. Microanal.* **25**, 563–582 (2019).

23. Karakulina, O. M., Demortière, A., Dachraoui, W., Abakumov, A. M. & Hadermann, J. In Situ Electron Diffraction Tomography Using a Liquid-Electrochemical Transmission Electron Microscopy Cell for Crystal Structure Determination of Cathode Materials for Li-Ion batteries. *Nano Lett.* **18**, 6286–6291 (2018).

24. Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep.* **7**, 44628 (2017).

25. Nederlof, I., van Genderen, E., Li, Y. W. & Abrahams, J. P. A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1223–30 (2013).

26. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

27. White, T. A. *et al.* CrystFEL : a software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).

28. 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).

29. White, T. A. *et al.* Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–689 (2016).

30. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

31. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

32. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

33. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

34. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.* **9**, 62–66 (1979).

35. Li, C. H. & Lee, C. K. Minimum cross entropy thresholding. *Pattern Recognit.* **26**, 617–625 (1993).

36. Brázda, P., Palatinus, L. & Babor, M. Electron diffraction determines molecular absolute configuration in a pharmaceutical nanocrystal. *Science (80-.).* **364**, 667–669 (2019).

37. Könnecke, M. *et al.* The NeXus data format. *J. Appl. Crystallogr.* **48**, 301–305 (2015).

38. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 85–97 (2018).

39. Kabsch, W. Processing of X-ray snapshots from crystals in random orientations. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2204–2216 (2014).

40. White, T. A. Processing serial crystallography data with CrystFEL : a step-by-step guide. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 219–233 (2019).

41. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science (80-.).* **336**, 1030–1033 (2012).