Serial protein crystallography *in cellulo*

Dissertation
Submitted for the Fulfillment
of the Degree

Doctor rerum naturalium
(Dr. rer. nat.)

by

Janine Mia Lahey-Rudolph

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Faculty of Mathematics, Informatics and Natural Sciences
Department of Chemistry

Hamburg
February 2024

Reviewers of Dissertation:	**Prof. Dr. Lars Redecke**
	Dr. Henning Tidow
Date of print:	26.02.2024
Date of Disputation:	05.07.2024
Oral defense committee:	**Prof. Dr. Markus Fischer**
	Prof. Dr. Henry Chapman
	Prof. Dr. Lars Redecke
	Dr. Henning Tidow
	Prof. Dr. Robin Santra
Dissertation publication approval:	15.07.2024

The practical part of this work was conducted
under the supervision of
Prof. Dr. Lars Redecke, Prof. Dr. Dr. Henry N. Chapman and Prof. Dr. Markus Fischer.
at the Center for Free-Electron Laser Science, DESY Campus in Hamburg,
and at the Institute of Biochemistry at the University of Lübeck
from October 2015 to October 2019.

List of publications

<u>First authored publications</u>

Lahey-Rudolph, J.M.; Schönherr, R.; Jeffries, C.M.; Blanchet, C.E.; Boger, J.; Ferreira Ramos, A.S.; Riekehr, W.M.; Triandafillidis, D.-P.; Valmas, A.; Margiolaki, I., Svergun, D. and Redecke, L.: Rapid screening of *in cellulo* grown protein crystals via a small-angle X-ray scattering/ X-ray powder diffraction synergistic approach. *J. Appl. Cryst*. **53**, 1169-1180 (2020).
DOI: 10.1107/S1600576720010687
Journal of Applied Crystallography impact factor (2020): 3.304
Title cover JAC issue 53
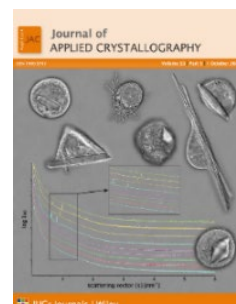
Lahey-Rudolph, J. M.; Schönherr, R.; Barthelmess, M.; Fischer, P.; Seuring, C.; Wagner, A.; Meents, A. and Redecke, L.:
Fixed-target serial femtosecond crystallography using *in cellulo* grown microcrystals *IUCrJ* **8**, 665-677 (2021).
DOI: 10.1107/S2052252521005297
IUCrJ impact factor (2021): 5.588

Schönherr, R.[#], Boger, J.[#], **Lahey-Rudolph, J.M.**[#], Harms, M., Kaiser, J., Nachtschatt, S., Wobbe, M., Duden, R., Koenig, P., Bourenkov, G.P., Schneider, T.R., Redecke, L.:
A streamlined approach to structure elucidation using in cellulo crystallized recombinant proteins, InCellCryst. *Nat Commun* **15**:1709 (2024)
DOI: 10.1038/s41467-024-45985-7
Nature Communication impact factor (2024): 14.919

Co-authored publications associated to PhD

1.   Schönherr, R. [#]; Klinge, M[#]; **Rudolph, J.M.**; Fita,K.; Rehders, D.; Lübber, F.;
     Schneegans, S.; Majoul, I. V.; Duszenko, M.; Betzel, C.; Brandariz-Nuñez, A.;
     Martinez-Costas, J.; Duden, R. and Redecke, L.:
     Real-time investigation of dynamic protein crystallization in living cells.
     Struct. Dyn. **2**, 041712 (2015).
     DOI: 10.1063/1.4921591
     Structural Dynamics impact factor (2015): 2.67

2.   Schönherr, R.; **Rudolph, J.M**., Redecke, L.:
     Protein crystallization in living cells. *Biol Chem*. **399**:751-772
     (2018).
     DOI: 10.1515/hsz-2018-0158.   – *review* -
     Biological Chemistry impact factor (2018): 3.014

Title cover Biol Chem *vol 399 issue 7*

3.   Wiedorn, M.O.; Oberthür, D.; Bean, R.; Schubert, R.; Werner, N.; Abbey, B.;
     Aepfelbacher, M.; Adriano, L.; Allahgholi, A.; Al-Qudami, N.; Andreasson, J.; Aplin,
     S.; Awel, S.; Ayyer, K.; Bajt, S.; Barák, I.; Bari, S.; Bielecki, J.; Botha, S.; Boukhelef, D.;
     Brehm, W.; Brockhauser, S.; Cheviakov, I.; Coleman, M.A.; Cruz-Mazo. F.;
     Danilevski, C.; Darmanin, C.; Doak, R.B.; Domaracky, M; Dörner, K., Du, Y.; Fangohr,
     H.; Fleckenstein, H.; Frank, M.; Fromme, P.; Gañán-Calvo, A.M.; Gevorkov, Y.;
     Giewekemeyer, K.; Ginn, H.M.; Graafsma, H.; Graceffa, R.; Greiffenberg, D.;
     Gumprecht, L.; Göttlicher, P.M; Hajdu, J.; Hauf, S.; Heymann, M.; Holmes, S.; Horke,
     D.A.; Hunter, M.S.; Imlau, S.; Kaukher, A.; Kim, Y.; Klyuev, A.; Knoška, J; Kobe, B.;
     Kuhn, M.; Kupitz, C.; Küpper, J.; **Lahey-Rudolph, J.M.**, et al.:
     Megahertz serial crystallography. *Nat Commun,* 9:4025 (2018).
     DOI: 10.1038/s41467-018-06156-7
     Nature Communication impact factor (2018): 11.88

4.   Nass, K.[#]; Redecke, L.[#]; Perbandt, M.; Yefanov, O.; Klinge, M.; Koopmann, R.; Stellato,
     F.; Gabdulkhakov, A.; Schönherr, R.; Rehders, D**.; Lahey-Rudolph, J.M.**; Aquila, A.;
     Barty, A.; Basu, S.; Doak, R.B.; Duden, R.; Frank, M.; Fromme, R.; Kassemeyer, S.;
     Katona, G.; Kirian, R.; Liu, H.; Majoul, I.; Martin-Garcia, J.M.; Messerschmidt, M.;
     Shoeman, R.L.; Weierstall, U.; Westenhoff, S.; White, T.A.; Williams, G.J.; Yoon, C.H.;
     Zatsepin, N.; Fromme, P.; Duszenko, M.; Chapman, H.N.; Betzel, C.:
     In cellulo crystallization of *Trypanosoma brucei* IMP dehydrogenase enables the
     identification of genuine co-factors. *Nat Commun,* 11:620 (2020).
     DOI: 10.1107/S2052252521005297
     Nature Communication impact factor (2019): 12.12

Table of contents

Zusammenfassung

Die Aufklärung hochauflösender dreidimensionaler Strukturen von Proteinen mittels Röntgenstrukturanalyse erfordert Proteinkristalle zur Signalverstärkung. Proteinkristalle wurden in lebenden Zellen aller Domänen beobachtet. Ergänzend zu konventionellen Kristallisationsmethoden wurde in der vorliegenden kumulativen Arbeit ein systematischer Ansatz für die Kristallisation rekombinanter Proteine in lebenden Insektenzellen verfeinert und ausgeweitet.

Dabei wurde insbesondere die Detektion der Kristalle *in cellulo* verbessert, indem eine Kleinwinkelstreuung/ Pulverdiffraktions Methode zum Screenen von Zellkulturen entwickelt wurde. Dafür wurden mit rekombinanten Baculoviren infizierte Zellkulturen an einer Hintergrund-optimierten Synchrotron-Strahlführung vermessen und mit Mockvirus infizierten und mit uninfizierten Insektenzellkulturen verglichen. Sind Kristalle in der Zellkultur vorhanden, so liegen bei dieser Messung Beugungssignale in zufälliger Orientierung vor. Dadurch werden auf dem Detektor schwache unvollständige Debye-Scherrer Ringe aufgenommen, die durch Summation mehrerer Bilder, radialer Mittlung und Hintergrundsubtraktion in 1D-Plots der Streuwinkelkurve als Peaks erscheinen. Es konnte gezeigt werden, dass bei einer minimalen Kristallbildungsrate von unter einem bis etwa sechs Prozent ausreicht, um innerhalb von Sekunden Proteinmikrokristalle in der Zellkultur zu identifizieren. Die minimal notwendige Kristallbildungsrate ist v.a. abhängig vom expremierten Zielgen. Um die Effizienz und Sensitivität der Methode bestimmen zu können wurden dafür Proteine gewählt, die *in cellulo* deutlich mit Lichtmikroskopie sichtbare Kristalle bilden. Die Dynamik der Kristallbildung in den Zellen wurde im Ensemble untersucht. Dadurch konnte der optimale Zeitpunkt für das serielle Kristallographieexperiment ermittelt werden.

Serielle Kristallographie (SX) Beugungsdatenaufnahme mittels einer Synchrotron Strahlungsquelle und Freie Elektronenlasern wurde *in cellulo* - also ohne vorherige Extraktion und Reinigung der Kristalle aus den Zellen - etabliert, da sich gezeigt hat, dass kein erhöhtes Hintergrundsignal bei Datenaufnahme in lebenden Zellen auftritt, und die umgebende Zelle die Proteinkristalle vor etwaigen Schäden durch das Isolationsverfahren und vor Dehydration schützt. Am Synchrotron wurde anhand verschiedener Konstrukte der Proteine *N. crassa* HEX-1 und *T. brucei* Inosinmonophosphatdehydrogenase (*Tb*IMPDH) die Datenaufnahme in MicroMeshes (MiteGen) mit einem helikalen Linienscan gezeigt und die Strukturen gelöst. Auch die Datenaufnahme bei Raumtemperatur (RT) wurde demonstriert, wobei es durch die Verwendung von CrystalDirect™ Platten nach Beschichtung mit Poly D-Lysin möglich war, sowohl eine rekombinante Bakulovirus (rBV) Infektion, Kristallwachstum als auch die Datenaufnahme *in cellulo* und *in situ* durchzuführen. Dies vermeidet Schäden an potentiell fragilen Kristallen. Diffraktionsdatensammlung bei RT kann alternative Konformere aufdecken und ermöglicht automatisierbare Datenaufnahme

an bis zu 96 Proben ohne Trägerwechsel. SX *in cellulo* am Freie Elektronenlaser LCLS wurde mit einem *fixed-target* Verfahren erfolgreich demonstriert. Im Vergleich zur Probenlieferung mit einem Liquid Jet entfallen die Risiken von Verstopfung einer Düse oder Sedimentieren der Probe im Reservoir. Somit kann eine hohe Trefferquote erzielt werden.

Interessanterweise ermöglicht Kristallisation in der lebenden Zelle die Bindung von genuinen Kofaktoren, welche in physiologischer Konzentration in den Kristall diffundieren oder kokristallisieren. Wird auch die Datenaufnahme in intakten Zellen durchgeführt, können diese Kofaktoren bei ausreichend starker Bindung und Datenqualität identifiziert werden. Auf diese Weise konnte gezeigt werden, dass die inhibierende Konformation der *Tb*IMPDH ein Phosphat an der IMP Bindungsstelle inkorporiert, und in den kanonischen Nukleotidbindungsstellen GDP und ATP mit einer alternativen Konformation.

Für die Lösung der Struktur nach erfolgreicher Diffraktionsdatensammlung gibt es die Möglichkeiten des molekularen Ersatzes, sei es mit einer bereits experimentell gelösten homologen Struktur oder seit AlphaFold2 auch anhand eines *in silico* Modells. Um Bias zu vermeiden wurde angestrebt schwere Atome in intrazelluläre Kristalle einzubringen, anhand derer *de novo* Phasierung getestet werden könnte. Es wurden erfolgreich *Tb*IMPDH Konstrukte mit einem Lanthanid-Bindetag (LBT) intrazellulär kristallisiert. Diese Kristalle zeigten eine geringere Stabilität. SAD-Phasierung von Test-Diffraktionsdaten der *Tb*IMPDH mit terminalem LBT war nicht möglich, weil Terbium intrazellulär nicht ausreichend spezifisch bindet und sich nach Kristallisolation die Auflösung der Diffraktionsdaten durch Terbium-Soaking reduziert.

Als alternativen Ansatz wurde die Inkorporation von Selen als Selenomethionin während der Genexpression getestet. Durch rBV Infektion der mit einem sehr hohen Virustiter (MOI 10) konnte eine erhöhte Inkorporation von Selen in den Kristallen mittels Massenspektrometrie für *Neurospora crassa* HEX-1 nachgewiesen werden. Dieser Ansatz ist vielversprechend für eine zukünftige SAD-Phasierung, sollte jedoch für weitere Proteine validiert werden.

Die Resultate dieser Dissertation leisten einen Beitrag zum Methodensprektrum der Strukturbiologie, indem die *in cellulo* Kristallisation für eine größere Gruppe Wissenschaftler durch vereinfachte Detektion und serielle Datenaufnahme in intakten Insektenzellen als Ergänzung zu konventioneller *in vitro* Kristallisation zugänglich gemacht wird.

Abstract

Elucidation of high-resolution three-dimensional protein structures via X-ray crystallography requires protein crystals for signal enhancement. Protein crystals have been observed in living cells from all domains of life. This cumulative dissertation refines and expands a systematic approach for crystallization of recombinantly introduced proteins in living insect cells, as a complementary method to conventional crystallization approaches.

In particular, detection of crystals *in cellulo* has been optimized by development of a small angle X-ray scattering/ powder diffraction synergistic method to screen cell cultures. To this end, recombinant baculovirus infected cell cultures were measured at a low-background optimized synchrotron beamline. Their X-ray scattering profile was compared to that of mockvirus-infected and uninfected insect cells. If protein crystals are irradiated, they are in random orientation within the cells. Consequently, weak incomplete Debye-Scherrer rings are recorded, which by summation of multiple detector images, radial averaging, and background subtraction show as peaks in the 1D-plots of the scattering curve. It could be shown that a crystallization efficiency in cells of less than one to six per cent suffices to identify within seconds if there are protein microcrystals within a cell culture. This is largely dependant on the expressed target protein. To this end, proteins were chosen that form clearly visible crystals identifieable by light microscopy. Thus, the efficiency and sensitivity of the established method could be elucidated. Additionally, the dynamic behaviour of crystal growth within the cell ensemble was investigated. This enabled determination of the optimal timepoint after infection for a serial crystallography (SX) experiment.

It became clear that data collection within living cells, meaning *in cellulo* with no prior extraction and purification of crystals, does not increase background noise. Going further, the surrounding cell even protects protein crystals from dehydration and damage during crystal extraction. In this work, SX data collection *in cellulo* was established at synchrotron radiation sources and using X-ray free-electron laser (XFEL) radiation. With synchrotron radiation, efficient data collection on MicroMeshes (MiteGen) with helical line scans is demonstrated on different constructs of the proteins *Neurospora crassa* HEX-1 (*Nc*HEX-1) and *Tb* inosine 5'-monophosphate dehydrogenase (*Tb*IMPDH), solving the structures. In addition, data collection at room temperature (RT) is demonstrated. Using CrystalDirect™ plates with a poly D-lysine coating for enhanced cell adhesion, everything from cell infection, crystal growth up to data collection in the intact cells can be done *in situ*. This way, any damage on potentially fragile intracellular crystals is prevented. Collecting at RT can also uncover potential alternative conformers. The method enabels automatable data collection on up to 96 sampels without changing the mount.

Successful SX *in cellulo* on XFELs is demonstrated with a *fixed-target* approach on micro-patterned silicon chips. Compared to liquid jets, risks of nozzle clogging or crystal sedimentation in the sample reservoir are eliminated, reducing sample consumption and increasing hitrates.

Interestingly, crystallization inside cells enables binding of genuine co-factors abundant at physiological concentrations. Ligands can either diffuse into or co-crystallize with the crystal. Collecting diffraction data in viable insect cells can lead to identification of such co-factors, provided sufficiently strong binding and data quality. This way it can be shown that the inactive conformation of *Tb*IMPDH incorporates phosphate at the IMP binding site, as well as GDP and ATP in alternate conformations at canonical nucleotide binding sites.

Solving the structure after successful data collection can be done by phase retrieval with molecular replacement. Phases can be transfered from a experimentally phased homologue, or more recently from *in silico* models generated f.i. with *AlphaFold2*. In order to prevent bias, it is still desirable to introduce heavy atoms into intracellular protein crystals, which later could be used for experimental *de novo* phasing. Towards this end, *Tb*IMPDH constructs with terminal lanthanide binding tags (LBT) were successfully crystallized *in cellulo*, unfortunately leading to increased crystal fragility compared to the *Tb*IMPDH *in cellulo* crystals without LBT. Moreover, unspecific binding of the lanthanide terbium to other moelcules within the cell was confirmed, and, terbium soaking reduced resolution of diffraction data. As an alternate approach, the incorporation of selenium in the form of selenium methionine was tested on different targets during gene expression. It was hypothized and proven by mass spectrometry for *Nc*HEX-1 that a high virus titer (MOI 10) results in an increased selenium incorporation. The potential of this approach for a future phasing via SAD needs to be validated on additional proteins.

Results of this dissertation contribute to the method spectrum of structural biology, opening the approach of *in cellulo* crystallization to a wider group of researchers. Crystal detection and serial data collection in living cells has been simplified and optimized, rendering the method an alternative crystallization method.

1 Introduction

The three-dimensional structure of a macromolecule is the key to its function because the specificity of active sites and binding pockets depends on the conformation. In order to understand how molecular mechanisms function, one needs to determine the atomic structure of the molecules. The method needs to fit to the scientific question, in particular to the particle size. For example, a propidium iodide (PI) stain is appropriate to determine the survival rate in a cell culture because PI intercalates between bases of nucleic acids but cannot cross an intact plasma membrane. To count the proportion of stained cells, a simple light microscope is sufficient. To investigate the three dimensional structure of single proteins, however, bond-lengths of a C-C single bond in the order of 154 pm (1.54 Å) have to be resolved.[1] To study interactions of such spatial resolution, the use of wavelengths equal or smaller than this typical inter-atomic distance is required, as described by Abbe's law: To resolve two objects from another, one is limited to approximately half the probing wavelength. With a typical photon energy of 12.7 keV at a macromolecular crystallography beamline, the X-ray probe has a wavelength of 0.976 Å and thus fulfils this criterion. This cumulative dissertation contributes to the establishment and optimization of methods to determine protein strutures at atomic resolution.

1.1 Major methods in structural biology

Besides X-ray protein crystallography, also nuclear magnetic resonance (NMR) spectroscopy and cryo electron microscopy (EM) can deliver structures of macromolecules of medium to high-resolution. Each method has advantages and disadvantages, as subsequently described.

Nuclear magnetic resonance (NMR) spectroscopy exploits the intrinsic magnetic field, the spin, of uneven atomic nuclei in a strong externally applied magnetic field. A spinning nucleus like a proton can adopt two states, either in-line (α-spin) or opposite (β-spin) the external magnetic field. When a pulse in form of a radio frequency (RF) is applied, some of the nuclei will be excited to the β-spin state if the frequency exactly matches the energy difference between the two states (which is the Larmor frequency).[2] To obtain a resonance spectrum, a constant magnetic field and variable RFs are applied to the macromolecules in solution. The nuclei are partly shielded by small electrical fields provided by the electrons around the nuclei, the degree of shielding depending on the specific electron density. Therefore, the chemical environment influences the Larmor frequency, seen in NMR spectra as chemical shifts that allow structural investigations. NMR techniques are often used to study protein-protein or protein-ligand interactions and their kinetics[3], but also provide useful tools for the investigation of macromolecular structure and dynamics in aqueous solution at RT.[4,5] NMR models of small molecules (30-40 kDa) are easily obtainable. Structure elucidation of proteins larger than 50 kDa is possible, but requires costly and tedious labeling

with ^{13}C, ^{15}N or [17]O, because of problems with overlapping peaks in the NMR spectra of large proteins.[6]

Electron microscopy (EM) and in particular cryo-EM has become an increasingly popular method in the structural biology community, due to huge and ongoing progress in terms of sample preparation and preservation in vitreous (non-crystalline) ice, improved electron optics, direct electron detectors, improved data processing software and faster computers.[7] In 2017, the Nobel Prize in chemistry was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"[8]. An electron beam together with a system of electron lenses is used to directly image biomolecules. Thousands of different single particles in different orientations that were immobilized in a thin layer of vitreous ice are fused with a computational approach to a 3D mass density map into which the atomic model of the macromolecule is fitted. To date it is possible to image macromolecular assemblies up to 2 Å resolution based on the overall shape of the molecule determined with cryo-EM.[9]

However, to date X-ray protein crystallography is still *the* method to obtain high-resolution structure models of proteins, proven by over 215,000 entries in the protein data bank (PDB, **Figure 1**) since the structure of sperm whale myoglobin was reported by Kendrew *et al*. in 1958.[10,11]

**Figure 1 Distribution of structure models of proteins and protein-nucleic acid complexes in the PDB**.
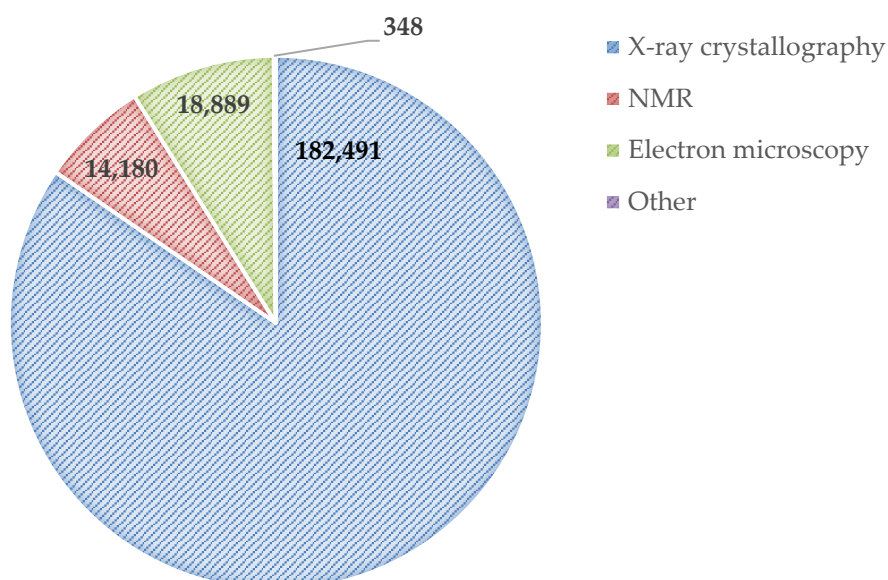
Shown as pie chart sorted by experimental method. Information extracted from RCSB PDB[10] on February 17th, 2024.

As of February 2024, around 85 % of experimentally determined and deposited structures were solved by X-ray crystallography, 6.6 % by NMR, and almost 9 % by EM. The number

of NMR structures deposited per year decreases, while the number of new EM-derived structures rises.

The basic setup of any crystallographic X-ray diffraction experiment consists of an X-ray source, at least one crystal with ordered planes denoted as h, k and l, and an X-ray detector (**Figure 2**). Most of the X-rays pass straight through the crystal, but a few scatter at the electron clouds of the macromolecules arranged in the crystalline lattice.

X-rays have a weak scattering cross section compared to electrons used in EM. This fact can be overcome by combining the signal of many weakly scattering biomolecules in a three-dimensional lattice, the crystal, so that the scattered X-rays are either amplified or distinguished by constructive or destructive interference according to Bragg's law. Part of the primary X-ray waves are scattered from the planes of the crystal and when the photons reach the detector, the position and intensity of every peak is recorded leading to a distinct peak pattern. [12,13]
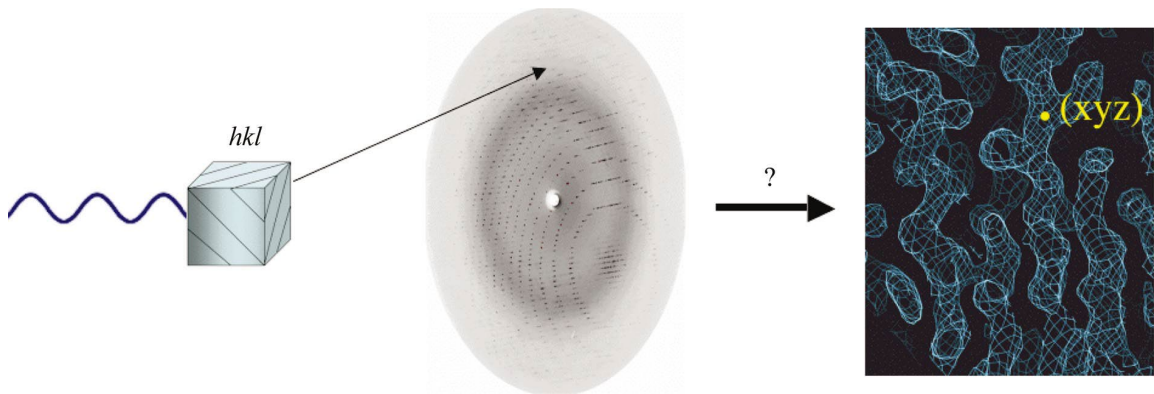
Figure 2 A generalized diffraction experiment.
X-rays scatter at the electron clouds of macromolecules arranged in a crystalline lattice. Positive interference of the scattered waves results in distinct spots on the detector, following Bragg's law. In this process the phase of the incident X-rays is lost. Therefore, the phases must be retrieved via an inverse Fourier transform to calculate the electron density map and subsequently model the structure into it. Image courtesy of Taylor, 2003.[14]

$|F_{hkl}|$, the amplitude of the wave, is proportional to the square root of the measured peak intensity. The electron density at a point in space (xyz) in the unit cell can be calculated by summing over all planes the sum of all scattered waves from a plane (hkl). The amplitude F_{hkl} depends on the number of electrons in the planes and on the phase of the wave.[14]

The following function mathematically describes the electron density ρ at a position (xyz).

$$\rho(xyz) = \frac{1}{V} \cdot |F_{hkl}| \cdot \exp(i\alpha_{hkl}) \cdot \exp(-2\pi\,ihx + ky + lz),$$

where V is the volume of the unit cell, and the phase associated with the structure-factor amplitude $|F_{hkl}|$ is denoted as α_{hkl}.

Amplitudes are measured. However, as mentioned in **Figure 2**, the phase of the incoming X-rays is lost. To retrieve the electron density map via an inverse Fourier transform and to

model the structure into it, phases thus have to be recovered first. This is denoted as the phase problem in crystallography. Methods to solve this are briefly explained in chapter 1.5.

Radiation damage

X-rays cause radiation damage. This is a major challenge noted early after discovery of X-ray imaging techniques by Wilhelm Röntgen in 1895. Radiation damage is caused by photoionization, followed by photoreduction due to diffusion and inelastic scattering of the photons. This leads to the reduction of bound metals, to bond-breakage and finally to the destruction of the molecule and the crystalline order. On a global level this is recognized by an overall drop in diffraction power and an increase of the temperature factors. Radiation-induced changes in the electron density are visible on a local level. Some amino acids (aa) are more prone to radiation damage than others and therefore act as useful damage indicators: disulfide bridges are cleaved, tyrosine residues dehydroxylated, glutamates and aspartates decarboxylated and methionines lose their C—S bond.[15–17]

Only scattered photons give rise to the diffraction pattern, whereas most pass through without any interaction, and many are absorbed by the macromolecule, which is effectively limiting the energy that can be deposited in a crystal. This in turn defines a lower limit for the crystal size that can be meaningfully investigated by synchrotron radiation. The limiting dose that can be deposited in a given crystal to allow data collection without damaging effects increased to the `Garman limit´ of approximately 30 MGy[18,19], when data collection was shifted from RT to cryogenic temperatures about thirty years ago.[20] Still, the challenge that only a small part of the incoming photons are scattered remains. Radiation damage limiting methods in data collection would be beneficial especially for small crystals with an accordingly small scattering power. On the other hand, photoelectron escape and secondary radiation damage produced by inelastic scattering events within the crystal may play a lesser role in very small crystals under favourable conditions, as theoretically claimed a decade ago[21,22] and recently experimentally determined.[23]

Consequently, during a diffraction experiment, the amount of energy deposited into any radiation-sensitive sample should be determined, to collect data of the highest possible quality. RADDOSE is a program developed by the group of Elspeth Garman for radiation dose estimation in protein X-ray crystallography. It considers the incident X-ray flux density and beam properties, the photon energy, the composition of the macromolecule and of the buffer, as well as the geometry of the crystal and the data collection strategy, to name the most important variables. Recently, also versions to model and monitor radiation damage for e.g. small angle X-ray scattering (SAXS) data collection, time- and space-resolved experiments or serial femtosecond X-ray crystallography (SFX) have been implemented.[12–14] Throughout this PhD work, RADDOSE[27,28] was used to calculate the absorbed energy dosage of irradiated crystals.

The standard method in macromolecular crystallography is oscillation data collection, where a single crystal is rotated in small increments during X-ray exposure, in order to sample the complete reciprocal space (**Figure 3 A**).[29] However, a more even distribution of dose per crystal volume can be achieved when adding elements of translation and offsets to the data collection strategy (**Figure 3 B-E**).[28,30–32]
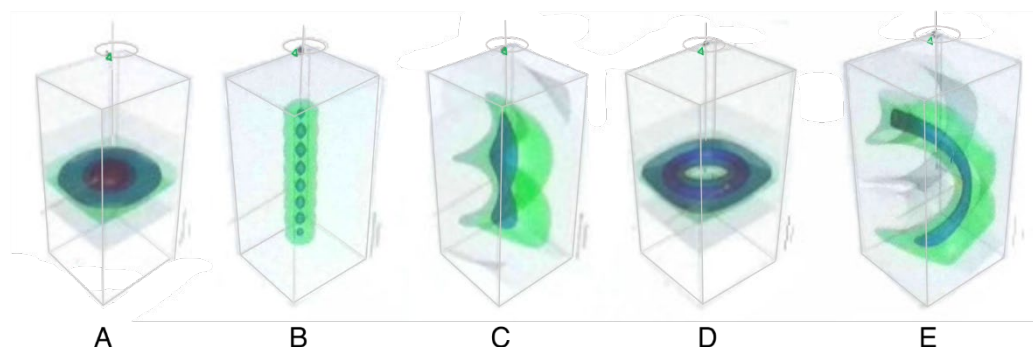
Figure 3 Dose distribution vs exposure strategy.
The distribution of energy dose in a crystal calculated with RADDOSE-3D is visualized in a green to blue heatmap, darker colours indicating areas of higher dose: 0.0001 MGy (grey), 5 MGy (green), 10 MGy (light blue), 20 MGy (dark blue), 30 MGy (red). Calculated doses apply for a 100 µm x 200 µm x 100 µm Gaussean beam with a FWHM of 20 µm x 20 µm using a 1 mm x 1 mm rectangular collimation[33].
(A) 360° oscillation data collection, without translation or offset elements. Radiation damage clusters in the crystal center. (B) An exposure strategy with 20 µm translation. (C) A helical data collection approach, 0.2 µm translation per degree. (D) A donut dose distribution resulting from an oscillation exposure and in addition a 30 µm offset. (E) A helical scan with a 30 µm offset and 0.2 µm translation per degree.
Image kindly provided by Elspeth Garman at the ECM32 in 2019.

Major developments in synchrotron radiation brilliance and the invention of XFEL radiation, enable the use of ever smaller crystals for diffraction data collection. Now, diffraction can be recorded from µm- to even nm-sized crystals. However, they are too small to collect a complete dataset by single crystal-rotation, because of the local radiation damage. Usually, thousands of diffraction patterns taken from µm- or nm-sized crystals at random orientation are sampled to form a merged dataset.[34,35] This approach was termed serial crystallography (SX), with major developments ongoing in terms of sample delivery methods and data processing algorithms. Nowadays, proof-of-principle studies with model proteins using e.g. a pink beam indicate that the number of required crystals can be further lowered to a few hundred. [36–40] If small rotational series can be recorded from single crystals with high symmetry, as little as 80 crystals can be sufficient, as demonstrated for Cathepsin B (CatB) isolated from Sf9 insect cells using a synchrotron radiation source.[41]

Radiation sources: Synchrotrons and free-electron lasers

Synchrotrons

A synchrotron radiation source is a particle accelerator or storage ring in the shape of a polygon. Charged elementary particles like electrons (e.g. at PETRAIII at DESY in Hamburg) or protons (e.g. at the Large Hadron Collider, LHC at CERN, near Geneva) are accelerated to relativistic speed in a synchronized manner. The path-curvature produced by strong magnets in the storage ring forces the electrons to emit electromagnetic radiation *i.e.* X-rays that can be focused and shaped to a beam with aspired characteristics, using dedicated beamline optics. Wigglers or undulators are insertion devices. Their name derives from their feature to *wiggle* electrons through a high undulating magnetic field produced by permanent magnets of alternating polarity. The electrons in the storage ring lose energy through emittance of X-ray radiation, which needs to be compensated for for continuous X-ray generation. In modern synchrotrons, electrons are produced in an electron gun, then pre-accelerated with a linear accelerator also called linac, and then continuously fed into the storage ring in bunches (**Figure 4 A**).
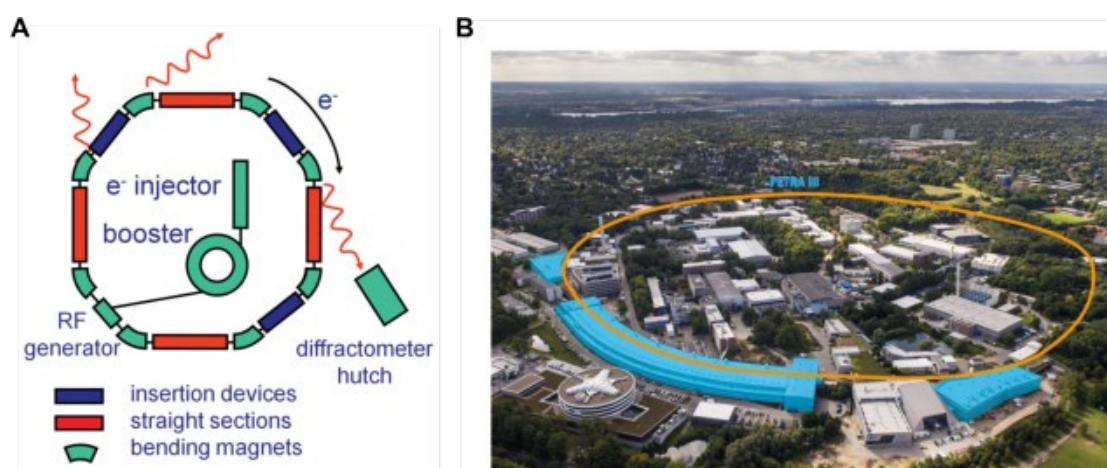
Figure 4 Schematics of a synchrotron.
A Basic components and their location in a synchrotron storage ring are depicted. Electrons are generated in an electron gun or electron injector. They are pre-accelerated in a booster ring before being injected to the storage ring accelerator. Electrons lose energy when they pass bending magnets, as the Lorentz force is changing the impulse of the electrons in the ring or rather polygon, leading to energy conversion into X-ray radiation of the circling electrons. A radio frequency (RF) generator compensates for this energy loss, providing a small "kick" on each pass. Beamlines are arranged in a tangential-line manner from the ring. Here, beamline-optics narrow down the polychromatic X-rays to a focused monochromatic beam with the aspired characteristics that finally reach the diffraction equipment in the experimental hutches. Left image from Biomolecular Crystallography by Bernhard Rupp, © 2009-2014 Garland Science/Taylor & Francis LLC. **B** The Positron-Electron Tandem Ring Accelerator PETRA III highlighted on the DESY Campus (Hamburg). Radiation is produced whenever electron bunches are transversely accelerated, as with the bending magnets and undulators signified in the left panel as insertion devices. PETRAIII supports 14 beamlines with 30 experimental stations. Image courtesy by DESY.

The pulses at PETRA III (**Figure 4 B**) last about 100 ps and can be ordered in bunches of 8 ns, 16 ns, or 192 ns, depending on the desired application. Synchrotron radiation is exceptionally bright over a wide energy spectrum that is determined by the critical energy of the source.[42]

Synchrotrons can be classified into four generations of development, with main characteristics listed in **Table 1**. The Positron-Electron Tandem Ring Accelerator 3rd generation syncrotron source (PETRAIII) is a 3rd generation storage ring at DESY. Currently, an upgrade to PETRAIV with damping wigglers and new undulators is in the planning phase, promising an increased brightness by a factor 500, using the same geometry as PETRAIII (Harald Reichert, DESY User Meeting 2024). All diffraction experiments with synchrotron radiation described in this work were measured at PETRAIII.

Table 1 Classification of synchrotron radiation sources and their main characteristics.

Generation of synchrotrons	Main characteristics
1st generation	Parasitic use of particle accelerators
2nd generation	Dedicated machines with damping wigglers
3rd generation	Synchrotrons including undulators
4th generation	Hybrid multi-bend achromats (HMBA)

It has been a long way and joint effort of multiple disciplines from parasitic use of bremsstrahlung (photons emanating from circular particle acceleration due to bending magnets) in synchrotrons of the 1st generation, built to study particle physics, to sophisticated facilities and beamlines dedicated for protein crystallography and material science. The storage ring PETRAIII harbors four beamlines that can be used for protein X-ray scattering experiments: P11 (DESY bio-imaging and diffraction), P12 (SAXS), P13 (macromolecular crystallography with crystal diameters > 20 µm), P14 (micro-focus, SX) and P14.2 (characteristics of P14 with flexible setup possibilities dedicated to time-resolved crystallography, TRX). The following Table 2 summarizes their beam characteristics.

Table 2 Beamline characteristics of PETRAIII undulator photon sources P11, P12, P13, P14 and P14.2.
Beam sizes are given in µm² (VxH). Using the PILATUS6M detector, shutterless rotation data can be collected at 25 Hz, the sensor thickness is 45 µm and count rate is 10^6 ph/s. SEC-SAXS = in-line size-exclusion chromatography. TRX = time-resolved crystallography

Beam-line	Min. crystal dimension	Beam sizes	Photon flux and energy tunability	Possible specials	Detector	Mounting
P11[43]	< 4 µm (or tissue)	Full-flux 4 x 9 µm² up to 300 x 300 µm²; Micro-beam 1 x 1 µm²	10^{13} ph/s at 12 keV; 5.5 – 30 keV > 2 x 10^{11} ph/s	Pink beam; coherence preservation; tapedrive SX[44]	EIGER2 X 16M (Vortex-EM)	Manual or automated: StäubliTX60L robotic arm + DeepPlunger

P12[45]	For macro-molecules, nanoparticles and bio-nano-conjugates	0.2 x 0.12 mm	10^{13} ph/s 4.1 - 20.6 keV = 3 – 0.6 Å	Low-background optimized; High-throughput; SEC-SAXS	PILATUS 2M & (since 2018) 6M EIGER	BioSAXS sample changer
P13[46]	20 μm	Micro-beam 30 x 20 μm²; Defocused full beam 150 x 70 μm²; Collimation down to 10 x 10 μm².	5 x 10^{12} ph/s at 12 keV 4.5 - 17.5 keV equivalent to 0.7 - 2.7 Å	Low energy collection < 5 keV with He cone	PILATUS 6M	Manual or MARVIN robot
P14	5 μm	Micro-beam 5 x 10 μm² Collimated beam Up to 300 μm	10^{13} ph/s at 12 keV 6 - 20 keV equivalent to 2.1 - 0.6 Å	*In situ* data collection; Serial data collection with helical line-scans	Since 2018: EIGER 16M Before: PILATUS 6M	Manual or MARVIN robot with SPINE pins and pucks
P14.EH2	5 μm	15 x 10 μm²	$2 \cdot 10^{12}$ ph/s 12.7 keV equivalent to 0.98 Å	TRX ms - μs; chopper disc providing short pulses; fibre optics laser	PILATUS 2M or EIGER 4M	Manual in a flexible sample environment, no goniometer

X-ray free-electron lasers

In contrast to synchrotrons, XFEL are linear accelerators. They produce radiation pulses with a high peak brilliance, high coherence and deliver as many photons in one single pulse in the time-range of a femtosecond as a third-generation synchrotron source provides in one second.[47] A good and simple explanation of the physical principles of an XFEL that results in the formation of microbunches can be found in a topical review by Ribic et al., 2012.[48] The invention of self-amplified spontaneous emission (SASE) mode of hard XFELs has enabled a revolution in peak-brilliance[49,50] that implies the possibility for macromolecular crystallography to collect high-resolution diffraction patterns from weakly diffracting matter such as μm-sized crystals made up of proteins[51,52], among them membrane proteins[53,54]. A typical X-ray bunch of an XFEL is 10^{12} photons per bunch, corresponding to an energy deposition that will destroy the sample in several tens to hundreds of femtoseconds.[34]

However, the pulse duration is so short, that the diffraction data are recorded before most chemicophysical changes associated with radiation damage can manifest.[55] This approach is known as "diffraction-before-destruction" and applies in serial femtosecond crystallography (SFX)[56]. The concept to keep the pulse-duration shorter than radiation damage dynamics was

first described by Neutze et al. in 2000[57]. It allows to probe samples with significantly more photons than possible using synchrotron radiation. In favourable cases, SFX is possible even from crystals with less than 100 unit cells.[51,58] Ultimately, the experimental goal for XFELs is diffraction and structure elucidation using single particles.[59]

Protein crystallization

In order to apply X-ray crystallography until single particle diffraction from proteins is possible, the biological macromolecule has to adopt to a well-ordered crystalline state. This is often a tedious, time- and resource-consuming process to achieve, as there is no inherent relationship between the primary structure of a protein (the aa sequence) and the crystallization conditions for proteins. Crystallographers raster the phase diagram for every protein by varying choice and concentration of precipitant, pH, temperature or additives (**Figure 5**) in order to find suitable crystallization conditions.[60,61] Even though robot assisted screening is frequently used to date to reach supersaturation of the purified protein, in combination with year-long dedication of research time, crystallization of a specific protein can totally fail. The standard crystallization methods are (i) micro-batch, (ii) vapour diffusion, (iii) dialysis and (iv) free interface diffusion.
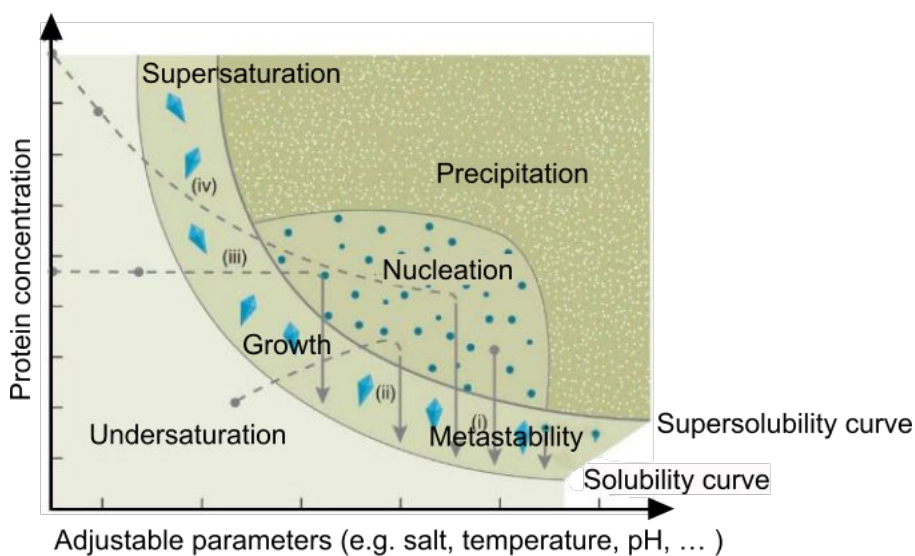
Figure 5 Phase diagram of a protein.
The phase diagram shows the ambient condition or adjustable parameter against protein concentration and classifies resulting states of the protein as a function of the solubility curve. Crystal nucleation occurs in a zone of supersolubility. Crystals ideally grow bigger starting from nuclei in the metastable zone. There are different established methods of protein crystallization. Ways within the phase diagram of (i) micro-batch, (ii) vapour diffusion, (iii) dialysis and (iv) free interface diffusion trials are depicted with dashed arrows. Image modified from (Govada and Chayen, 2019).[60,62]

Before the start of my PhD project, it was already clear that some proteins crystallize in living cells under favourable conditions. Chapters 1.2 and 1.3 review the current knowledge on protein crystallization in living cells, either as a native process or as a consequence of recombinant gene expression. Subsequently the proteins *Trypanosoma brucei* cathepsin B and IMPDH, *Nc*HEX-1 and human nicastrin are introduced as models used in this work to

optimise data collection approaches of intracellular protein crystals, detection methods, and X-ray structure solution.

1.2   Native *in cellulo* crystallization

Protein crystallization in living cells has been observed with surprising frequency over the past decades, often as a native assembly process. The earliest observation of protein crystals in human tissue and in seeds of the Brazil nut was recorded in the 1850s.[63,64] Crystals have been found in all kindoms of life and without an inclination towards a specific cellular compartment,[65–70] as we summarized in 2018.[71] Most intracellular crystals have been identified by EM due to their dense packing and regular morphology, with unclear identity and function of the assembled entity. Native *in cellulo* crystallization is often beneficial for the organism. Specific properties of the crystalline state of the specific protein can be favourable, such as stability or triggerable formation or disassembly of the crystal lattice, as highlighted below. However, intracellular protein crystals can also occur under pathological settings.
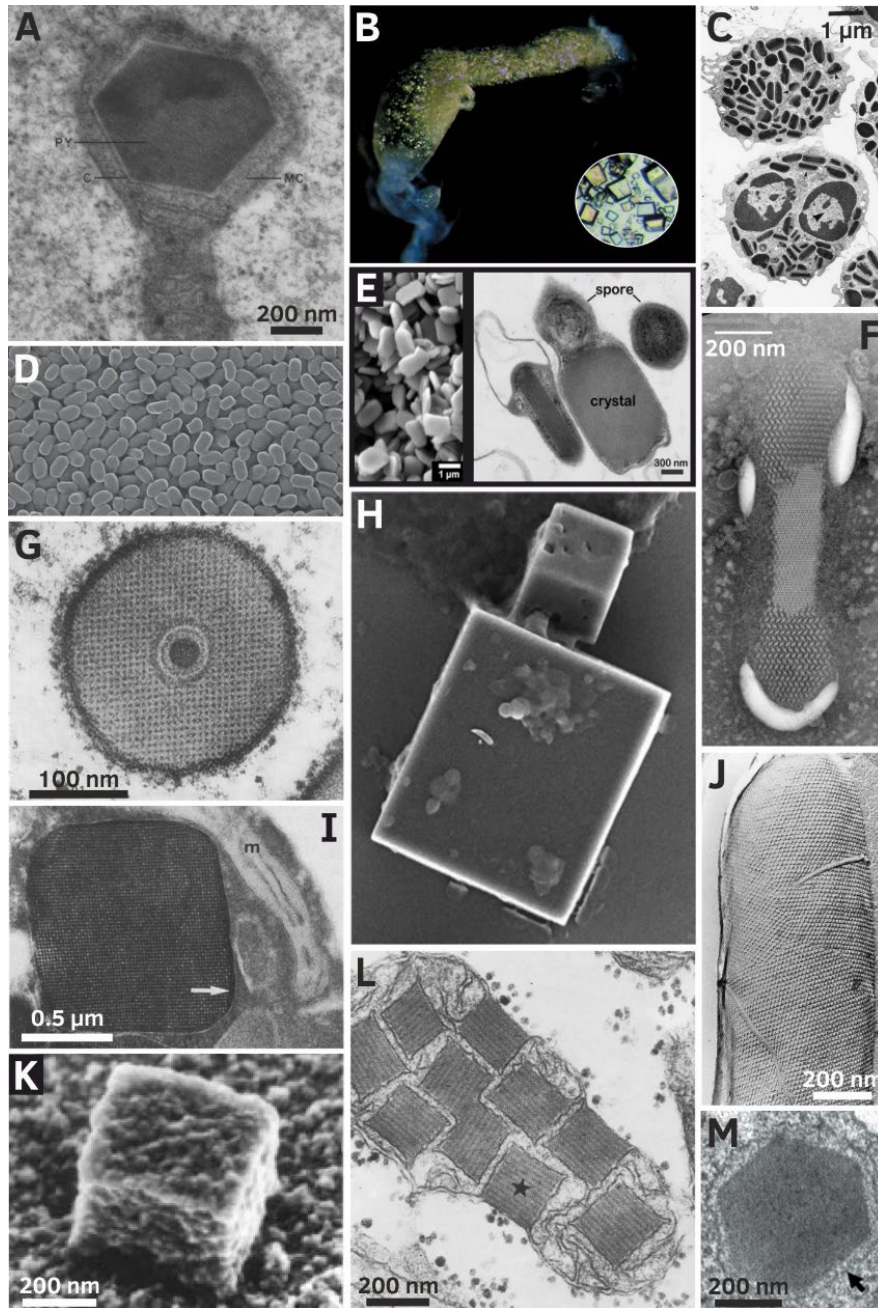
Figure 6 Protein crystals natively crystallized in living cells show different morphologies.
(A) A hexagonal crystal embedded in the mitochondrial cortex of a *Rana pipiens* oocyte that develops yolk platelet crystals.[72] (B) *In vivo*-grown Lili-Mip crystals in the embryo midgut of *D. punctata* are triclinic.[73] (C) MBP-1 crystals vesicle-enclosed in eosinophil granules have an orthorhombic lattice.[74] (D) Granulovirus (CpGV) OBs purified from infected larvae of *C. pomonella* cotain cubic crystals.[31] (E) Rectangular platelet crystals of the Cry3A toxin from *B. thuringiensis subsp. morrisoni* expressed in *Bt subsp. israelensis*, before (left panel) their isolation from bacterial cells and after (right panel)[75] (F) An expansion-intermediate of the crystalline core from a secretory granule of *Tetrahymena thermophile*.[76] (G) A granulovirus rod in its capsule within an infected larva of *Plodia interpunctella*.[77] (H) Isolated cubic baculoviral polyhedra from G25D *Autographa californica* multicapsid nucleopolyhedrovirus (*Ac*MNPV).[78] (I) Crystalline alcohol oxidase in peroxisomes in methanol-grown cells of the yeast *Hansenula polymorpha* are cubic.[79,80] (J) A bacterial cell is covered by a crystalline S-layer coat.[81] (K) Catalase-containing a crystalline cubic-shaped peroxisomal core extracted from 5-days-old sunflower cotyledons.[82] (L) Paracrystalline mitochondrial inclusions in normal human skeletal muscle.[83] (M) Hexagonal HEX-1 crystal in a Woronin body of the fungus *Neurospora crassa*.[84] Image from Schönherr, Rudolph and Redecke (2018).

One of the functions assigned to native intracellular protein crystals is protein storage. The packing of proteins inside the crystal lattice can be a space-efficient mode of either permanent concealment or temporary storage[71]. Changes in the environment can trigger the disassembly of protein crystals, thus releasing high concentrations of the active protein. This is useful for nutrient storage and often observed in nature. Oviparous species such as *Drosophila melanogaster*[85], frogs[86], bony fish[87] or the mosquito *Aedes aegypti*[88] contain crystalline yolk proteins in developing oocytes that serve as a constant supply of nutrition. Another striking example was found in the viviparous cockroach *Diploptera punctata*[89,90]. In the midgut of its embryo, crystals form from incoming liquid milk in a heterogenous mixture, a very effective compression of nutrition. In 2016, Banerjee et al. solved the structure of these milkproteins, at atomic resolution, from isolated *in vivo* crystals by applying SFX techniques.[73]

In addition, virulence factors like the parasporal toxin of *Bacillus thuringiensis*[75,91] and *Lysinibacillus sphaericus*[51] profit from the crystalline stage as a possibility of long-term survival sheltered from proteolytic degradation. Insect poxviruses form crystalline spindles[92] that disassemble quickly upon ingestion by an insect. This causes an intoxication process finally allowing bacterial endospores to germinate[93] or damage the chitin-rich paratropic matrix within the gut of the insect, enabling poxvirus infection[94]. Virulence factors can have an economic impact or are an interesting target in agriculture[93], because they are usually host-specific, exhibit few side-effects and are biodegradable. High-resolution X-ray crystallography structures have been solved from *in cellulo* grown crystals of poxvirus fusolin[94], and from the *Bt*Cry3A toxin BinAB[51,95,96] and *Bt*Cry11Aa and *Bt*Cry11Ba toxins[91].

Furthermore, proteins involved in secretory pathways frequently adopt crystalline states or very dense amorphous structures in vesicles. The advantage is that these structures can readily be dissolved, triggered by an external signal[97] like a pH-shift[98] or calcium ion release[76], segregating the protein out of the cell. Effects range from regulatory switch[99], safe storage of cytotoxic proteins of the immune system[74,100] to a remarkable directed pushed movement of the cell induced by the quick expansion of the crystalline cargo of granulines.[69,76,101]

In addition to protein storage, native *in cellulo* crystal lattices can protect and stabilize. Many viruses form paracrystalline arrays that, by providing a stable matrix in the infected host cells, facilitate an ordered virion assembly. In addition to EM studies, these arrays have recently been investigated both by SFX and cryo-eletron tomography.[58,102] Cytoplasmic (CPV) and nuclear polyhedrosis viruses (NPV), granular viruses as well as entomopoxviruses encapsulate their virions during the late state of infection inside a stable crystalline coat that consists of either polyhedrin, granulin, or spheroidin and protects the virions from rough environmental conditions[103,104] to remain infectious over years[105]. Once taken up by insect larvae, the crystalline coat dissolves in the alkaline gut and frees the infectious viral particles[71,106]. Polyhedra are the best-studied examples of native intracellular

protein crystallization. Two unrelated insect viruses, CPV of the family *Reoviridae* and NPV of the family *Baculoviridae* produce polyhedra[78,104,107]. Meanwhile, three different NPV and nine different CPV polyhedrin high-resolution structures have been solved[78,107–110]. They exhibit an extremely low water content of only 19 %, which can explain the remarkably high stability of these crystals. This, in combination with a very high crystal-growth efficiency in insect cells has made them a popular target for the development of crystallographic methods[109,111,112]. In procaryotes and archaea, the mechanical stability of crystalline structures is also known to be involved in various functions like DNA protection, cell wall stabilization and compartmentalization.[113–118]

The third functional observation on naturally occuring intracellular protein crystals is solid state catalysis, the assembly of active enzymes in a crystalline lattice.[71] In the peroxisomes of yeast cells nourished with methanol, active alcohol oxidase forms crystals[67], sometimes in composition with other enzymes[119]. The same phenomenon occurs in higher plant peroxisomes[82] and rat liver cell microbodies[120].

In addition to *in cellulo* crystallization as a native and regulated process with often beneficial function for the organism, proteins that are soluble under normal circumstances can 'accidentally' crystallize inside cells as a consequence of either a non-silent mutation or because of environmental changes. This phenomenon is often associated with disease. It is mostly unclear whether crystal formation causes the disorder, facilitates it or whether crystallization is merely a harmless side-effect[71,121].

Although intracellular protein crystallization is a fascinating and beautiful phenomenon by itself, for a structural biologist it would gain relevance if it was not only restricted to a limited number of proteins determined to assemble in a crystalline lattice naturally, but if crystallization of interesting protein targets could be artificially induced in cells. Increasing evidence of *in cellulo* crystallization of recombinant proteins in combination with the advance of radiation sources and of detector technique foster this interest.

1.3   *In cellulo* crystallization of recombinant proteins

There are several reported examples of *in cellulo* protein crystallization after recombinant gene expression, observed so far either in plant cells, in chicken and mammalian cells, and in insect cells infected with recombinant baculovirus. Furthermore, *B. thuriengiensis* proteins have been found to crystallize in bacterial cells[91,122], while hexagonal crystals of *N. crassa* HEX-1 have been observed using TEM in yeast cell, pointing towards a more general phenomenon.[71,123]

To date, five recombinant proteins were reported to crystallize in mammalian cells:

1)   human IgG[124],
2)   a genetically engineered coral fluorescent protein called 'crystallizable and photo-activatable' (Xpa)[125],
3)   a complex of the metazoan-specific kinase PAK4 and its endogenous inhibitor Inka1; The protein complex still forms diffracting crystals with green fluorescent protein (GFP) fused to the inhibitor[126],
4)   the nonstructural reoviral protein μNS fused to GFP, although crystallinity of these ordered inclusions was not proven.[127]
5)   human neuraminidase[128–131].

In insect cells, the afore-mentioned various versions of very robust and well-diffracting viral polyhedrin form crystalline inclusion bodies as a native property of the baculovirus. Therefore, polyhedrin is not considered as an *in cellulo* crystallizing recombinant protein in this work, although admittably crystallization only occurs in infected cells. Five other recombinant proteins have been reported to crystallize in either Sf9, Sf21 or High Five insect cells before starting this work:

1)   a heterodimeric phosphatase calcineurin, with one human and one fungal subunit[132],
2)   firefly (*Photinus pyralis*, *Pp*) luciferase[101]
3)   the IMPDH from *T. brucei*[133,134],
4)   the reoviral fusion protein GFP-μNS also reported in mammalian cells[101,127], and
5)   the glycosylated cysteine-protease cathepsin B from *T. brucei*.[41,133]

Two of them, cathepsin B and IMPDH, have been used in this thesis as model proteins to optimize serial *in cellulo* crystallography in living insect cells. They were supplemented by fungal HEX-1 and the ectodomain of human nicastrin that previously had not been reported to crystallize in insect cells. The following sections introduce these model proteins in more detail.

Cathepsin B from *Trypanosoma brucei*

Human African trypanosomiasis (HAT), widely known as sleeping sickness, is a severe parasitic disease transmitted by tsetse flies that is endemic in 36 sub-Saharan countries and is caused by the human pathogen *T. brucei* (*Tb*)[135]. After extensive efforts in isolating *Tb*-infected patients, to date *Tb* is causing an estimated number of actual cases below 10 000

while the population at risk comprises 65 million people. The pathogen first multiplies in subcutaneous tissues, blood and nymph during the haemo-lymphatic stage, where the infection mostly stays undetected. In the second, meningo-encephalic stage, the parasite crosses the blood-brain barrier and causes severe neurological damage that leads to coma and the eventual death of the host.[136]

The papainlike cysteine protease cathepsin B (CatB) is involved in host-protein degradation and infectivity of *Tb*, and is a validated drug target for the development of new safe and effective treatments against trypanosomiasis[137–140]. More specifically, CatB catalyzes peptide degradation in the lysosome, both as an exo- and an endoprotease, by means of a catalytic triad consisting of Cys-42, His-282, and Asn-302. Release of specific aa is essential for metabolic homeostasis of the parasite.[141,142] During translation, the 22 aa-long N-terminal signal peptide is released, and preproCatB is directed into the rough endoplasmic reticulum and as glycosylated proCatB further to the Golgi apparatus. At neutral pH, the CatB propeptide acts as natural inhibitor of the enzyme, with stable binding to the active center through several interactions with the substrate binding region S2′ and the highly flexible occluding loops. However, proCatB is directed to the acidic environment of the lysosome by receptor-mediated transport and here, the proenzyme is activated by removal of the propeptide. His-194 is protonated and thus free for a salt-bridge interaction with Asp-115. Thereby the S2′ binding pocket is blocked, preventing binding of the propeptide.[144,145]

The structure of glycosylated *Tb* proCatB determined using isolated *in cellulo* grown crystals was the first novel structural bioinformation solved by applying SFX at a XFEL, and confirmed the feasibility of the "diffraction-before-destruction" approach. Significant differences in propeptide binding between human and *Tb* CatB render the enzyme an attractive drug target for trypanosomiasis.[41] The successful determination of a protein structure from *Tb* CatB crystals grown in insect cells strongly motivated the work presented in this thesis.

Inosine 5′-monophosphate dehydrogenase from *Trypanosoma brucei*

IMPDH is a key enzyme for *de novo* purin nucleotides synthesis. Inosine 5′-monophosphate is the precursor of both adenine and guanine nucleotides and the product of *de novo* synthesised purine nucleotides. IMPDH catalyses the reaction from inosine 5′-monophosphate to xanthosine 5′-monophosphate (XMP). This is the rate-limiting step in guanosine nucleotide synthesis. XMP is subsequently converted to guanosine monophosphate (GMP) by the guanosin monophosphate synthetase. GMP is not only needed for the synthesis of DNA and RNA. In the form of GTP it is pivotal as the energy source for translation, as a precursor for glycosylation, signalling molecules and allosteric regulators, and as a cofactor for G-proteins[144,145].

In particular, strongly proliferating cells have an increased demand for guanine nucleotides, and in addition, IMPDH is involved in the regulation of cellular proliferation by p53. This renders IMPDH a promising drug target against cancer.

A number of IMPDH inhibitors, both for immunosuppressive chemotherapy (mycophenolic acid and mizoribine) and antiviral chemotherapy (ribavirin) have been developed and are in use[146,147]. However, there are salvage pathways with phosporibosyltransferases and/ or nucleoside phosphotransferases and kinases that can also produce guanine nucleotides. Susceptibility to IMPDH inhibitors depends on the ratio between salvage and IMPDH/GMPS pathway activity of the organism in question. With the exception of a few protozoan parasites, IMPDH is present in every organism. There are two human IMPDH isoforms that share 84 % sequence identity. In general, *hIMPDH2* is upregulated during proliferation and transformation of cells, while *hIMPDH1* is constitutively expressed at low levels.[145]

It was previously published that *Tb*IMPDH forms intracellular protein crystals after recombinant gene expression in Sf9 insect cells.[148] Due to remarkable differences in the metabolic pathway for the production of purine nucleotides between parasites and its human hosts, *Tb*IMPDH was proposed as a suitable target for anti-trypanosomal drugs against HAT. This was strongly supported by enzymatic differences of *Tb*IMPDH compared to human IMPDH of both isotypes.[149] Especially the K_m for NAD$^+$ of *Tb*IMPDH is considerably higher than that of mammalian enzymes and evidently exhibits a different NAD$^+$ binding mode. In contrast to mammalian cells that are able to synthesize purine nucleotides *de novo*, *T. brucei* is auxotrophic for purines and therefore critically depends on salvage mechanisms that include the IMPDH-catalyzed reaction to acquire purine nucleotides from the extracellular environment. A *Tb*IMPDH-specific inhibitor is thus postulated to disrupt an essential biological pathway that of the parasite's life cycle[149]. Structural insights into *Tb*IMPDH may therefore pave the way towards the development of new specific inhibitors that are needed because of the limitations of current HAT treatment and the emergence of drug-resistant trypanosomal strains.[150,151]

HEX-1 from *Neurospora crassa*

Ascomycete fungi grow by invasive hyphal tip growth, where individual cells of the multicellular hyphae are separated from each other by perforate septa, allowing cell-to-cell communication.[152] This causes a serious threat if the syncytium is wounded, because this would lead to an extensive cytoplasmic bleeding. However, fungi can protect their cells by septal pore occlusion, where a plug is formed that seals the cell and thereby prevents cytoplasmic bleeding.

HEX-1 is the major component of peroxisome-derived Woronin bodies in ascomycete fungi, such as in *Neurospora crassa* (*Nc*). Upon cell injury, HEX-1 self-assembles and forms a crystalline structure that stabilizes the Woronin bodies. They in turn rapidly translocate to the septal pore, acting as a 'plug'.[153]

The application of sitting drop vapor diffusion generated recombinant HEX-1 crystals. These have been used to solve the structure of *Nc*HEX-1 by X-ray crystallography at 1.8 Å in 2003.[84] The HEX-1 structure is remarkably similar to that of the eukaryotic initiation factor 5A (eIF-5A). Only the aa required for *Nc*HEX-1 self-assembly and the N-terminal tag required for peroxisomal targeting are absent in eIF-5A. Yuan *et al.* suggested that the protein's new function, septal pore sealing, evolved from an ancestral eIF-5A gene dublication[84].

Throughout this thesis, *in cellulo* grown crystals of a cytoplasmic construct of *Nc*HEX-1, named HEX-1 cyto (vector map **Supplemental Figure 1**), serves as a model to evaluate different data collection and processing methods. The amino acid sequence is identical with the native *Nc*HEX-1 except for an N-terminal glycine, and a C-terminal adenine residue, artefacts of the applied blunt-ending cloning strategy. The wildtype protein is produced by the construct named HEX-1 ori.

Human nicastrin

γ-Secretase is an intramembrane aspartyl protease that cleaves type I membrane proteins. It is an integral membrane protein complex required for Notch signalling pathways that are involved in cell fate determination. The γ-Secretase is important for processing of the amyloid precursor protein (APP) to amyloid beta (Aβ) peptides of various lengths. A subset of these Aβ peptides – mainly Aβ40 and Aβ42 – can form stable extracellular aggregates in the brain, which leads to the formation of amyloid plaques. This triggers a neurodegeneration cascade and ultimately leads to a massive loss in neurons and synapse degeneration, thus closely linking γ-Secretase to Alzheimer's disease.[154] The protein complex is composed of four subunits, namely anterior pharynx defective 1, presinilin, presinilin enhancer 2, and nicastrin.

Nicastrin is a single-pass transmembrane glycoprotein responsible for substrate binding and selectivity. A crystal structure of nicastrin from *Dictyostelium purpureum* at a resolution of 1.95 Å is available[155], but this isoform has a very low sequence identity of only 23 % with human nicastrin. A high-resolution X-ray crystallographic structure of human nicastrin does not exist so far. However, such a structure would help to understand the molecular basis of the disease and to enable the development of specific activity and selectivity modulators for γ-secretase that could act as potent drugs to treat Alzheimer's disease. Bai *et al.* determined an atomic structure of human γ-secretase at 3.4 Å resolution by single-particle cryo-electron microscopy.[156] Recently, Zhou *et al.* determined a 2.6 Å cryo-EM structure of a mutant human γ-secretase cross-linked to APP C83, including the nicastrin domain.[157] In this work (chapter 3.3.1), the ectodomain of nicastrin (Nct ECD SKL) was crystallized in insect cells. Nct ECD is essential for substrate recognition by the γ-secretase.[158,159] The Nct ECD SKL misses the C-terminal transmembranous helix, which increases the solubility of the protein domain. Instead, the construct includes a C-terminal translocation signal to peroxisomes, consisting of the three amino acids serine, lysine and leucine, hence "SKL".

Viral life cycle of *Baculoviridae*

The genes of the aforementioned proteins are integrated into insect cells via a Bac-to-Bac expression vector system (Invitrogen) to allow *in cellulo* crystallization of the recombinant proteins with higher efficiency than can be achieved by direct transfection[101]. The baculovirus *Ac*MNPV of the family *Baculoviridae* is a Baltimore group I virus with dsDNA that replicates in the host nucleus. The genome of *Baculoviridae* is packed within the rod-shaped viral capsids, which are themselves embedded in a lipid envelope. Two phenotypes can be distinguished in the infection cycle: On the one hand the occlusion-derived virus (ODV) and on the other hand the budded virus (BV). ODVs are responsible for the primary infection and occluded in a protein matrix usually consisting of polyhedrin.
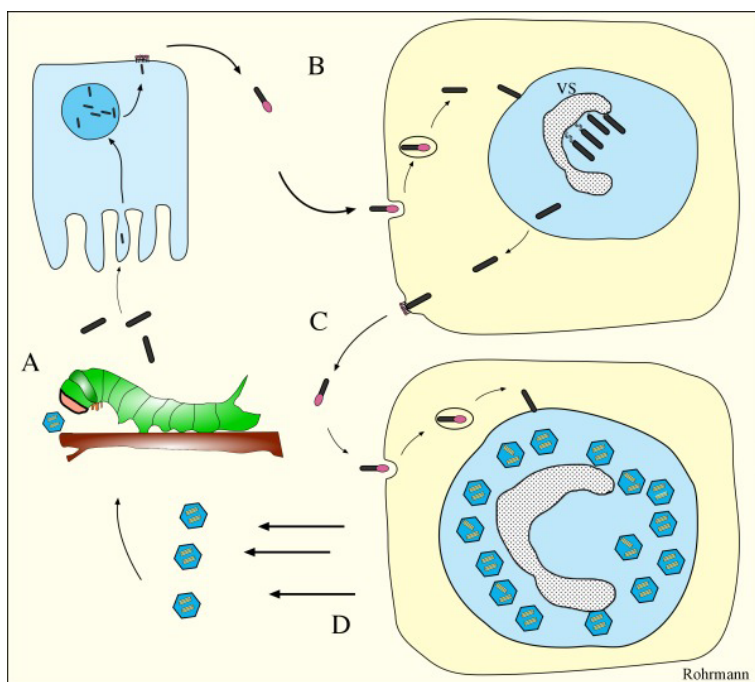
Figure 7 Life cycle of a baculovirus causing systemic infection in insects.
A An insect ingests occlusion bodies (OB) that dissolve in the insect midgut. Subsequently, ODV are released which in turn infect epithelial cells of the midgut. **B** A systemic infection is initiated by out-budding of the virions in a basal direction. **C** More BV are produced early during the systemic infection and viral spread occurs. **D** Occluded virions are produced late in infection cycle and are released upon cell lysis; VS, virogenic stroma. Reprinted with courtesy from Rohrmann (2013)[160].

The baculovirus infection can be divided into three phases:

- ❖ Early phase   (0–6 h post infection, *p.i.*),
- ❖ Late phase   (6–24 h *p.i.*)
- ❖ Very late phase   (18–24 up to 96 h *p.i.*)

Expression of the heterologous genes that are placed under the transcriptional control of the polyhedrin promotor initiates at the beginning of the very late phase of infection.[161]

The viral life cycle of *Ac*MNPV in nature starts with ingestion of OB into the gut of a host, such as an insect larva, followed by primary infection of cells of the gut lumen (**Figure 7**).

The release of ODV is followed by infection of epithelial cells in the midgut of the insect. Now, the infection becomes systemic with budding of virions towards the basallamina. BV are highly infectious for insect cell cultures and promote viral spread upon release during secondary infection.[162]
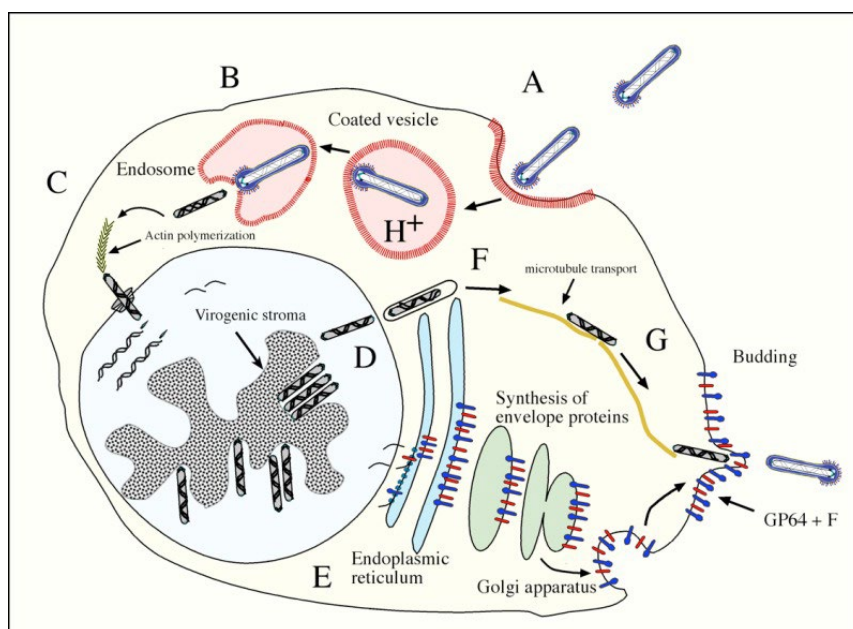
Figure 8 Budded virus infection of a BV Group I virus.
A BV of Group I attach to clathrin coated pits via GP64 receptors and are endocytosed. **B** Acidification of the endocytic vesicle causes a conformational switch of GP64. This induces the virion envelope and the endosomal membrane to fuse, thereby releasing the nucleocapsid into the cytoplasm. **C** The nucleocapsid passes a nuclear pore complex to enter the nucleus. **D** Transcription and DNA replicaion takes place in the nucleus. New nucleocapsids assemble in the virogenic stroma. **E** Viral envelope proteins that are synthesized and posttranslationally modificatied follow the secretory pathway to the plasma membrane. **F** Seperately, nucleocapsits upon their exit from the nucleus presumably are transiently coated. **G** BVs are transported along microtubules in association with the motor protein kinesin to the envelope-protein-rich plasma membrane and obtain envelopes through budding. Reprinted with courtesy from Rohrmann (2013).[160]

In the course of half an hour the virus enters the cytosol through endosomes via clathrin-mediated endocytosis, as depicted in **Figure 8**.[163] Viral DNA replication and production of virions occurs approximately 10 to 20 h *p.i.* in the nucleus, followed by viral release and commencing viral spread within the host organism/ cell culture. 24–48 h after the primary infection, large amounts of matrix proteins, namely polyhedrin and p10 are produced, forming OB that can protect the virion outside of a host (**Figure 7**). In the Bac-to-Bac system (Invitrogen) which is used in the workflow for intracellular protein crystallization described in this work, the baculoviral polyhedrin gene is knocked-out and replaced by the gene of interest, which is produced in large quantities under to the control of the strong polyhedron promotor.[164] Finally, dying cells release virus particles, ready to eventually infect a new host.[165]

1.4 Data collection from isolated intracellular crystals

The first indication that living cells could indeed be used as a crystallization chamber for recombinant proteins suitable for structure elucidation was provided in 2013. Redecke and others demonstrated, that a complete high resolution diffraction dataset can be collected using isolated (then-called '*in vivo*') cathepsin B crystals from rBV infected insect cells.[41,134] The known examples of 'unnatural' intracellular protein crystals all showed exceptional stability and robustness, even surviving harsh treatment like ultracentrifugation, vortexing and low percentages of detergents. In addition, there was the general assumption that the crowded environment within a cell would give a high background in any diffraction experiment if crystals were irradiated within the cell.[121] At the start of this doctoral project in 2015, it became clear that not all intracellular protein crystals show the remarkable stability of polyhedra, CatB or IMPDH. Both firefly luciferase and Xpa crystals grown in insect and mammalian cells, respectively, quickly dissolve after disruption of the cellular membrane.[71,101,125] This highlights that a change of the chemical environment associated with crystal isolation from the cells poses a strain to the crystals. Thus, also crystal diffraction quality might suffer, even in cases where the apparent morphology does not change upon crystal isolation. This strongly motivates the establishment of data collection strategies in viable cells, as established within this PhD thesis.

Currently, the determination of entirely new unbiased structures from *in cellulo* collected diffraction data is limited by the necessity of somehow retrieving the phase of the incident X-ray beam. Strategies to solve the phase problem are available for crystals in solution.

1.5 Phasing approaches

When X-ray diffraction data are collected from a crystal, the intensities of the diffracted waves are scattered from a series of planes, denoted by (h,k,l) in the crystal. From the intensities, amplitudes of the scattered waves may be derived. However, the phase information is lost during diffraction data recording; this is generally known as the crystallographic 'phase problem'.[166]

Multi-layer lenses that can focus hard X-rays are under development, but not routinely available yet.[167,168] Such lenses could allow direct imaging at high energies, as already possible with visible light (optical lenses), electrons (magnets) and also soft X-rays (Fresnel zone plates). However, computational imaging by Fourier transformation is possible. This requires both amplitudes and phases of the diffracted X-ray waves.[169] Thus, a starting model within the "radius of convergence for refinement"[169] needs to be identified by other means.

Molecular replacement (MR), and the experimental methods single-wavelength anomalous dispersion or diffraction (SAD) and multi-wavelength anomalous dispersion (MAD) are the most prominent methods to retrieve phases and are described here. Furthermore, there are isomorphous replacement methods: Single isomorphous replacement (SIR) or, if there is

more than one derivative crystal this is termed multiple isomorphous replacement (MIR). In isomorphous replacement methods one uses the scattering difference between atoms at specific known sites in the molecule and their substituents to gain definite phase information using a Harker diagram.[170]

Recently, a computational folding prediction revolution with AlphaFold has started, adding MR with a predicted new structure as a serious alternative to new experimental phase elucidation. This was made possible only by the extensive use of the PDB database by the structural biology community.[171,172]

Molecular replacement

The term 'molecular replacement` is used to describe the employment of a known molecular model to solve the unknown crystal structure of a related molecule or related domains in the crystal. It enables a first estimate of the phases of an unknown structure from that of a known homologue structure. By rotation and translation functions the starting model is placed within the crystallographic unit cell to best fit the measured diffraction amplitudes, followed by automatic chain tracing methods.[173–175]

By experience, given a sequence identity of ~ 40 %, molecular replacement is very likely to succeed. If a structure undergoes larger conformational changes, separate domains should be split up for model search. At a sequence identity 30-40 %, MR is usually still possible, but more difficult and challenges arise disproportional to the sequence identity. In minor cases, with the aid of programs like MR-Rosetta that use density- and energy guided structure optimization, MR may succeed with less than 30 % sequence identity.[176]

MR is the easiest way to retrieve phases of an unknown structure. However, since phases dominate electron density reconstruction, there may be a phase bias influencing the structure. For truly new structures *de novo* phasing becomes necessary.

De novo phasing

Single-wavelength anomalous diffraction (SAD) and multi-wavelength anomalous diffraction (MAD) are experimental phasing methods that exploit scattering resonances from diverse elements, which are strongest in heavy metals such as Fe, Se or Terbium.

The scattering power from anomalous centers differs significantly from all other atoms. This difference in diffraction is exploited in SAD/MAD to isolate and refine the position (in phase and amplitude) of a few anomalous scatterers (A) from the total scattering power. Once scattering components of (A) are known, from the phases of this substructure it becomes possible to derive the unknown phases of diffracted waves from the total structure, using an algorithm that was optimised by Hendrickson et al in the 1980s- 1990s[177] and *e.g.* a Patterson map from the anomalous centers.[169] This provides a phase probability distribution.[178]

Anomalous scattering is wavelength dependent. At higher energies > 7 keV (as used in experiments within this thesis), the anomalous scattering from O or P atoms is negligible. In the absence of (A), $|^{\lambda}F(\mathbf{h})|$ (the structure factor for each position $\mathbf{h}$) is independent of the wavelength. In practice this means that Friedel mates' structure factor values $|^{\lambda}F(\pm\mathbf{h})|$ are identical, as are their symmetry equivalents. Symmetry is broken in anomalous scattering.

In MAD, multiple datasets are collected from the same or equal crystals using different wavelengths, usually below the X-ray absorption edge, at the edge, and well away from the edge of the anomalous scattering species (http://skuld.bmsc.washington.edu/scatter/AS_periodic.html). The combined information is subsequently used to break the trigonometric phase ambiguity. For example, the K-edges of sulfur and selenium are at 2.472 keV and 12.6578 keV, respectively.[179] Presently there are practical difficulties for the use in crystallography for absorption edges below 6 keV, corresponding to X-ray wavelengths below 2 Å. Frequently, over 50 % of a protein crystal consists of solvent molecules that possess very different properties than the protein itself. Density modification techniques such as solvent flattening, solvent flipping, sphere of influence or histogram matching much improve the phase determination.

Heavy atom labeling of purified intracellular crystals was demonstrated by soaking experiments, co-crystallization and successful *de novo* structure determination severalfold.[51,73,78,104,111] However, intracellular labeling in living cells would require either incorporation of a selenium-containing amino acid into the protein that forms the crystal or binding of heavy metal ions at specific positions within the crystal lattice. This in turn requires a preceding passage of the metal ions over at least one cellular membrane.

2 Aim of study

Protein crystallography is the primary method that has enabled the successful determination of over 215,000 protein structures stored in the PDB, freely available to the public.[10] It is now possible to collect X-ray diffraction data in a serial manner from smaller protein crystals than previously possible, crystals in the very low μm or even high nm-range. For decades, crystallographers have concentrated on optimizing methods to grow protein crystals as large as possible. Now, SX approaches need a large number of small protein crystals to completely sample the reciprocal space and spread radiation damage. Some proteins are extremely challenging to crystallize with conventional methods, independent from their function and importance.[60,180] Thus, there is a large demand for alternate methods of protein crystallization.

Parameters that influence protein crystallization in insect cells remain largely unknown, and the number of yet undetected protein crystals using the expression system insect cells is presumed to be high. In the scope of this thesis, the detection of protein crystals in living cells should be improved by establishing a new high throughput XRPD/SAXS approach. The sensitivity and robustness of the method should be quantified and evaluated on the model systems *Tb*CatB, *Nc*HEX-1, *Photinus pyralis* luciferase and *Tb*IMPDH, using mock-infected and uninfected insect cells as controls.

Since the isolation of protein crystals from their cellular surrounding can impact the diffraction quality, applied data collection and mounting strategies for SX experiments should circumvent crystal isolation and purification. To protect the *in cellulo* crystals from environmental challenges, the delivery of crystals within intact insect cells to the X-ray beam should be established. Within this work, *in cellulo* data collection at cryogenic conditions should be investigated at DESY on Mitegen MicroMeshes and for additional screening purposes, at RT in CrystalDirect™ plates. Sample preparation and diffraction data collection should also be established and evaluated for SFX. As groundwork for *in cellulo* SFX using liquid jets, diffraction data collection on isolated intracellular crystals should be established for rBV *Nc*HEX-1 and rBV human Nct ECD SKL infected Sf9 cells. Establishing a FT-approach *in cellulo* with multi-patterned silicon chips should be done with *Nc*HEX-1 *in cellulo* crystals. Collected diffraction data should be processed with state-of-the-art SX processing software like CrystFEL.

Molecular replacement as an easy way to solve the phasing problem can only be applied if a protein with a related structure is available. Towards solving completely new structures with the *in cellulo* methodology, where crystals are surrounded by membrane(s), complicating soaking, this study aims to find an answer to the question: How could heavy metals be integrated to *in cellulo* crystals to enable direct phasing? This should be evaluated with two experimental approaches, one using the lanthanide terbium and LBT, the other SeMet incorporation during gene expression and crystal formation in insect cells.

SeMet incorporation should be quantified, and incorporation positions should be mapped using the mass difference between sulfur and selenium by MS.

All in all, this work aims to expand a workflow for *in cellulo* crystallography that increases the usability of the *in cellulo* approach for the larger scientific community of structural biologist.

3 Results

The established and applied *in cellulo* crystallization pipeline within the working group of Prof. Lars Redecke at the University of Lübeck consists of seven major steps:

1. Cloning the gene of interest into a vector that is compatible for bacmide generation;
2. Transfection of insect cells, generation and amplification of a potent virus stock by sequential infection of Sf9 insect cells;
3. rBV infection of High Five cells;
4. Detection of crystals within the living insect cells;
5. Diffraction data collection at a synchrotron or XFEL facility;
6. Data processing, merging and retrieving phases via MR;
7. Finally, similar to the workflow for single-crystal crystallography, model building and iterative structure refinement.

This work has focused on steps four to seven, ranging from the challenging detection of protein crystals *in cellulo*, serial data collection challenges using either 3rd generation synchrotron sources or XFEL radiation, processing the large amounts of resulting diffraction data with crystallographic software suites, to model building and finally structure refinement. The chapters of this cumulative dissertation consequently reflect the InCellCryst pipeline steps. The first part (chapter 3.1) presents the pipeline in its optimized version as published in *Nature Communications* (Schönherr, Boger and Lahey-Rudolph *et al*, 2024)[181]. Parameters that were found to influence crystal growth in living insect cells are elucidated, and additional protein targets that were channelled through InCellCryst are mentioned. Furthermore, data collection strategies for intracellular protein crystals using synchrotron radiation, directly measuring in the viable cell, together with processing results using both CrystFEL and XDS are presented. The next cumulative chapter 3.2 introduces detection of intracellular protein crystals via *in cellulo* powder diffraction coupled with small angle X-ray scattering, irradiated directly in the cell culture, published in *Journal of Applied Crystallography*[182].

Chapter 3.3 toggles first SFX experiments using *in cellulo* crystals, including the first solved structure collected with fixed-target SFX, published in *International Union of Crystallography journal*[183] and further proof-of-principle experiments with intracellular crystals and liquid jets.

The final monographic chapter 3.4 presents the successful incorporation of heavy(er) atoms into these membrane-surrounded crystals as a prerequisite to direct phasing methods.

3.1 A streamlined approach to structure elucidation using *in cellulo* crystallized recombinant proteins, InCellCryst.

A streamlined approach to structure elucidation using in cellulo crystallized recombinant proteins, InCellCryst

Robert Schönherr[1,9], Juliane Boger [1,9], J. Mia Lahey-Rudolph [1,2,8,9], Mareike Harms[1], Jacqueline Kaiser [1], Sophie Nachtschatt [1], Marla Wobbe[1], Rainer Duden[3], Peter König[4,5], Gleb Bourenkov [6], Thomas R. Schneider [6] & Lars Redecke [1,7] ✉

With the advent of serial X-ray crystallography on microfocus beamlines at free-electron laser and synchrotron facilities, the demand for protein microcrystals has significantly risen in recent years. However, by in vitro crystallization extensive efforts are usually required to purify proteins and produce sufficiently homogeneous microcrystals. Here, we present InCellCryst, an advanced pipeline for producing homogeneous microcrystals directly within living insect cells. Our baculovirus-based cloning system enables the production of crystals from completely native proteins as well as the screening of different cellular compartments to maximize chances for protein crystallization. By optimizing cloning procedures, recombinant virus production, crystallization and crystal detection, X-ray diffraction data can be collected 24 days after the start of target gene cloning. Furthermore, improved strategies for serial synchrotron diffraction data collection directly from crystals within living cells abolish the need to purify the recombinant protein or the associated microcrystals.

The crystallization of recombinant proteins directly within the producing cell is considered to be a rare process of structured protein assembly. It was initially reported by Fan et al. in 1996, who showed the crystallization of a heterodimer of calcineurin in insect cells following coinfection with two recombinant baculoviruses[1] (BVs). More than a decade later the crystallization of the C-terminal domain of the avian reoviral μNS fused to EGFP in CEF cells as well as of a human IgG in CHO cells was reported[2,3].

Due to the small volume of intracellular crystals that was suggested at the time to be limited by the dimensions of the producing cell[4], they harbor low diffraction capabilities. Only with the implementation of microfocus beamlines at third and fourth generation synchrotrons and the commissioning of X-ray free-electron lasers, combined with the development of novel serial data collection strategies[5,6], intracellular crystals become a target for structural analysis. In 2007, the first structure of a recombinantly produced but natively crystallizing protein, cypoviral polyhedrin, was published[6], followed in 2013 by the structure of trypanosomal cathepsin B (CatB), a protein that does not natively form intracellular crystals[7]. Today, a significant number of intracellularly grown protein microcrystals have been discovered, both in native environments and as a consequence of recombinant protein production in host cells[8,9]. Some of them have

[1]Institute of Biochemistry, University of Lübeck, Lübeck, Germany. [2]Center for Free-Electron Laser Science (CFEL), Hamburg, Germany. [3]Institute of Biology, University of Lübeck, Lübeck, Germany. [4]Institute of Anatomy, University of Lübeck, Lübeck, Germany. [5]Airway Research Center North (ARCN), University of Lübeck, German Center for Lung Research (DZL), Lübeck, Germany. [6]European Molecular Biology Laboratory, Hamburg Unit c/o Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. [7]Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. [8]Present address: X-ray technology lab, TH Lübeck - University of Applied Sciences Lübeck, Lübeck, Germany. [9]These authors contributed equally: Robert Schönherr, Juliane Boger, J. Mia Lahey-Rudolph. ✉e-mail: redecke@biochem.uni-luebeck.de

successfully been used to elucidate protein structures[6,7,10–27]. However, due to the limited available literature and a high threshold for starting a new and largely unexplored method, the intracellular protein crystallization approach is still used by only a small fraction of structural biologists. To better exploit the cellular crystallization capabilities at quasi-native conditions and to broaden the user community, a streamlined approach for the generation and detection of intracellular crystals, as well as for their application in protein structure elucidation is needed.

In this context, two initial approaches have been published in recent years. Boudes et al.[15] proposed in 2016 a simple pipeline employing Sf9 insect cells and a baculovirus expression vector system (BEVS) for recombinant protein production. The infected cells are screened for intracellular crystals by bright-field microscopy, followed by enrichment of crystal-carrying cells using flow cytometry and try-pan blue staining for improved visibility at the beamline. After pipetting the cells onto a mesh grid support and flash-cooling in liquid nitrogen without cryoprotection, samples are mounted at a micro-focus synchrotron beamline. Single cells are sequentially centered in the X-ray beam to collect partial diffraction datasets.

Tang et al.[28] extended this approach in 2020 by designing a gateway-compatible baculovirus expression vector library for high-throughput gene expression in insect cells. Large existing Gateway clone libraries should allow the rapid and cost-effective construction of expression clones for mass parallel protein production, while their plasmid collection also supports the attachment of several fusion tags for different research applications. This pipeline further includes advanced SONICC[29] and TEM[30] techniques to screen for microcrystal formation within the infected cells. However, no strategies for diffraction data collection have been proposed.

To address this issue and improve the existing pipelines we present an advanced approach for protein structure elucidation from crystals growing in living insect cells, denoted as InCellCryst (Fig. 1). InCellCryst combines the high-throughput approach of Tang et al.[28] with improved enrichment protocols for crystal containing cells, as first suggested by Boudes et al.[15], and improved X-ray diffraction collection strategies. For this we chose soluble proteins known to crystallize in insect cells with native localizations in different cellular compartments[7,17,20,31,32]. Most significant hallmarks of InCellCryst extending the previous approaches include (i) a broadly applicable and highly versatile cloning system; (ii) the possibility to direct the target protein into different cellular compartments and thus different environmental conditions, enabling a systematic screening for optimal intracellular crystal growth comparable to in vitro crystallization screenings; (iii) the establishment of X-ray diffraction data collection strategies directly in viable insect cells, applying our previously established fixed-target approach[25] that allows serial helical line scans using high frame-rate detectors and synchrotron radiation; and (iv) the application of state-of-the-art data processing software on the collected serial diffraction data. Furthermore, by reducing cloning-dependent artificial amino acids and by rendering the isolation of the crystals from the cells unnecessary for diffraction data collection, the pipeline has a minimal impact on the native protein structure. InCellCryst opens a quick and easy route to efficiently exploit the crystallization capability of living cells for structural biology.

Results

Simple and versatile cloning systems for intracellular crystallization screening of target proteins

Crystallization of recombinant proteins reportedly occurs in several insect cell compartments. For example, calcineurin, EGFP-µNS, and inosine-5′-monophosphate dehydrogenase (IMPDH) crystallize within the cytoplasm of infected insect cells[1,20,31,32]. In contrast, firefly luciferase forms crystals within peroxisomes[31], and CatB crystals are observed within the endoplasmic reticulum[33].

Inspired by these results, we aimed to exploit protein crystallization in different cellular compartments as a screening parameter, that could be comparable to buffer variations in conventional crystallization screenings. To that end, we developed a 1st generation and further optimized 2nd generation cloning system for target genes based on the Bac-to-Bac system (Invitrogen) and its pFastBac1 plasmid (pFB1), where different cellular localization sequences and fusion tags, together with start- and stop-codons, are encoded on the modified plasmid (Supplementary Table 1). To keep the cloning simple and efficient, while minimizing the impact on the protein structure, we used a ligation-based approach that minimizes artificial amino acids added to the recombinant protein. In the 1st generation of the system, target gene cloning is achieved using blunt-end ligation of the PCR-amplified gene into an EheI restriction site integrated between sequences coding for different N- and C-terminal localization sequences. The 2nd generation cloning system (denoted as v2) uses cohesive ends produced by restriction enzymes KpnI and NheI for ligation, significantly improving the cloning efficiency, at the expense of two artificial residues at the N- and C-terminus of the target protein. In addition to localization sequences, several fusion tags were introduced into the cloning systems to extend the range of downstream applications, enabling protein localization as well as purification (Supplementary Table 1).

### Optimized procedures for recombinant baculovirus generation

After cloning, parts of the recombinant pFastBac1 plasmids are transposed into the baculoviral genome using Tn7 transposition in E. coli DH10EmBacY cells[34]. This strain carries a bacmid encoding an EYFP reporter gene, allowing for a quick and easy fluorescence-based evaluation of cell transfection, viral infections, and gene expression. Moreover, the deleted viral cathepsin and chitinase genes in the BV genome reduce target protein degradation and cell lysis[34].

To produce high-titer stocks of the recombinant BVs, it proved most efficient to isolate the recombinant bacmid using the ZR Bac DNA Miniprep Kit (Zymo Research), followed by transfection of Sf9 insect cells using the ESCORT IV transfection reagent (Merck) according to an optimized procedure. A sufficiently high viral titer can be produced by Sf9 cells in two steps: after 5 days of incubation the transfection supernatant is used to infect another 5 mL suspension culture of Sf9 cells. After 4 days of incubation at continuous shaking, the supernatant can be harvested and titrated using an endpoint dilution assay on High Five cells. Using the EYFP fluorescence as a marker allows for a sensitive detection of single infected cells. Two different insect cell lines are used, since High Five cells are much more susceptible to baculoviral infection than Sf9 cells[35] and usually indicate a viral titer that is between one and two orders of magnitude higher than that detected on Sf9 cells. Sf9 cells, on the other hand, produce orders of magnitude more viral particles and thus serve as versatile virus producing cells (Fig. 2).

Optimization of intracellular crystallization based on MOI, infection time, and insect cell lines used

As initially shown by Fan et al. for intracellular calcineurin crystals[1], crystallization efficiency, and crystal sizes may depend on the insect cell line used for crystal production. High Five cells produced larger calcineurin crystals at a higher rate than Sf9 or Sf21 cells. To test whether this is a general phenomenon or dependent on the specific target protein, we produced several proteins in Sf9 and High Five cells. Although the infection rate at an MOI of 1 was above 95% in both cell lines, two to ten times more crystal-containing cells were observed in High Five cell cultures (Fig. 3a). Furthermore, High Five cells consistently produced crystals with a 2.5 to 7 times larger volume compared to that in Sf9 cells (Fig. 3b), confirming previous observations by Fan et al.[1].

Almost all cells within a culture should be simultaneously infected for optimal crystal production. To achieve this, an MOI of 1 is sufficient, based on our optimized virus titration assay (Fig. 3c).
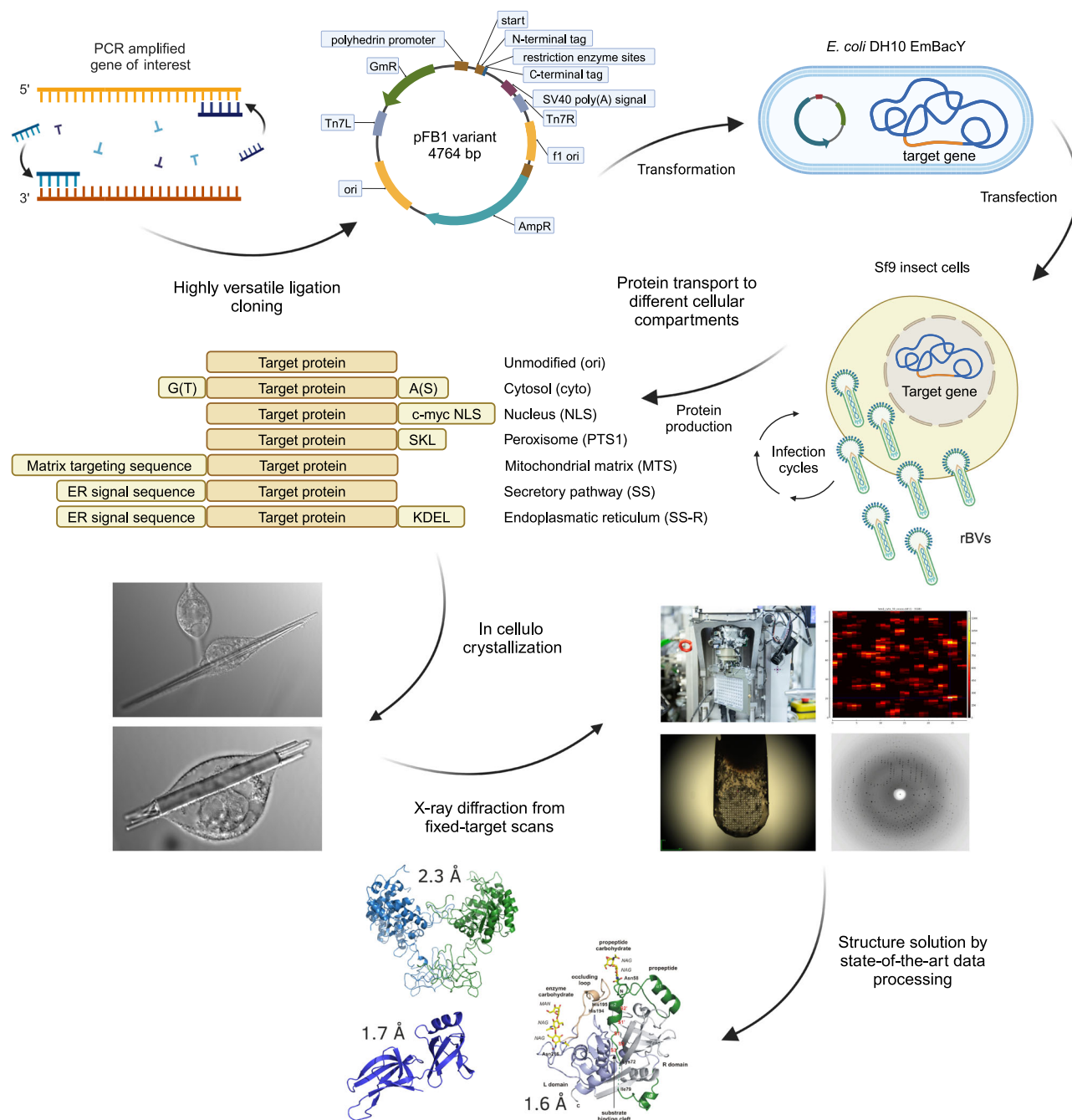
Fig. 1 | The InCellCryst pipeline. The gene of interest is amplified by PCR and ligated into modified pFastBac1 plasmids. After transformation of *E. coli* DH10EmBacY cells, recombination with the bacmid takes place. The recombinant bacmid is isolated and Sf9 insect cells are transfected for BV generation. After high titer viral stock production, High Five insect cells are infected and used for high yield target gene expression. This eventually leads to the crystallization of the target protein within one of the cellular compartments, depending on the transport signaling tag fused to the target protein sequence. Crystal-containing cells are directly used for serial diffraction data collection at RT or 100 K at a synchrotron source or an XFEL. Serial diffraction data is finally processed to elucidate the structure of the target protein. rBVs recombinant baculoviruses.

The BEVS works on a transient basis. After infection, the insect cells replicate the recombinant virus and the target gene expression is started about 18 hours post infection (hpi). A few days later, infection associated cytopathic effects lead to cell death[36]. This implies that the time point after virus infection is crucial to harvest the maximum fraction of living and crystal-containing cells, optimal for X-ray data collection. Thus, we analyzed the crystallization of several target proteins over a time course of 9 days using light microscopy (Fig. 3d). First intracellular crystals were detected at the earliest 36 hpi and at the latest 72 hpi, depending on the target protein, while the highest fraction of crystal-containing cells was obtained between 72 and 96 hpi. Afterwards, the crystal-containing cell fraction consistently decreased. The comparatively fast decline in cells producing EGFP-μNS might be attributed to the use of the Bac-to-Bac bacmid (Invitrogen), compared to the EmBacY bacmid[34] used to generate all other recombinant BVs. As mentioned above, deletion of the viral cathepsin and chitinase genes reduced target protein degradation and cell lysis, improving the stability of the intracellular crystals.

As conclusion, X-ray diffraction data collection should optimally be performed 96 hpi, independent from the crystallizing target
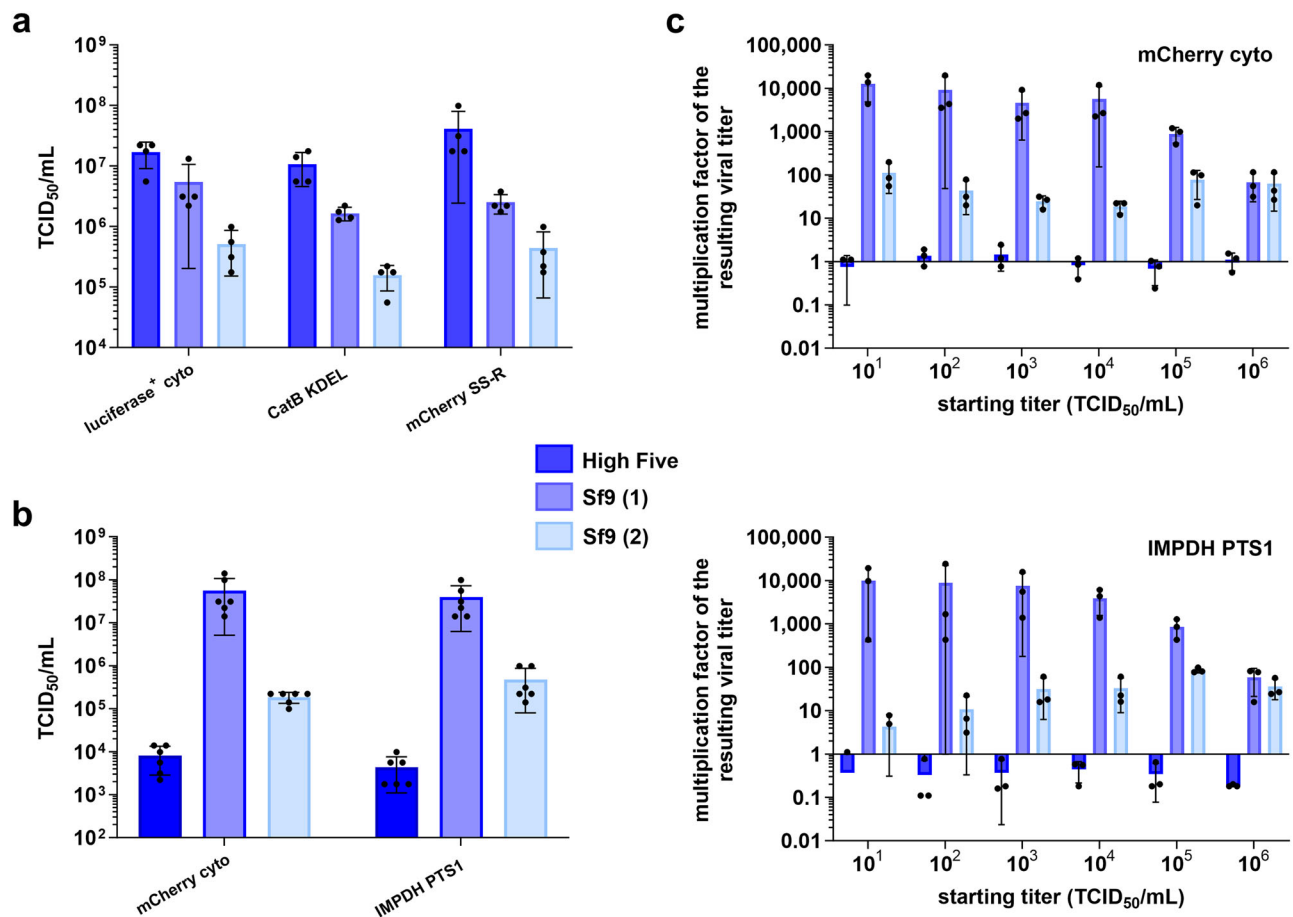
Fig. 2 | Optimization of virus stock production. a Comparison of the susceptibility of different insect cell lines for recombinant baculovirus (rBV) infection. Three different rBVs were titrated on High Five cells as well as on two Sf9 cell lines obtained from different sources. The averaged $TCID_{50}$-values of four independent experiments are presented as mean values ± SD. High Five cells exhibit the highest apparent titer due to their increased susceptibility. **b** Comparison of virus production of High Five and Sf9 cell lines. Two different rBVs (with/ without crystal production capability) were amplified on the denoted cell lines in three independent experiments and titrated twice on High Five cells. For the infection, a titer of $1 \times 10^4 \, ml^{-1}$ was used. The resulting virus stock was harvested 4 days after infection of 0.45×10^6 cells in a 12-well plate. The averaged $TCID_{50}$-values are presented as mean values±SD. **c** Comparison of virus productions depending on the initial infection titer. Experiment design as described in **b**. Infection titers varied between 1×10^1 and $1 \times 10^6 \, ml^{-1}$. The differences between infection and harvesting titers are shown as amplification factors. High Five cells do not produce a noticeable amount of new infectious viral particles, while Sf9 cells are shown to be highly productive. Different clones of the same cell line can exhibit considerable differences in their virus production capabilities. Data of three independent experiments are presented as mean values ± SD.

protein. Intracellular crystal production can be linearly upscaled to increase the yield of crystals of the same size-range and quality, a stark benefit for serial X-ray diffraction data collection. However, this is limited by the need for semi-adherent cells for optimal crystal production.

**Efficient detection of intracellular protein crystals**

A bottleneck for any intracellular crystallization pipeline is the question whether the target protein is crystallizing within the cell. Proof of crystallinity can only be established by diffraction of X-rays or by visualization of the crystal lattice. However, the generation of a second harmonic (SHG) signal[28] and the detection of regular forms and straight edges within the chaotic cellular environment can serve as strong indicators for crystal growth. Ordered structures even in the sub-micrometer size range (down to about 500 nm in diameter) can be most conveniently screened in a cell using a high NA objective in combination with differential interference contrast (DIC) (Fig. 4a). With fluorescent microscopes that are widely available, immunofluorescent staining of the target protein within the infected cells using a specific antibody represents another possibility to detect regular arrangements (Fig. 4b). If the target protein itself is tagged with a

fluorescent protein, crystal-like structures can be readily detected using standard or confocal fluorescence microscopy (Fig. 4b, c).

Intracellular crystals can grow in at least one dimension to the micrometer size-range, exceeding the diameter of the cell body several fold. This is mainly limited by the protein production capability of the individual cell and the protein half-life in the living system. Since a high local protein concentration is required to obtain the conditions for crystal nucleation and growth, the size of individual crystals depends on how much correctly folded protein can be produced to balance the crystal growth and protein degradation rates. However, they can also occur as nanocrystals with edges of less than 100 nm in length. A well-established technique for the detection of such tiny crystals is TEM (Fig. 4d–g). The detection of crystalline structures in thin sections, stained with standard heavy metal contrasting techniques, is accomplished by the characteristic high protein density compared to the surrounding material of retained soluble proteins (Fig. 4d). Visualization of the crystal lattice can confirm the crystallinity of the detected structures at the same time (Fig. 4e–g).

To verify the crystallinity of structures detected within cells by X-ray diffraction, we have established two different approaches. X-ray powder diffraction (XRPD) of a concentrated pellet of living cells
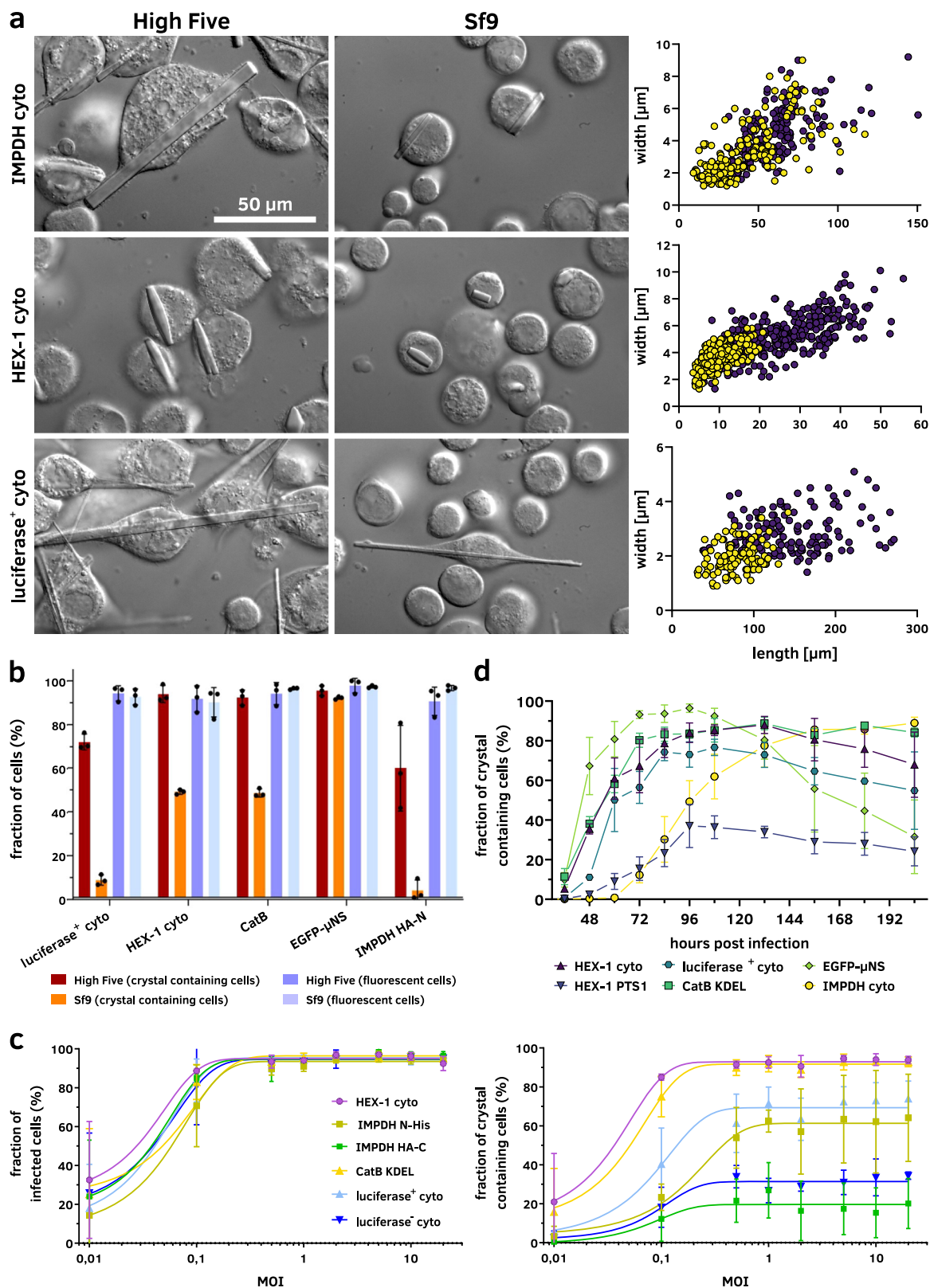
within a 3.5 mm capillary at a microfocus synchrotron beamline results in visible Debye-Scherrer rings with a resolution up to that of the water ring (approximately 3–4 Å), if long exposure of the sample is combined with a helical line scan and a collimated beam (Fig. 4h, i). By combining XRPD with small angle X-ray scattering (SAXS) we previously developed a far more sensitive approach for the detection and analysis of intracellular protein crystals[37]. The specific Bragg diffraction

detectable as peaks in radially averaged 1D plots of the SAXS scattering signal of the cells provides a characteristic fingerprint of the intracellular crystals, corresponding to partial Debye-Scherrer rings that contain information on the unit cell parameters of the detected crystals. Here, we extended this method to evaluate the impact of different tags and localization sequences on the structure of crystals from the same target protein. For HEX-1 variants, large differences between the

Fig. 3 | Crystallization capabilities of insect cell lines. a Visual comparison of crystals in Sf9 and High Five cells, illustrating the size differences between crystals of the same protein produced in the two cell lines. Representative images of three independent experiments are shown. High Five cells produce significantly larger crystals than Sf9 cells while the morphology remains identical. Measurements of several hundred crystals in the right-side panels illustrate the large size differences. Yellow dots, Sf9 cells; violet dots, High Five cells. **b** High Five cells also show improved crystallization capabilities compared to Sf9 cells characterized by a higher fraction of crystal-containing cells within an infected culture, while no differences are visible regarding the fraction of infected, EYFP-producing cells. **c** An

MOI of 0.5 is sufficient for infection of and protein production in nearly all cells within a High Five cell culture. However, a MOI of 1 should be used to maximize protein crystallization within the cells. A higher MOI does not improve in cellulo crystallization, independent of the crystallizing protein. **d** Visible crystal growth in High Five cells is detectable from 36 h after infection with an MOI of 1 onwards. Growth characteristics are variable but follow a comparable scheme with the maximal fraction of crystal-containing cells obtained between 72 and 96 hpi. The only exception is IMPDH cyto, showing a continuous increase in crystal containing cells. **b**–**d** Data of three independent experiments are presented as mean values ± SD.

specific fingerprints are visible (Fig. 4j), indicating different crystal morphologies, while highly comparable fingerprints have been obtained for all tested IMPDH variants (Fig. 4k). The detectable peak intensity correlates with the diffractive volume in the capillary; thus, the observed peaks are weak for mitochondrial targeted and ER localized HEX-1 protein (Fig. 4j).

Impact of target protein modifications on intracellular crystallization

To evaluate the cellular compartment as a crystallization screening parameter, we cloned the genes encoding IMPDH, CatB, HEX-1, luciferase, and EGFP-µNS into different pFB1 screening vectors encoding specific cellular translocation signals. The environmental conditions and properties of the individual compartments modulated the shape, size, and order of the crystals, as well as the fraction of crystal-containing cells (Fig. 5). In other cases, crystallization was fully compartment dependent. The most significant differences have been observed between the secretory pathway and the other cellular compartments, likely caused by incorrect glycosylation of the target proteins in the ER/Golgi. CatB did not crystallize any more when retained in the cytosol, while IMPDH, EGFP-µNS, and luciferase showed no indications for crystallization when co-translated into the ER. HEX-1 is the only protein observed so far that forms crystals in all cellular organelles tested.

IMPDH crystallized with or without a PTS1 signal (peroxisomal import) in the cytosol, forming thick needles with a square base, similar to the unmodified version of the protein (Fig. 5a). The morphology of these crystals is comparable to that of the His-tagged version that was initially found to crystallize in insect cells[20]. The unit cell parameters of all IMPDH crystals remained unchanged as verified by SAXS-XRPD (Fig. 3k). CatB only crystallized in the ER. However, if a C-terminal retention signal is fused to the protein, crystals started to grow earlier and to a larger diameter. Luciferase only crystallized after cytosolic protein production and the inactivation of its native PTS1 signal resulted in a strong increase in crystallization efficiency, while the crystal morphology was not affected. HEX-1 crystallized in all compartments tested but showed different crystal shapes depending on the fused intracellular translocation tags. Even unmodified HEX-1 formed crystals of two different shapes. A predominant hexagonal, block-like morphology that is comparable to, although much larger than, the woronin bodies found in the native fungus[38], as well as a bipyramidal shape that was not reported to occur natively. The disruption of the native PTS1 signal by cloning into the pFB1 cyto and cyto v2 vectors transformed the hexagonal crystal blocks into a spindle-like shape. Fusion of the nuclear localization signal leads to a predominant fraction of bipyramidal crystals with only few spindle-like crystals present. Addition of a PTS1 signal, as well as fusion of the KDEL retention signal, exclusively resulted in bipyramidal HEX-1 crystals, while the translocation into the ER produced mostly spindle-like crystals that were considerably smaller than the cytosolic crystals. The smallest HEX-1 crystals grew when targeting signals to the mitochondrial matrix were fused (Fig. 5b). Although the mentioned HEX-1 constructs showed only three different crystal morphologies, at least four different fingerprints are obtained by SAXS-XRPD analysis (Fig. 3j),

characterized by the scattering curves of HEX-1 ori, cyto, HA-C and NLS v2, respectively. EGFP-µNS did not crystallize in the secretory pathway, but in all other compartments tested (Fig. 5c). Compared to the unmodified fusion protein, morphological differences are mostly confined to the crystal size. All EGFP-µNS crystals exhibit a needle-like morphology with a hexagonal cross-section. Since very weak diffraction at low resolution was previously recorded using cytosolic EGFP-µNS crystals[31,32], we additionally tested the intracellular crystallization of both parts of the fusion protein alone, as well as the fusion of µNS with other fluorescent proteins (mCherry and mScarlet-I). However, none of those constructs showed any indication of intracellular crystallization (Supplementary Fig. 1).

Cell sorting enriches crystal containing cells in a culture

Since some proteins crystallize with low efficiency, it is crucial to enrich crystal containing cells within a culture to ensure the efficient use of X-ray beamlines. Boudes et al. proposed to use the side scatter channel of a cell sorter to directly select crystal-containing cells. When compared to uninfected cells, high side scatter values correlate with polyhedrin crystal-containing cells[15]. To implement this approach, we tested forward scatter (FSC) and side scatter (SSC) channels for their reliability to select cells containing crystals of different target proteins. Indeed, a distinct difference in the side scattering behavior of baculovirus-infected cells compared to uninfected cells is visible. However, comparing infected cells producing a soluble protein with infected cells containing protein crystals does not show selectable differences in either the forward or the side scattering signal of the cells, even for cells harboring very large crystals (Fig. 6a). Further, there is no selectable population visible in either channel when plotted against the target protein level as correlated by EYFP fluorescence (Fig. 6b). Therefore, we investigated the EYFP production itself as a parameter to reliably select crystal-containing cells. Since both EYFP- and target genes are controlled by the polyhedrin promotor we hypothesized that a higher EYFP production would correlate with a higher production of the target protein and therefore result in a higher crystallization probability. Since there is a strong positive correlation of both protein production levels (Fig. 6c), we sorted cell cultures containing different target protein crystals based on the EYFP-fluorescence level and evaluated the fraction of crystal-containing cells in each population. This strategy successfully leads to an enrichment of crystal-containing cells, depending on the EYFP-fluorescence, with the enrichment factor primarily depending on the EYFP-gate settings (Fig. 6d).

Collection of X-ray diffraction data from crystals in viable insect cells

While some in cellulo grown crystals, such as CatB or IMPDH, were shown to be very robust and can readily be isolated from the insect cells for diffraction data collection[7,12], others such as luciferase grown in insect cells or XPA grown in mammalian cells suffer from the environmental changes associated with cell damage[12,31]. Collecting X-ray diffraction data from crystals directly within the living cells can mitigate those problems. We previously established in cellulo data collection strategies at room temperature (RT) using silicon chips at XFELs and synchrotron sources[17,39]. Here, we extend the methodology
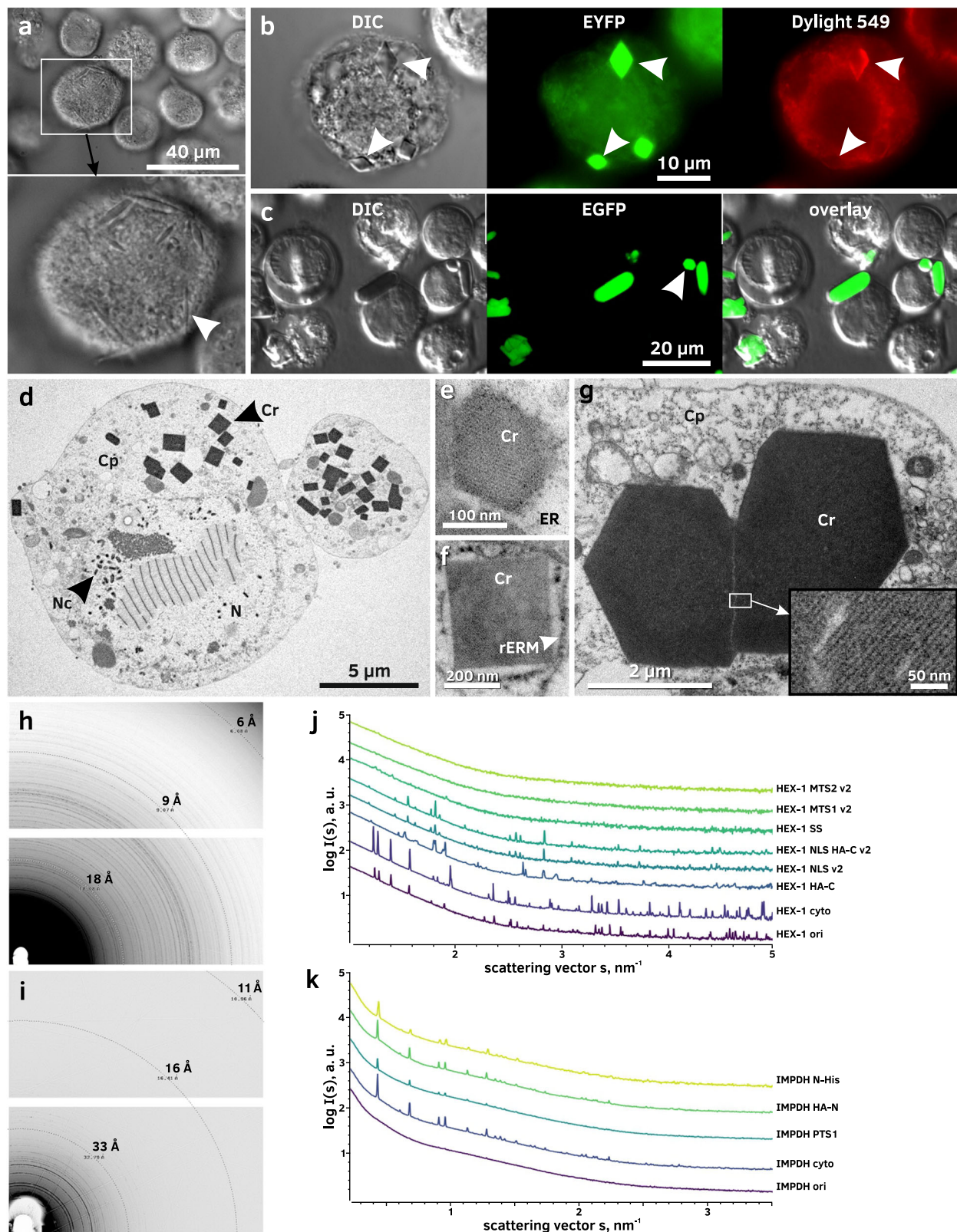
to serial cryo-crystallography using MicroMesh mounts (Fig. 7a) and serial in situ data collection on CrystalDirect™ plates[40] at RT (Fig. 7b) at a synchrotron source. The latter can be particularly advantageous for highly sensitive crystals that are already damaged when cells are transferred to the sample holder.

For data collection from intact crystal-containing cells at 100 K, we carefully transferred the cells onto a MicroMesh mounted on a light microscope stage. Cells were then covered with 40 % PEG200 diluted in cell culture medium and manually frozen in liquid nitrogen. This procedure takes less than 30 s, from adding the cryoprotectant to freezing. Together with keeping the cells in a 90% humidity environment, this ensures the integrity of the cells, allowing diffraction data collection from crystals within their growth environment.

Fig. 4 | Methods for detection and analysis of intracellular protein crystals.
a Imaging of HEX-1 SS crystals (white arrow) in High Five cells using differential interference contrast (DIC). **b** Immunofluorescence labeling of HA-tagged HEX-1 HA-C crystals (white arrows) in High Five cells with a DyLight 549-conjugated antibody. **c** Confocal fluorescence imaging of EGFP-μNS in Sf9 cells. **d** TEM of High Five cells producing luciferase⁺ cyto. Crystals are visible in high contrast due to their comparatively high protein density. **e** TEM of nanocrystals of a baculoviral protein located within the ER of a Sf9 cell showing a fine crystal lattice grating. **f** TEM of a CatB crystal surrounded by a ribosome studded membrane (rER) in a Sf9 cell. **g** TEM of EGFP-μNS crystals in High Five cells showing defects in the crystal lattice. **h, i** Powder diffraction images of High Five cells containing crystals of HEX-1 cyto **h** or IMPDH HA-N **i**. For IMPDH background subtraction was done using *adxv* to enhance the visibility of Debye-Scherrer rings. **j, k** SAXS curves of crystal-containing High Five cell suspensions. Peaks arise from incomplete Debye-Scherrer rings on the detector images. Graphs correspond to cells producing HEX-1 **j** and IMPDH **k** proteins fused to different tags and localization sequences. In contrast to IMPDH, crystals of HEX-1 variants give diverse fingerprints, implying differences in the unit cell parameters. Cr protein crystal, Cp cytoplasm, ER endoplasmic reticulum, N nucleus, Nc nucleocapsid, rERM membrane of the rough ER. Representative micrographs of three independent experiments are shown.

At RT we collected data from HEX-1 cyto crystals grown directly on CrystalDirect™ plates. Since the plate is mounted in a vertical position at the goniometer, the hydrophobic crystallization foil was coated with poly-D-lysine and the cell culture medium was supplemented with 25% FBS to achieve proper cell adhesion. Prior to mounting, the supernatant cell culture medium was removed from the wells to minimize background scattering. To prevent the sample from drying, the lid was replaced by another crystallization foil. During 6 hours of data collection the sample did not show any reduction in diffraction power.

Serial diffraction datasets of the crystals contained in viable cells was performed using a helical grid scan approach[25,41], available at the EMBL P14 beamline located at the PETRA III storage ring (DESY, Hamburg, Germany).

In cellulo diffraction data collection leads to improved biological data

By collecting X-ray diffraction data from crystals directly diffracted in viable High Five insect cells, we were able to solve the structures of HEX-1 and IMPDH variants using the software suite *CrystFEL*[42]. HEX-1 ori and IMPDH ori structures were additionally elucidated using *XDS*[43] on small rotational datasets identified from the same raw data applying custom-made scripts (Supplementary Data 2 and 3). Details on data collection and structure refinement are summarized in Supplementary Table 2.

As mentioned above, HEX-1 crystals showed different morphologies inside the cytoplasm or nucleus. Whereas HEX-1 cyto and cyto v2 both formed spindle-like as well as bipyramidal crystals, HEX-1 ori crystals exhibit a hexagonal block shape (Fig. 5b). For HEX-1 ori and HEX-1 cyto crystals diffracted at 100 K one distinct unit cell population was found, whereas two unit cell populations were identified for HEX-1 cyto v2. When diffracted at RT another unit cell population of HEX-1 cyto was detected, differing in the c-axis. The unit cell parameters of HEX-1 cyto collected at RT are enlarged compared to those diffracted at 100 K, however, comparable to recently published structures of HEX-1 cyto, also recorded at RT[17,39]. All HEX-1 structures (Fig. 7c, Supplementary Fig. 2) show an increased flexibility between Phe145 and Ser151, the C-terminal loop region after the α-helix. However, when superimposed to the previously reported SFX HEX-1 structure (PDB 7ASX), only minor deviations were observed. A maximum RMSD of ~ 2 Å was calculated for C_{alpha} atoms in the loop between Ser61 and Gln66.

The IMPDH ori and cyto structures (Fig. 7d) were determined by molecular replacement using the coordinates of the human IMPDH isoform 1 A-chain monomer (PDB 1JCN). All elucidated IMPDH variants only show minor variations in flexible parts like the Cys-loop and the C-terminal region (mean RMSD 0.37 Å, max RMSD ~ 7 Å) compared to the C_α atoms of the previously published IMPDH N-His structure (PDB 6RFU), which corresponds to a closed inhibited IMPDH conformation. All IMPDH structures show natural ligands at the canonical binding sites 1 and 2 in the Bateman domain. However, instead of guanosine monophosphate (GMP) found in the IMPDH N-His structure solved from isolated crystals[20], the structures solved in this study allow a clear identification of GDP in the canonical binding site 2, stabilized by

hydrogen bonds to Lys133 and Arg101 that coordinate the α-phosphate and β-phosphate of GDP (Fig. 7e, Supplementary Fig. 3). In addition, a phosphate is bound to the IMP binding site of the catalytic domain of the IMPDH structures.

As observed in the previously published IMPDH structure solved by SFX[20], the canonical binding site 1 incorporates an ATP molecule. However, we observed two conformations of the bound ATP, where the gamma phosphates of the adjacent ATP molecules are either opposing or facing each other (Fig. 7f). The base and ribose are stabilized by interactions with Thr180, Thr174 or Asp158 and His200, whereas the different phosphate groups are interacting mainly with Thr156, Lys157, Arg219, Gly201, solvent molecules and the main chain. A detailed view of all interactions is depicted in Supplementary Fig. 3.

Based on these results, in cellulo diffraction data collection clearly improves the bioinformation on natively bound ligands and their conformation compared to the diffraction of isolated in cellulo or conventionally grown crystals.

Discussion

In this study, we present InCellCryst, an advanced pipeline for cloning and recombinant gene expression in insect cells that allows, at best, to collect serial X-ray diffraction data for structure elucidation of the crystallized protein within a month. Our pipeline is based on insect cells for their performance as an efficient crystal production facility[8], able to grow sufficiently large crystals for synchrotron diffraction data collection[25]. Insect cells can be grown in suspension and as semi-adherent cultures with seamless switching between both, and efficient gene shuttles are available in the form of modifiable BEVSs producing a massive amount of the target protein nearly synchronized throughout the culture[36], without significant biological safety hazards (BSL 1).

We have demonstrated the pipeline on five different proteins (IMPDH, HEX-1, Cathepsin B, Luciferase and EGFP-μNS), representing a selection of soluble proteins of different source organisms that crystallize in insect cells in their native compartment or, in case of peroxisomes, readily in the cytoplasm or nucleus of the cells. Additionally, even artificial fusion proteins like EGFP-μNS can crystallize in different compartments of the insect cells. Importantly, as shown for IMPDH and HEX-1, InCellCryst can produce crystals from native protein sequences, without purification tags or even remnants of proteolytically removed tags, within a quasi-native cellular environment. This leads to crystal structures that are as close to the native state of the protein as possible.

We established cellular compartments as suitable screening parameters that provide differing chemical environments to maximize the chance for crystallization of the target protein. Compartment screening could also lead to a diversification of produced crystal morphologies and sizes that can be used in different downstream applications, e.g. phasing approaches using diffraction data from multiple crystal forms[44]. It has to be emphasized that the established mitochondrial matrix targeting signals successfully translocate a target protein into the mitochondrial matrix, which has not been described so far.

To ensure an efficient use of beam time at X-ray facilities, crystal density in the culture is maximized by using High Five cells for crystal
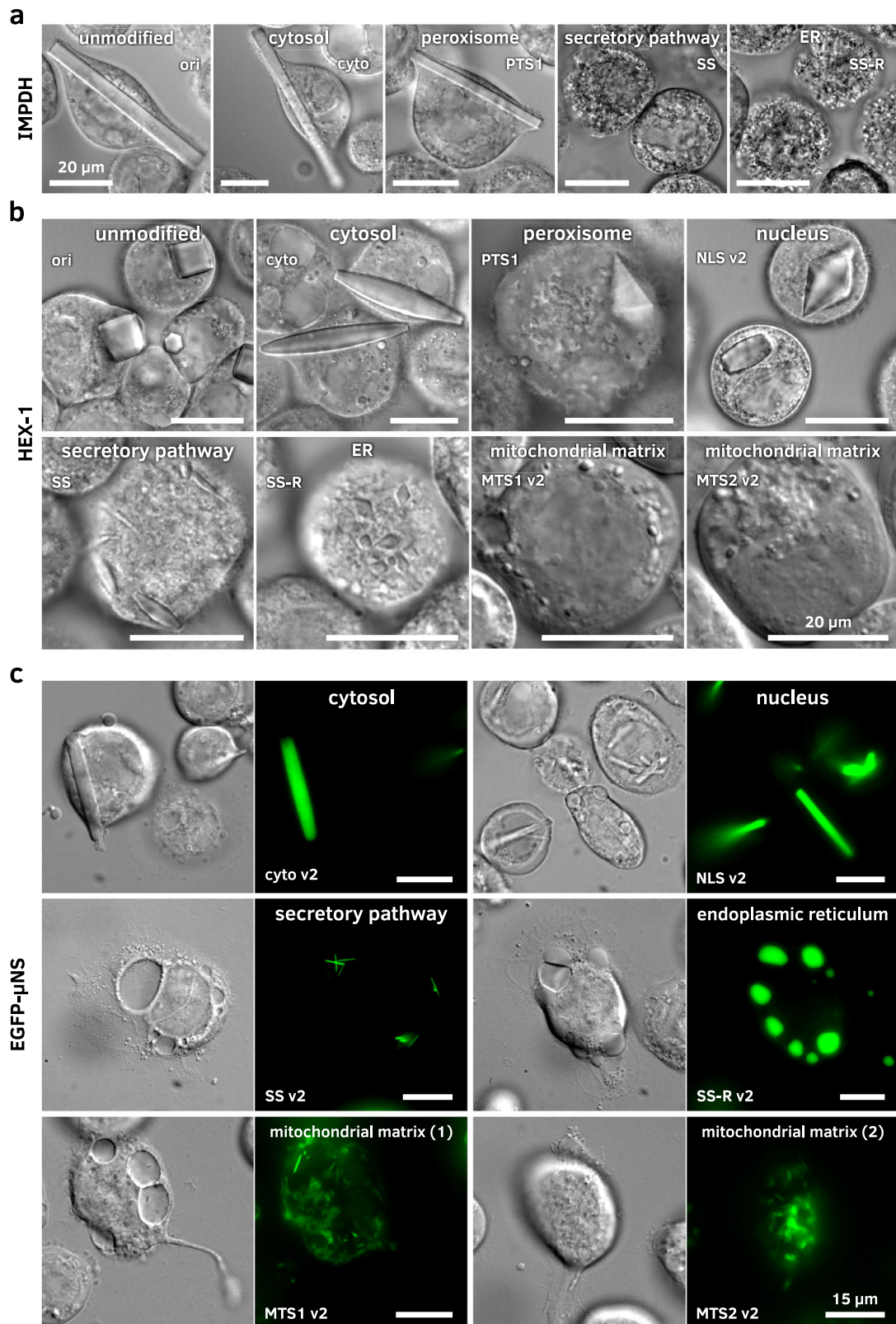
production, by standardizing the viral stock titration, and by using a cell sorter for selecting crystal containing cells from the infected culture. A comparable selection approach has been described by Boudes et al., employing the side scatter of infected insect cells[15]. However, as shown here, this parameter only selects baculovirus-infected cells, since the scattering behavior does not depend on the presence of protein crystals. In contrast, the EYFP marker protein represents a

reliable score for the amount of target protein produced in each cell that directly correlates with the crystallization probability, qualifying it as an advanced selection marker.

The established loading technique for MicroMesh mounts results in a cell monolayer with a minimal culture medium volume. Compared to alternative mounting techniques like silicon chips[45–47], the background signal is significantly reduced to a level observed for

Fig. 5 | Crystal morphology depends on the cellular compartment. rBVs were used to infect High Five **a**, **b** or Sf9 cells **c** at an MOI of 1. Imaging followed 4 dpi on a Nikon Ti2-E or Ts2R-FL microscope equipped with 100x objectives using the DIC contrast mode and EGFP wide-field fluorescence. Size bars for all images represent 20 µm **a**, **b** and 15 µm **c**, respectively. Representative images of three independent experiments are shown. **a** Compartment screening of IMPDH. Crystallization success and crystal morphology depend on the target organelle. The unmodified IMPDH (ori) as well as variants without the native (cyto) or with an artificial PTS1 motif crystallize within the cytosol, but no crystallization can be observed when the protein is translated into the ER. Crystal morphology is comparable in all compartments that enable crystallization. **b** Compartment screening of HEX-1. The

unmodified HEX-1 crystallizes in blocky hexagons. The N- and C-terminal addition of single amino acids (cyto) leads to spindle-like crystals in the cytosol. Additional amino acids at the C-terminus (translocation tags for peroxisome, nucleus and ER) result in a shift to mostly bipyramidal crystals. Differences in the compartmental environment also result in different crystal size distributions as visible for the ER, the secretory pathway, and the mitochondrial matrix. **c** Compartment screening of EGFP-µNS. Crystallization occurs in all tested compartments except for the ER. Without retention in the ER, however, thin and needle-shaped crystals occur. Also, targeting the mitochondrial matrix using both MTS versions results in very fine, needle-shaped crystals.

micro-patterned polyimide well mounts[48]. Thus, automated sample handling combined with raster-scanning allows efficient serial X-ray diffraction data collection. The CrystalDirect™ plates, previously established for in situ data collection of conventionally crystallized proteins with automated harvesting[40], enable in our setting a time-efficient screening at RT for diffraction of new targets in varying crystallization conditions.

As shown here, serial synchrotron diffraction data can be processed by using *CrystFEL* and *XDS*. The use of fixed targets in combination with helical line scans combines the advantages of serial crystallography with information obtained by rotation data collection[25,49]. However, the identification of crystal wedges is a prerequisite for processing with XDS, which is mainly used for single crystal rotational datasets. CrystFEL is optimized for still images in high multiplicity collected at RT at XFEL sources and thus, can be directly applied to serial synchrotron data.

We were able to identify naturally bound ligands and their conformations in the canonical nucleotide binding sites of the IMPDH ori and cyto structures, at a resolution of 2.3 and 2.4 Å, respectively. The second binding site is unambiguously occupied by GDP, instead of the previously proposed GMP[20], demonstrating that crystal isolation and storage can alter the information obtained due to hydrolysis of the ligand. Thus, diffraction data collection inside the intact cell preserves the native state of the biomolecules. Moreover, the identification of bound ligands in the native cellular environment is unique for intracellular protein crystallization and cannot be replaced by in silico protein folding predictions like AlphaFold[50,51].

Although InCellCryst has been overcome significant obstacles, some challenges remain to be addressed in future studies. Because of optimization for versatility, minimal impact on the protein sequence, and cost efficiency, our coning system achieves only medium throughput. Furthermore, automated crystal detection within the cell culture should be implemented into the pipeline. Initial developments to use machine learning for cell culture analysis have already been started[52]. The identification of additional parameters that affect the intracellular crystallization, next to the cellular compartment, will further improve the success rate. In this context, the addition of chemicals to the cell culture medium during crystal growth needs to be systematically evaluated. Finally, the crystallization of membrane proteins remains the most significant challenge. Although there is initial evidence that infected cells can form 2D crystals of membrane proteins[53], the possibility of generating 3D membrane stacks with the embedded and highly concentrated target protein has to be investigated in the future.

In summary, InCellCryst ranges from cloning over efficient crystal production and detection in insect cells, and adapted diffraction data collection inside viable, crystal-harboring cells to structure elucidation. As a supplement to conventional methods, it opens the intracellular crystallization approach for the broad structural biology community, thereby increasing the success rate, particularly by the compartment screening option. The rich source of biomolecules in the cellular environment allows the identification of native ligands by co-crystallization at quasi physiological conditions, representing a unique feature of InCellCryst.

Methods

Vector construction

Cloning of pFastBac1 vectors of the 1st generation cloning system containing sequences coding for translocation signal peptides, start and stop codons, as well as an *Ehe*I restriction site, was performed by annealing sense and antisense DNA oligonucleotides, restriction with FastDigest *Bam*HI and *Hind*III (Thermo Scientific) and ligation (T4 DNA Ligase, Thermo Scientific) into the equally restricted pFastBac1 plasmid. For sequences longer than 35 bases, overlapping oligonucleotides with single stranded 5′ ends were annealed, overhanging ends were filled using *Taq* DNA polymerase (Thermo Scientific), and the generated dsDNA was cloned as described below.

Cloning of pFastBac1 vectors of the 2nd generation cloning system was done using synthesized dsDNA sequences (BioCat GmbH) and restriction cloning. In brief, fragments were digested using FastDigest *Bam*HI and *Hind*III restriction enzymes and ligated into the equally restricted pFastBac1 plasmids. For vectors encoding an mTurqoise2 tag, the gene was PCR-amplified using primers containing *Nhe*I and *Hind*III cleavage sites. After restriction, the fragment was ligated into equally digested pFastBac1 v2 cyto and MTS plasmids. All generated vectors were checked by sanger sequencing (LGC Genomics), the sequences are summarized in Supplementary Table 1.

Target gene cloning

Inosin 5′-monophosphate dehydrogenase (IMPDH). Cloning of a His-tagged version of the IMPDH from *Trypanosoma brucei* (NCBI Accession M97794), denoted here as IMPDH N-His, has been previously described[20]. In brief, after gene amplification by PCR using AccuPrime™ Taq DNA polymerase (Invitrogen) with trypanosome cDNA according to the manufacturer's instructions and subcloning (TOPO-TA cloning kit, Invitrogen) into XL1-Blue competent *E. coli* cells (Agilent 200249), plasmid DNA was purified (QIAprep spin miniprep kit, Qiagen) and digested with *Bam*HI and *Hind*III. The extracted agarose gel fragment (QIAquick gel extraction kit, Qiagen) was cloned into equally digested pFastBacHTb expression plasmid (Invitrogen) that provided an additional gene sequence encoding a sixfold His-tag and a TEV protease cleavage site fused to the N-terminus of the TbIMPDH gene. For cloning of the unmodified IMPDH gene, denoted as IMPDH ori, the pFastBac1 IMPDH N-His vector was digested with FastDigest *Bam*HI and *Not*I, followed by ligation of the equally digested gene.

Since the IMPDH gene contains an *Ehe*I restriction site, incompatible with the 1st generation cloning system, a silent mutation was introduced using complementary primers containing a mismatched base pair. Using the ALLin HiFi DNA Polymerase (highQu), both strands of the respective plasmid were amplified in 16 cycles in separate reactions, followed by template digestion with FastDigest *Dpn*I (Thermo Scientific). Both reactions were then mixed, heated to 95 °C for 5 min and cooled down to RT over 1 h. After purification of the PCR product using the GeneJET PCR Purification Kit (Thermo Scientific), chemically competent *E. coli* DH5α cells (NEB C2987) were transformed. The amplified plasmid DNA was extracted using the GeneJET Plasmid Miniprep Kit (Thermo Scientific).
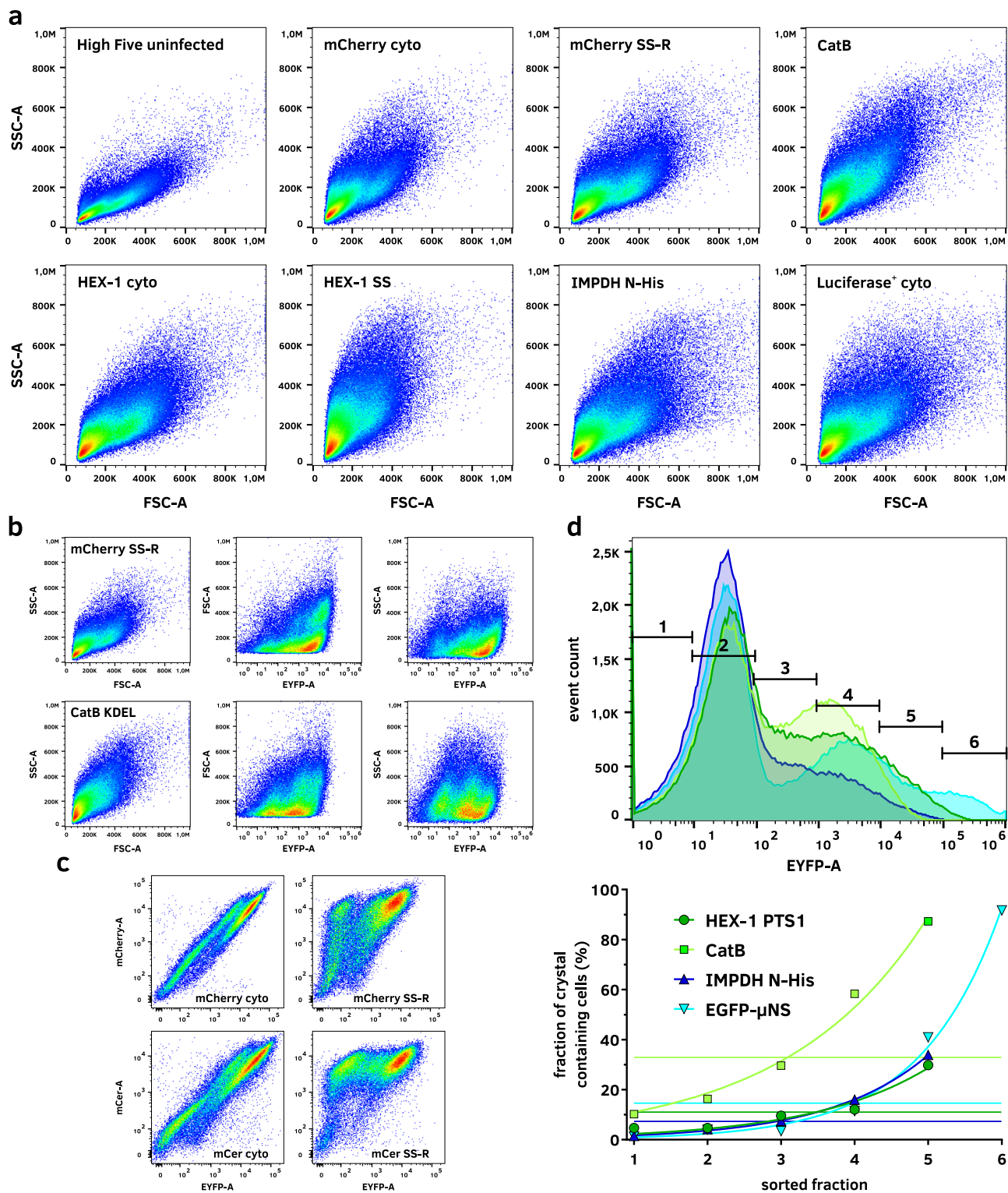
Fig. 6 | Enrichment of crystal containing insect cells using fluorescence-based cell sorting. a Forward (FSC-A) and side scatter (SSC-A) are no viable parameters for cell sorting due to a high similarity in the scattering behavior of crystal-containing and non-containing infected insect cells. **b** Evaluation of forward (FSC-A) and side scatter (SSC-A) against the fluorescence of EYFP does not produce a selectable population of crystal-containing cells that significantly differs between crystal-containing and non-containing cell cultures. **c** A high positive correlation is visible between EYFP production and cytosolic as well as endoplasmic target protein production in infected High Five cells. If the target protein is co-translated into the ER, the double negative population shifts towards a single positive population, indicating plasma membrane disruption, while the ER membranes remain intact. **d** FACS-based selection of infected cells based on their EYFP fluorescence allows an enrichment of crystal containing cells from the original culture. Straight lines in the lower plot indicate the fraction of crystal containing cells before cell sorting. Fractions are corresponding to the gates shown in the upper panel.
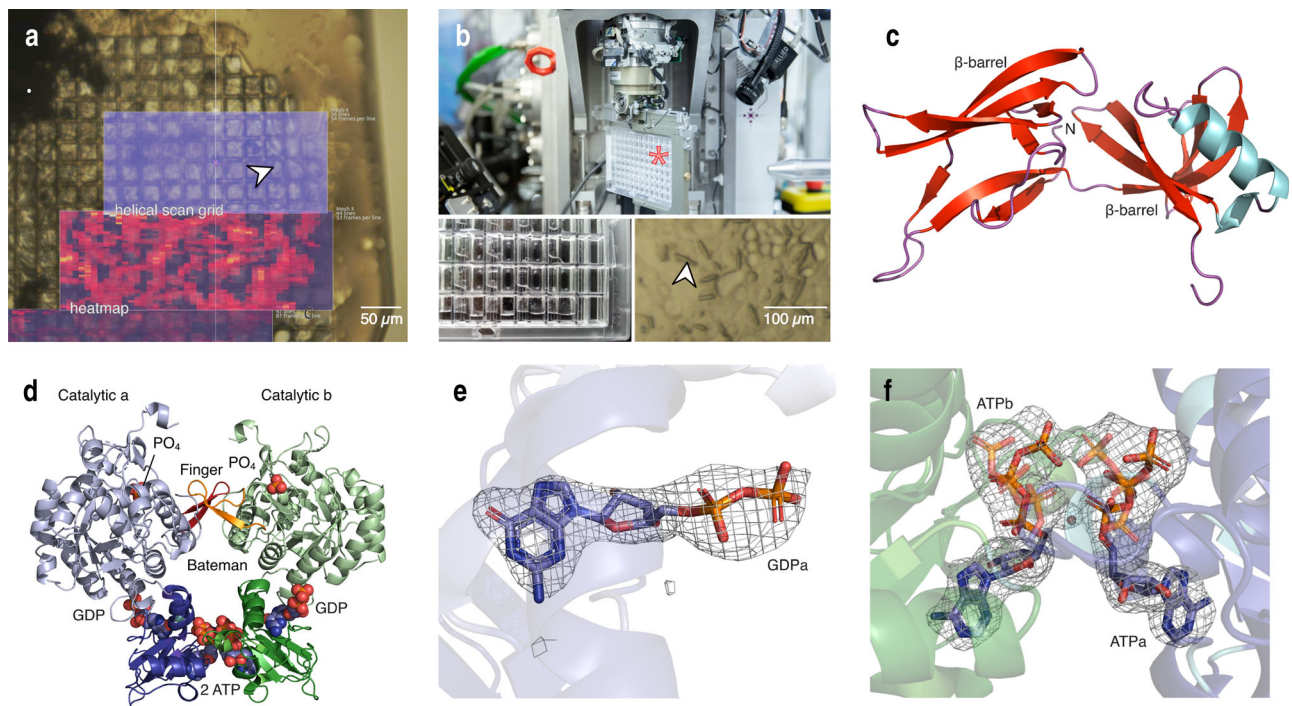
Fig. 7 | Intracellular crystal growth and data collection allow solving protein structures with genuine cofactors. a A MicroMesh mount loaded with High Five cells containing IMPDH ori crystals. Both the helical line scan and resulting *dozor* heat map generated at the P14 beamline (PETRA III, DESY, Hamburg) are shown. **b** A CrystalDirect™ plate for RT diffraction data collection (asterisk) is mounted on the Arinax MD3 diffractometer in an upright position (upper panel). Below, the loaded CrystalDirect™ plate is shown with HighFive cells carrying HEX-1 cyto crystals (close-up, right panel, arrow) adherent to the bottom foil. Next to each cell-containing well is a water-filled reservoir to keep humidity high and prevent cells from drying. **c** Cartoon representation of the HEX-1 ori structure. **d** IMPDH ori structure with natural ligands (ATP and GDP) in the Bateman domain and phosphates bound in the IMP binding site of the catalytic domain (spheres). Bateman domain and Finger domain in full colors, catalytic domains in light colors. **e** GDP within the canonical binding site 2 of the IMPDH ori Bateman domain. The omit map calculated with simulated annealing is shown at 3.0 sigma. **f** ATP within the canonical binding site 1 of the IMPDH ori Bateman domain. For clarity, phosphate moieties in the alternate conformation are shown in 50% transparency. The ligand omit map is shown at 3.0 sigma.

Moreover, primers were designed to amplify the IMPDH coding sequence excluding the native C-terminal PTS1 signal. The amplified fragment was cloned into the pFB1 cyto, PTS1, HA-N, SS and SS-R vectors of the 1st generation cloning system as mentioned above.

HEX-1. Cloning of the *Neurospora crassa* HEX-1 gene (NCBI Accession XM_958614, denoted as HEX-1 ori) was done by PCR amplification using primers containing *Bam*HI and *Hind*III restriction sites and ALLin HiFi DNA Polymerase (highQu). After digestion of the amplified DNA using FastDigest *Bam*HI and *Hind*III (Thermo Scientific), the fragment was ligated into the equally restricted pFastBac1 plasmid using T4 DNA Ligase (Thermo Scientific).

Cloning of the HEX-1 gene into the pFB1 cyto, PTS1, HA-C and SS vectors of the 1st generation cloning system was done by PCR amplification using primers excluding the N-terminal glycine, since this residue is already encoded in the vector sequences.

For cloning into the pFB1 v2 cyto, NLS, NLS HA-C, MTS1 and MTS2 vectors of the 2nd generation cloning system primers adding *Kpn*I and *Nhe*I restriction sites at the 5' and 3' ends of the HEX-1 gene, respectively, were used for PCR amplification. After digestion of the PCR product with FastDigest *Nhe*I and *Kpn*I (Thermo Scientific), the fragment was ligated into the equally digested vectors.

Luciferase. To establish compatibility with the 1st generation cloning system, an *Ehe*I restriction site within the gene coding for *Photinus pyralis* luciferase (NCBI Accession AB644228.1) was removed by introducing a silent mutation as described above. For gene amplification, two sets of primers were designed to either include (denoted as luciferase⁺) or exclude (denoted as luciferase⁻)

the native C-terminal PTS1 motif. Both PCR products were cloned into the pFB1 cyto and SS-R vectors of the 1st generation cloning system, the luciferase⁻ fragment additionally into the pFB1 PTS1 vector.

Cathepsin B (CatB). Cloning of the native pre-pro-form of *Trypanosoma brucei* cathepsin B (NCBI Accession XM_840086.1) has been previously described[33]. In brief, the gene was amplified by PCR using AccuPrime Taq DNA polymerase (Invitrogen) according to the manufacturer's instructions. After subcloning (TOPO-TA cloning kit, Invitrogen) into XL1-Blue–competent *E. coli* cells (Stratagene), plasmid DNA purification (QIAprep spin miniprep kit, Qiagen) and digestion with FastDigest *Bam*HI and *Xho*I (Thermo Scientific), the extracted gel fragment (QIAquick gel extraction kit, Qiagen) was cloned into equally digested pFastBac1 expression plasmid (Invitrogen). Moreover, a CatB construct with an additional C-terminal KDEL retention signal (denoted as CatB KDEL) as well as a construct without any signal sequence for cytosolic protein production (denoted as CatB -SS) was cloned by gene amplification using designed primers and ALLin HiFi DNA Polymerase (highQu). After PCR cleanup, ends were restricted using FastDigest *Bam*HI and *Xho*I (Thermo Scientific) for ligation into the equally restricted pFastBac1 plasmid.

EGFP-µNS. Cloning of the EGFP-tagged reoviral µNS protein (EGFP-µNS) has been previously described[2]. In brief, the pEGFP-C1 vector (BD Biosciences, Madrid, Spain) was used to express a fusion of *Aequorea victoria* EGFP to the N-terminus of the µNS region 448-605. *Eco*RI and *Bam*HI sites as well as start and stop codons were introduced at the

required positions in the avian reovirus S1133 M3 gene by PCR amplification using pGEMT-M3 as a template[54]. The PCR products were cut with EcoRI and BamHI (Thermo Scientific) and ligated to equally digested pEGFP-C1. The resulting plasmid pEGFP-C1-μNS(448-605) was again PCR amplified, introducing EcoRI and XbaI restriction sites, followed by ligation of the digested PCR products in the equally restricted pFastBac1 vector.

For the cellular compartment-screening using the 2nd generation cloning system, primers were designed adding KpnI and NheI restriction sites at the 5′ and 3′ ends of the EGFP-μNS coding sequence, respectively. After digestion with FastDigest KpnI and NheI (Thermo Scientific), the fragment was ligated into the equally digested pFB1 v2 cyto, NLS, SS, SS-R, MTS1 and MTS2 vectors.

For the screening of the impact of different fluorescent protein tags on μNS crystallization, the coding sequences of mCherry (NCBI Accession AY678264.1), mScarlet-I (NCBI Accession KY021424.1) and Xpa[12] were amplified using primers adding KpnI and SacI restriction sites at the 5′ and 3′ ends, respectively. After digestion with FastDigest KpnI and SacI enzymes (Thermo Scientific), the fragments were ligated into the equally digested pFB1 v2 EGFP-μNS cyto, NLS, SS, SS-R, MTS1 and MTS2 vectors.

For the crystallization screening of the μNS C-terminal domain (AA 448-605) without EGFP, the corresponding coding sequences was amplified with primers adding KpnI and NheI restriction sites at the 5′ and 3′ ends, respectively. After digestion with FastDigest KpnI and NheI (Thermo Scientific), the fragment was ligated into the equally digested pFB1 v2 cyto, NLS, SS, SS-R, MTS1 and MTS2 vectors.

Fluorescent proteins. The genes encoding the fluorescent proteins mCherry (NCBI Accession AY678264.1) and mCerulean (NCBI Accession KP666136.1) were PCR amplified and cloned into the pFB1 cyto and SS-R vectors of the 1st generation cloning system. Moreover, the EGFP coding sequence was amplified from the pFastBac1 EGFP-μNS vector and ligated into the pFB1 v2 cyto, NLS, SS, SS-R, MTS1 and MTS2 vectors, after digestion with FastDigest KpnI and NheI (Thermo Scientific).

1st generation cloning system. To be compatible with the 1st generation cloning system, PCR primers were designed to amplify the genes lacking start and stop codons. All genes were amplified using ALLin HiFi DNA Polymerase (highQu) and the purified fragments were ligated into the EheI-digested, modified pFastBac1 plasmids using T4 DNA Ligase (Thermo Scientific) supplemented with 0.2 μL FastDigest EheI (Thermo Scientific) per 20 μL ligation volume. Ligated plasmids were amplified in E. coli DH5α cells (NEB C2987) and correctness of the insertion was verified by Sanger sequencing (LGC Genomics).

2nd generation cloning system. To be compatible with the 2nd generation cloning system, PCR primers were designed to amplify the genes lacking start and stop codons while adding KpnI and NheI restriction sites at the 5′ and 3′ ends, respectively. Ligation of the digested PCR products into the equally restricted, modified pFastBac1 plasmids was performed using T4 DNA Ligase (Thermo Scientific). Ligated plasmids were amplified in E. coli DH5α cells (NEB C2987) and correctness of the insertion was verified by Sanger sequencing (LGC Genomics).

All primers used are listed in Supplementary Table 3.

Virus stock production
For the production of recombinant baculoviruses, competent E. coli DH10 (Invitrogen, used only for EGFP-μNS constructs) or E. coli DH10EmBacY[34] (Geneva Biotech) cells were transformed with previously cloned pFastBac1 plasmids according to the Bac-to-Bac manual (Invitrogen). The target gene replaced the viral polyhedrin gene using its promoter for high-yield expression. Recombinant bacmid DNA was purified using the ZR Bac DNA Miniprep Kit (Zymo Research) and correctness of the transposed sequence was analyzed by PCR using

pUC/M13 primers. Bacmid DNA was then used for lipofection of Spodoptera frugiperda Sf9 insect cells (Invitrogen B82501 and Merck 71104) grown in serum-free ESF921 cell culture medium (Expression Systems) at 27 °C using ESCORT IV reagent (Sigma Aldrich). In brief, 1 μg of bacmid DNA was used with 3 μL of ESCORT IV to transfect 5×10^5 Sf9 cells in 1 mL total volume in a well of a 12-well cell culture plate. 5 days after transfection the supernatant was harvested as the P1 stock and used to infect a 5 mL culture of 2×10^6 Sf9 cells/mL in an upright-standing 25 mL-cell culture flask that was incubated at 27 °C for 4 days with continuous shaking (100 rpm). The supernatant, representing the P2 virus stock, was harvested by centrifugation at $20,000 \times g$ for 30 s and subsequently used for virus titration and infection experiments in Sf9 and High Five cells.

Virus stock titration
To calculate the titer of the baculoviral stocks, a serial dilution assay was employed. In a 96-well plate, 180 μL of a cell suspension containing 3×10^4 High Five insect cells (Trichoplusia ni, Thermo Scientific B85502) in ESF921 cell culture medium were plated in each well and incubated for 30 min to let cells attach to the bottom. Then, a 1:10-dilution of the virus solution in ESF921 cell culture medium was prepared and 20 μL of this solution was added to 6 wells of the first row. For each serial dilution step the medium containing the virus was mixed in the well using a multi pipette and 20 μL of the supernatant was transferred into the next row. Pipette tips were discarded after each row and eight rows were prepared per titration. After 4 days at 27 °C, enhanced yellow fluorescent protein (EYFP) fluorescence of infected cells was evaluated and wells with at least two fluorescent cells were counted as positive. The virus titer was calculated using the TCID$_{50}$ (tissue culture infectious dose) according to a custom excel matrix (Supplementary Data 1) and the amount of viral stock used for an infection with a chosen MOI is calculated using the formula: virus stock [mL] = (MOI · cell number)/(0,69 · TCID$_{50}$/mL)[55].

Light microscopy
For detecting protein crystals within insect cells, infected cells were imaged either on a Nikon Ts2R-FL or on a Nikon Ti2-Eclipse microscope employing the differential interference contrast (DIC) mode. Images were taken using a Nikon Qi2 camera. Immunostained samples were imaged on the Nikon Ti2-Eclipse microscope using a SPECTRA X Light Engine (Lumencor). Confocal fluorescence images were collected on a Nikon Ti-Eclipse microscope equipped with a CSU X-1 spinning disk (Yokogawa), a Laser Combiner System, 400 series, and an iXon^{EM} + EMCCD camera (both from Andor Technology). Cell culture imaging for crystal counting was performed using a Zeiss Axio Observer.Z1 microscope equipped with an AxioCam MRm microscope camera (Carl Zeiss AG) employing the DIC mode.

Electron microscopy
For transmission electron microscopy (TEM), 0.9×10^6 infected Sf9 or High Five insect cells were harvested 4 dpi (days post infection) from the multi-well plate, pelletized for 3 min at $500 \times g$, fixed by resuspension in 1 mL cold cacodylate buffer (60 mM, pH 7.35) containing 2% glutaraldehyde, 6 g/L paraformaldehyde (PFA), and 0.3 g/L CaCl$_2$, followed by incubation for at least 24 h at 4 °C. At RT, the cells were then pelletized for 3 min at $900 \times g$ and the supernatant was carefully removed. After a 30 min wash in 120 mM sodium cacodylate buffer (pH 7), cells were postfixed in 1% osmium tetroxide (in 120 mM sodium cacodylate, pH 7) for 2 h. After two additional washing steps (15 min each) in sodium cacodylate buffer, cells were dehydrated in ethanol for 2×15 min at each step with increasing ethanol concentrations (30–100%; 10%-steps) and two 30 min incubations with 100% ethanol at the end. At the 70% ethanol step, cells were incubated overnight. Subsequently, samples were cleared in propylene oxide in two 30 min incubations. Embedding in Araldite was done

using a mixture of 10 mL Araldite M, 10 mL Araldite M hardener 964, 0.3–0.4 mL Araldite M accelerator 960 and 0.1–0.2 mL dibutyl phthalate. First the cells were incubated for 1 h in a mixture of Araldite with propylene oxide (1:2), then another hour in a 2:1 mixture. After removal of this supernatant, the pellet was left for 2 min to evaporate the rest of the propylene oxide. Finally, the pellet was carefully overlayed with 500 μL of the Araldite M mixture and left for hardening at 60 °C for 48 h. Sections were cut with a Leica Ultracut E microtome to 60–90 nm thickness and transferred onto TEM-grids (G2410C, Plano GmbH). Sections were then stained with a Leica EM AC20 for 30 min in 0.5% uranyl acetate (Ultrostain I, Leica) at 40 °C and for 7 min in 3% lead citrate (Ultrostain II, Leica) at 20 °C. Imaging was done using a JEOL JEM-1011.

Immunofluorescence staining

A total of 0.5×10^6 High Five cells were plated and infected on a round glass coverslip (Ø 25 mm, No. 1) in a 6-well cell culture plate in 2 mL ESF921 cell culture medium. Fixation of the cells was done 4 dpi by adding 700 μL fixation buffer (40 g/L PFA in 0.5× PBS, pH 7.4) directly to the well and incubation for 2 min at RT, followed by a medium exchange to pure fixation buffer and additional incubation for 15 min at RT. The cells were then washed 3× for 10 min each in PBS. Permeabilization and blocking was performed using 0.1% Triton X-100 and 50 g/L dry milk powder in PBS for 1 h at RT. Subsequently, the coverslip was washed twice with PBS, taken from the well and placed on a plastic holder. For staining, 200 μL of a 1:1000 dilution of a mouse αHA-epitope tag antibody (BioLegend 901501, clone 16B12) were carefully pipetted onto the coverslip and incubated for 1 h at RT. The coverslip was washed 3× in PBS and incubated for 1 h at RT in the dark with 200 μL of a 1:15,000 dilution of αMs DyLight 549 antibody (Jackson ImmunoResearch 115-585-003). After three wash steps in PBS, the coverslips were embedded in 90% glycerol on a microscopy specimen slide.

X-ray powder diffraction (XRPD)

A total of 9×10^5 cells were plated in 2 mL ESF921 cell culture medium in one well of a 6-well cell culture plate, rBV infected (MOI 1.0), and incubated at 27 °C. 4 dpi cells were gently flushed from the well bottom, transferred to a 1.5 mL reaction tube and left to settle down for 30 min at RT. 40 μL of the dense cell suspension were transferred from the tube bottom into a 3.5 mm wide Micro RT tube (MiTeGen) using a gel loading tip, followed by 2× centrifugation for 1 min at $100 \times g$ in a fixed-angle rotor with a 180° rotation in between runs. After removal of the supernatant the tube was cut to 20 mm in length and slipped on a Reusable Goniometer Base B1A (MiTeGen). X-ray scattering experiments followed immediately at the EMBL beamline P14 (PETRAIII, DESY, Hamburg). Using a photon energy of 12.7 keV with a photon flux of 5×10^{12} ph/s at the sample position, diffraction data were recorded at RT on an EIGER 16 M detector (DECTRIS, Switzerland). Using a collimated beam with a 75 x 75 μm focal spot, 300 detector frames were recorded per sample with a single-frame exposure time of 1 s, resulting in a total exposure time of 5 min per data set. During exposure the sample was vertically translated and rotated by 2° per frame. All frames were merged using *merge2cbf* from the XDS software suite and background subtraction was done using the *adxv* image viewer (version x86_64CentOS7). For that, 90% of the calculated background intensity was subtracted from each pixel using the moving average option of *adxv*.

Small angle X-ray scattering (SAXS-XRPD)

To perform the SAXS-XRPD approach[37], 9×10^5 cells were plated in 2 mL ESF921 cell culture medium in one well of a 6-well cell culture plate, rBV infected (MOI 1.0), and incubated at 27 °C for 4 days. The cells were gently flushed from the well bottom, transferred to a 1.5 mL reaction tube and centrifuged for 30 s at $270 \times g$. The cell pellet was resuspended in 25 μL of Tris-buffered saline (TBS; 20 mM Tris, 150 mM NaCl pH 7.0)

and 45 μL of the suspension were transferred into 0.2 mL sample tubes. X-ray scattering experiments followed immediately at the EMBL beamline P12 (PETRAIII, DESY, Hamburg)[56]. Using a photon energy of 10 keV with a photon flux of 1×10^{13} ph/s at the sample position, data [I(s) versus s, where s $= 4\pi \sin(\theta)/\lambda$ with 2θ as the scattering angle and λ as the X-ray wavelength] were recorded at a distance of 3 m between sample and detector using a PILATUS 6 M detector (DECTRIS, Switzerland). 30 μL of the samples were transferred into a temperature-controlled 1.8 mm quartz capillary using the automatic bioSAXS sample changer (Arinax)[57]. Using a focal spot of 0.2×0.12 mm (FWHM) in a fixed-position measurement at 20 °C, 40 detector frames were recorded per sample followed by 40 frames of the TBS buffer for subtraction during data analysis, all with a single-frame exposure time of 45 ms and a readout time of 5 ms, resulting in a total 2 s exposure time per data set. Data analysis and graph preparation was performed with ATSAS 3.0.[58]

Flow cytometry

Single cell-analysis of infected insect cells was performed at 4 dpi of 9×10^5 High Five cells in a well of a 6-well cell culture plate and incubation at 27 °C. Flow analysis of the correlation of protein production levels was carried out on a LSR II (BD Bioscience) equipped with 405 nm, 488 nm, and 561 nm laser lines. Forward and side scatter were measured using the 488 nm laser light. mCerulean fluorescence was excited using the 405 nm wavelength and mCherry fluorescence using the 561 nm wavelength. Photodiode sensitivity was adjusted individually to fit the fluorescence intensity of the respective channels. Forward and side scatter analysis as well as sorting of crystal containing cells was done using a Sony SH800S cell sorter equipped with 488 and 561 nm lasers. To compensate for the large cell diameter of infected High Five insect cells, a 130-μm microfluidic sorting chip was used. Cells were sorted in PBS sheath fluid into 5 mL FACS-tubes and analyzed on a Nikon Ti-2 Eclipse microscope equipped with a Nikon Qi2 camera using high resolution DIC. For that, cells were pipetted onto a glass coverslip out of the bottom of the tube after settling down.

CrystalDirect™ plate preparation

For in cellulo X-ray diffraction data collection at RT at the EMBL beamline P14 (PETRAIII, DESY, Hamburg), High Five insect cells were directly grown on CrystalDirect™ plates. CrystalDirect™ plates are modified 96-well vapor diffusion plates glued to a bottom crystallization support consisting of a 25 μm thick cyclic olefin copolymer (COC) film[40]. The plates were sterilized by UV light for 40 min and incubated with 75 μL/well of a 0.2 mg/mL poly-D-lysine solution for 1 h at RT. The wells were then washed twice with 100 μL PBS/well. 1×10^4 High Five cells per well were plated in 50 μL of ESF921 cell culture medium. Cells were left to adhere to the foil for 30 min. Subsequently, cells were infected with a rBV using an MOI of 1 by exchanging the medium with the virus stock diluted in 50 μL ESF921 cell culture medium supplemented with 25 % FBS. To avoid drying, the outermost row of wells was filled with 100 μL water and the plate was covered with the lid of a 96-well plate. After 4 days of incubation at 27 °C, the plates were prepared for the X-ray diffraction measurement. For that, 50 μL of water was filled into the reservoirs and the medium was completely removed from the cells. Then, the plate was immediately enclosed airtight with a second foil on top and directly mounted upright at the appropriate sample holder of the MD3 diffractometer at the P14 beamline.

Sample preparation for diffraction data collection at 100 K

For measurements at 100 K, 0.9×10^6 High Five cells were plated in 2 mL ESF921 cell culture medium in one well of a 6-well plate and infected with a rBV at an MOI of 1. After 4 days incubation at 27 °C, the insect cells carrying protein crystals were carefully resuspended in 1 mL ESF921 cell culture medium, transferred to a 1.5 mL tube and left to settle down for 10 min. 0.5 μL of the loose cell pellet was pipetted

onto a MicroMesh (700/25; MiTeGen), previously mounted on a goniometer base (B1A, MiTeGen), which was positioned by a custom-made holder in the optical focus of a standard upright cell culture light microscope. The excess medium was manually removed from the back of the MicroMesh using an extra fine liquid wick (MiTeGen). For cryoprotection, 0.5 μL of a 40% PEG200 solution diluted in ESF921 cell culture medium were pipetted onto the cells. Immediately, excess liquid was again removed from the back of the mesh using a liquid wick. During that procedure, the cells were kept in a 90 % humidity air stream to avoid drying of the sample. Subsequently, the cell-loaded MicroMesh was manually immersed in liquid nitrogen to ensure rapid freezing. For storage and transport to the beamline the samples were kept suspended in liquid nitrogen.

X-ray diffraction data collection

Data collection was conducted at the EMBL microfocus beamline P14 at the PETRAIII storage ring (DESY, Hamburg). Data collection was controlled using the Hamburg version of the mxCuBE v2 user interface[59]. For measurements at 100 K, the crystal-carrying insect cells loaded on a MicroMesh were mounted on a mini-kappa goniostat attached to an MD3 diffractometer using the MARVIN Sample Changer. Samples were kept at 100 K by a gaseous nitrogen stream. To collect complete datasets using the EIGER 16 M detector (DECTRIS, Switzerland), meshes were scanned employing a serial helical grid scan strategy[25,41]. For that, rasters were predefined across the mesh surface with a defined spacing between data collection points. During diffraction data collection in each vertical line, the goniostat was rotated and translated continuously. At the end of each line, the mesh was translated to the side and the rotation and translation direction was inverted. For RT measurements the CrystalDirect™ plates were mounted onto a plate holder attached to the MD3 diffractometer and intracellular crystals were illuminated using the same serial helical line scan strategy as described above. Details for the specific parameters used to collect each dataset are presented in Supplementary Table 4.

Data processing and structure determination

Collected diffraction patterns were processed either applying the *CrystFEL* software suite (versions 0.9.1, 0.10.0, and 0.10.1) written for snapshot serial crystallography[42,49], or using *XDS*[43], originally designed to process single-crystal monochromatic diffraction data recorded by simple rotation of the crystal, after identifying small rotational datasets.

Data processing using *CrystFEL*. *CrystFEL* versions 0.9.1 (HEX-1 RT), 0.10.0 and 0.10.1 (HEX-1 cyto v2, HEX-1 ori, IMPDH cyto and IMPDH ori) were used to confirm peak detection, as well as for indexing and integration. Diffraction data from all datasets were merged using *CrystFEL* version 0.9.1.

In version 0.10.0, the *CrystFEL* GUI was used to confirm peak detection, while the command line was used for final indexing, integration and merging of the diffraction data. Bad pixels and spaces between the detector panels were manually flagged in the geometry file or were excluded by masks generated using *CsPadMaskMaker* (https://github.com/kbeyerlein/CsPadMaskMaker). The exact detector distance was refined in iterative cycles manually and by using the *detector-shift* (*CrystFEL*) script, respectively. *Peakfinder8* was used for peak detection using the parameters displayed in Supplementary Table 5, leading to the hit and indexing rates indicated in Supplementary Table 2. To determine the unit cell, all patterns were indexed using *mosflm-latt-nocell*, for the final indexing *xgandalf*[60], *mosflm-latt-nocell, mosflm-nolatt-cell*[61], *TakeTwo*[62] and *XDS* were invoked. A second apparent unit cell population of HEX-1 cyto, differing in the c-axis, and of HEX-1 cyto v2, differing in all three axes, was cleaned from the stream-file using the *CrystFEL*-script *stream-grep*.

Before merging the Bragg reflection intensities from single crystal diffraction patterns using *partialator* with one to three rounds of post-

refinement using the partiality correction *unity, offset* or *xsphere* and *push-res* 1 to 2, reflections from precipitated salt were removed from the stream file by applying resolution dependent thresholds for the maximum reflection intensities. The *peakogram-stream* (*CrystFEL*) output was used to determine the values for filtering and to confirm the successful removal of the powder diffraction signal. *CrystFEL* hkl-files were converted to MTZ-files using *create_mtz* or *get_hkl*. Figures of merit were calculated using *check_hkl* and *compare_hkl*.

Data processing with *XDS*. To process data collected by helical line scans using *XDS* small rotational data sets needed to be defined. We used a custom-made script supplied here as Supplementary Data 2 and 3. In brief, the pattern with the maximum spot count given by *dozor*[40,63] was defined as the crystal center and adjacent frames were added to the dataset if they contained a minimum number of spots. To distinguish whether two crystal wedges belong to the same crystal, the orientation of overlapping crystals was compared. Crystals with a deviation of more than 6 degrees between the unique axis' were defined as unique. Crystals were merged and quality parameters calculated using *XSCALE*[43].

Diffraction power over time in CrystalDirect™ plates

To determine whether the diffraction capability of crystals during ~6 h of data collection in CrystalDirect™ plates changed, the sorted *stream* of the HEX-1 cyto (RT) dataset was divided into four parts with an equivalent number of chunks using *truncate_stream*. The average and best resolution as determined by CrystFEL for each fourth of the dataset was then determined (*ave-resolution* script).

Refinement and model building

Phases were retrieved by Molecular Replacement with *Phaser*[64–66] using the A-chain of human type I IMPDH, crystallized by hanging drop vapor diffusion (PDB 1JCN) with two copies in the asymmetric unit, or using the HEX-1 structure previously obtained from crystals obtained by sitting drop vapor diffusion as a search model (PDB 1KHI), respectively. Structure refinements for all generated datasets were carried out using *PHENIX6* (version 1.19.2-4158)[65,66] and iterative cycles of manual model building in *Coot*[67] (version 0.9.7).

Simulated-annealing omit maps were calculated in *PHENIX*. Applying the *FastFourierTransform* program, electron density maps with *ccp4* extensions were saved and loaded in the PyMOL Molecular Graphics System (version 4.5.0, Schrödinger, LLC) for graphical illustrations of contoured omit maps.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding authors upon request. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession codes 8C51 (IMPDH cyto); 8C53 (IMPDH ori, processed with CrystFEL); 8CGY (IMPDH ori, processed with XDS); 8CD5 (HEX-1 ori, processed with CrystFEL); 8CGX (HEX-1 ori, processed with XDS); 8CD4 (HEX-1 cyto, 100 K); 8C5K (HEX-1 cyto, RT); and 8CD6 (HEX-1 cyto v2). We have referred to the previously published PDB Codes 7ASX; 1JCN [https://www.wwpdb.org/pdb?id=pdb_00007asx]; 6RFU; and 1KHI. The source data underlying Figs. 1, 3, and 6d are provided as a Source Data file. Source data are provided with this paper.

Code availability

The code of two scripts developed to improve serial X-ray diffraction data processing using XDS are provided in the Source Data file (Supplementary Data 2, XDS-script to identify crystal wedges; Supplementary Data 3, XDS-script to check for overlapping crystals).

References

1. Fan, G. Y. et al. In vivo calcineurin crystals formed using the baculovirus expression system. *Microsc. Res. Tech*. **34**, 77–86 (1996).
2. Brandariz-Nuñez, A., Menaya-Vargas, R., Benavente, J. & Martinez-Costas, J. Avian reovirus microNS protein forms homo-oligomeric inclusions in a microtubule-independent fashion, which involves specific regions of its C-terminal domain. *J. Virol*. **84**, 4289–301 (2010).
3. Hasegawa, H. et al. In vivo crystallization of human IgG in the endoplasmic reticulum of engineered Chinese hamster ovary (CHO.) cells. *J. Biol. Chem*. 286, 19917–19931 (2011). .
4. Doye, J. P. K. & Poon, W. C. K. Protein crystallization in vivo. *Curr. Opin. Colloid Interface Sci*. **11**, 40–46 (2006).
5. Yamamoto, M. et al. Protein microcrystallography using synchrotron radiation. *IUCrJ* **4**, 529–539 (2017).
6. Coulibaly, F. et al. The molecular organization of cypovirus polyhedra. *Nature* **446**, 97–101 (2007).
7. Redecke, L. et al. Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser. *Sci. (80-.)* **339**, 227–230 (2013).
8. Schönherr, R., Rudolph, J. M. & Redecke, L. Protein crystallization in living cells. *Biol. Chem*. **399**, 1–22 (2018).
9. Mudogo, C. N., Falke, S., Brognaro, H., Duszenko, M. & Betzel, C. Protein phase separation and determinants of in cell crystallization. *Traffic* https://doi.org/10.1111/tra.12711 (2020).
10. Ginn, H. M. et al. Structure of CPV17 polyhedrin determined by the improved analysis of serial femtosecond crystallographic data. *Nat. Commun*. **6**, 1–8 (2015).
11. Baskaran, Y. et al. An in cellulo-derived structure of PAK4 in complex with its inhibitor Inka1. *Nat. Commun*. **6**, 1–11 (2015).
12. Tsutsui, H. et al. A diffraction-quality protein crystal processed as an autophagic cargo. *Mol. Cell* **58**, 186–193 (2015).
13. Banerjee, S. et al. Structure of a heterogeneous, glycosylated, lipid-bound, in vivo-grown protein crystal at atomic resolution from the viviparous cockroach Diploptera punctata. *IUCrJ* **3**, 282–293 (2016).
14. Colletier, J. P. et al. De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, 43–47 (2016).
15. Boudes, M., Garriga, D., Fryga, A., Caradoc-Davies, T. & Coulibaly, F. A pipeline for structure determination of in vivo-grown crystals using in cellulo diffraction. *Acta Crystallogr. Sect. D. Struct. Biol*. **72**, 576–585 (2016).
16. Gati, C. et al. Atomic structure of granulin determined from native nanocrystalline granulovirus using an X-ray free-electron laser. *Proc. Natl Acad. Sci. USA* **114**, 2247–2252 (2017).
17. Lahey-Rudolph, J. M. et al. Fixed-target serial femtosecond crystallography using in cellulo grown microcrystals. *IUCrJ* **8**, 665–677 (2021).
18. Tetreau, G. et al. De novo determination of mosquitocidal Cry11Aa and Cry11Ba structures from naturally-occurring nanocrystals. *Nat. Commun*. **13**, 4376 (2022).
19. Oberthuer, D. et al. Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep*. **7**, 1–10 (2017).
20. Nass, K. et al. In cellulo crystallization of Trypanosoma brucei IMP dehydrogenase enables the identification of genuine co-factors. *Nat. Commun*. **11**, 620 (2020).
21. Coulibaly, F. et al. The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc. Natl Acad. Sci. USA* **106**, 22205–10 (2009).
22. Ji, X. et al. How baculovirus polyhedra fit square pegs into round holes to robustly package viruses. *EMBO J*. **29**, 505–514 (2010).
23. Ji, X. et al. Polyhedra structures and the evolution of the insect viruses. *J. Struct. Biol*. **192**, 88–99 (2015).
24. Axford, D., Ji, X., Stuart, D. I. & Sutton, G. In cellulo structure determination of a novel cypovirus polyhedrin. *Acta Crystallogr. Sect. D. Biol. Crystallogr*. **70**, 1435–1441 (2014).
25. Gati, C. et al. Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* **1**, 87–94 (2014).
26. Sawaya, M. R. et al. Protein crystal structure obtained at 2.9 A resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl Acad. Sci. USA* **111**, 12769–74 (2014).
27. Chiu, E. et al. Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization. *Proc. Natl Acad. Sci. USA* **112**, 3973–3978 (2015).
28. Tang, Y., Saul, J., Nagaratnam, N. & Garcia, J. M. M. Construction of gateway - compatible baculovirus expression vectors for high - throughput protein expression and in vivo microcrystal screening. *Sci. Rep*. https://doi.org/10.1038/s41598-020-70163-2 (2020).
29. Wampler, R. D. et al. Selective detection of protein crystals by second harmonic microscopy. *J. Am. Chem. Soc*. **130**, 14076–14077 (2008).
30. Hall, C. E. Electron microscopy of crystalline edestin. *J. Biol. Chem*. **185**, 45–51 (1950).
31. Schönherr, R. et al. Real-time investigation of dynamic protein crystallization in living cellsa) Real-time investigation of dynamic protein crystallization. **041712**, (2015).
32. Nagaratnam, N. et al. Enhanced X-ray diffraction of in vivo-grown μnS crystals by viscous jets at XFELs. *Acta Crystallogr. Sect. F. Struct. Biol. Commun*. **76**, 278–289 (2020).
33. Koopmann, R. et al. In vivo protein crystallization opens new routes in structural biology. *Nat. Methods* **9**, 259–262 (2012).
34. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol*. **172**, 45–54 (2010).
35. Wilde, M., Klausberger, M., Palmberger, D., Ernst, W. & Grabherr, R. Tnao38, high five and Sf9-evaluation of host-virus interactions in three different insect cell lines: Baculovirus production and recombinant protein expression. *Biotechnol. Lett*. **36**, 743–749 (2014).
36. Rohrmann, G. F. *Baculovirus Molecular Biology. National Center for Biotechnology Information (US)*. (2019).
37. Lahey-Rudolph, J. M. et al. Rapid screening of in cellulo grown protein crystals via a small-angle X-ray scattering/X-ray powder diffraction synergistic approach. *J. Appl. Crystallogr*. **53**, 1169–1180 (2020).
38. Yuan, P. et al. A HEX-1 crystal lattice required for Woronin body function in Neurospora crassa. *Nat. Struct. Biol*. **10**, 264–270 (2003).
39. Norton-Baker, B. et al. A simple vapor-diffusion method enables protein crystallization inside the HARE serial crystallography chip. *Acta Crystallogr. Sect. D. Struct. Biol*. **77**, 820–834 (2021).
40. Zander, U. et al. Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallogr. Sect. D. Struct. Biol*. **72**, 454–466 (2016).
41. Shahsavar, A. et al. Structural insights into the inhibition of glycine reuptake. *Nature* **591**, 677–681 (2021).
42. White, T. A. et al. CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr*. **45**, 335–341 (2012).
43. Kabsch, W. Xds. *Acta Crystallogr. Sect. D. Biol. Crystallogr*. **66**, 125–132 (2010).
44. Metz, M. et al. Macromolecular phasing using diffraction from multiple crystal forms. *Acta Crystallogr. Sect. A* **77**, 19–35 (2021).
45. Coquelle, N. et al. Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr. Sect. D. Biol. Crystallogr*. **71**, 1184–96 (2015).
46. Mueller, C. et al. Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct. Dyn*. **2**, 054302 (2015).
47. Roedig, P. et al. A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. *Sci. Rep*. **5**, 1–11 (2015).
48. Guo, G. et al. Sample manipulation and data assembly for robust microcrystal synchrotron crystallography. *IUCrJ* **5**, 238–246 (2018).

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

50. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

51. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

52. Kardoost, A. et al. Convolutional neural network approach for the automated identification of in cellulo crystals. *bioRxiv* 2023.03.28.533948 https://doi.org/10.1101/2023.03.28.533948 (2023).

53. Gualtieri, E. J. et al. Detection of membrane protein two-dimensional crystals in living cells. *Biophys. J.* **100**, 207–214 (2011).

54. Tourís-Otero, F., Cortez-San Martín, M., Martínez-Costas, J. & Benavente, J. Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of sigmaNS and lambdaA to microNS inclusions. *J. Mol. Biol.* **341**, 361–374 (2004).

55. O'Reilly, D. R., Miller, L. K. & Luckow, V. A. *Baculovirus expression vectors: a laboratory manual*. (Oxford University Press, 1994).

56. Blanchet, C. E. et al. Highly automated and time-resolved BioSAXS at the P12 beamline of EMBL Hamburg. **23**, 2015 (2015).

57. Round, A. et al. BioSAXS Sample Changer: A robotic sample changer for rapid and reliable high-throughput X-ray solution scattering experiments. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **71**, 67–75 (2015).

58. Manalastas-Cantos, K. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355 (2021).

59. Oscarsson, M. et al. MXCuBE2: The dawn of MXCuBE collaboration. *J. Synchrotron Radiat.* **26**, 393–405 (2019).

60. Gevorkov, Y. et al. XGANDALF - Extended gradient descent algorithm for lattice finding. *Acta Crystallogr. Sect. A Found. Adv.* **75**, 694–704 (2019).

61. Powell, H. R., Johnson, O. & Leslie, A. G. W. Autoindexing diffraction images with iMosflm. *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1195–1203 (2013).

62. Ginn, H. M. et al. TakeTwo: an indexing algorithm suited to still images with known crystal parameters. *Acta Crystallogr. Sect. D., Struct. Biol.* **72**, 956–965 (2016).

63. Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M. W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr. D. Biol. Crystallogr.* **71**, 1757–1767 (2015).

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

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

66. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D., Struct. Biol.* **75**, 861–877 (2019).

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

Author contributions

R.S. and L.R. conceived the experiments, which were designed with J.M.L., J.B., P.K., R.D., G.B. and T.R.S.; Supervised by L.R., R.S., J.B., J.M.L., M.H., S.N., J.K., M.W. performed gene cloning, insect cell culture, and rBV generation, as well as intracellular crystallization experiments; TEM experiments were performed by P.K.; Crystal-containing cell samples were prepared and characterized by R.S., J.B., J.M.L., M.H., J.K., S.N. and M.W. under supervision of L.R.; X-ray diffraction experiments were carried out by J.M.L., R.S., J.B. and L.R.; Beamline setup was done by T.R.S. and G.B. The sample delivery system was developed by T.R.S. and G.B. and operated by R.S., J.M.L., J.B. and L.R. J.B. and J.M.L. processed the diffraction data and performed molecular replacement, refined the structures, and calculated the electron density maps. The manuscript was prepared by R.S., J.M.L, J.B. and L.R. with discussions and improvements from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-45985-7.

Correspondence and requests for materials should be addressed to Lars Redecke.

Peer review information *Nature Communications* thanks Jose Gavira and Elena Andreeva for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

3.1.1 Synopsis

Intracellular crystallization is a niche technique in structural biology.[101] This is mostly due to lacking know-how and the fact that the scientific community lacks elaborate and rapid procedures for *in cellulo* crystallization of unknown targets. A pipeline suggestion from Boudes and coworkers in 2016 included baculovirus expression vector system (BEVS) recombinant protein expression in Sf9 cells, bright-field microscopy and enrichment of polyhedrin crystal-containing (or rather, of infected) cells by fluorescence associated cell sorting (FACS), as well a a trypan-blue stain during data collection and heavy metal soaking (of plasma membrane-transmissive, and respectively dead cells).[111] Four years later, Tang and coworkers added a high-throughput *in cellulo* crystallization method which exploits gateway-compatible BEVS cloning libraries, including SONICC and TEM detection of crystals in infected Sf9 cells[184].

The proposed InCellCryst pipeline fills this know-how gap by presenting greatly improved procedures for insect cell *in cellulo* crystallography: The expressed target gene can be directed to different compartments (**Figure 9**) by cloning into customized InCellCryst compartment screening pFB1 vectors.
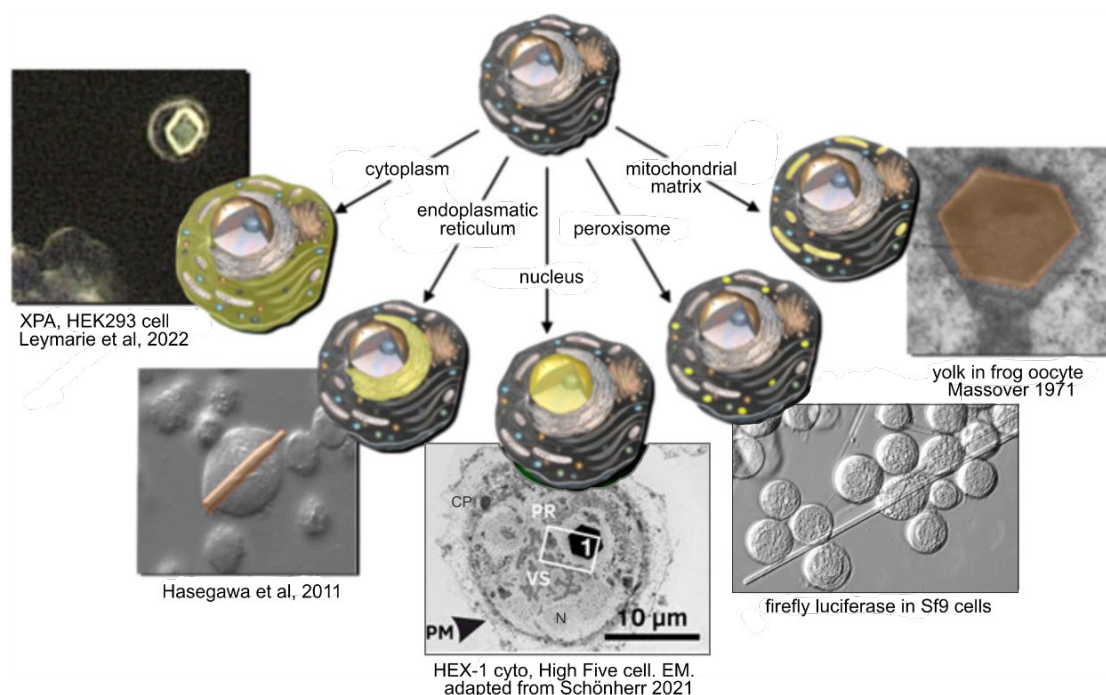
Figure 9 Compartment screening cloning strategy for crystallization in living cells.
Synthesized proteins are directed to cytoplasm, ER, nucleus, peroxisome or mitochondria in order to maximise crystallization screening success. Organelles are highlighted ochreous, crystals that are not fluorescent themselves coppery. Image adapted from Leymarie et al (2022)[185]. PM – plasma membrane, CP – cytoplasm, N – nucleus, VS – virogenic stroma, PR – peripher region.

Cloning vectors are designed to include translocation sequences recognized for the secretion pathway (ER following Golgi apparatus and secretory vesicles), transport to the ER

including a 'KDEL' retention signal, the C-terminal 'SKL' sequence for transport to peroxisomes, a mitochondrial targeting sequence for transport to the mitochondrial matrix, or the lack of a translocation sequence for the expressed genes to stay in the cytoplasm. If proteins bearing a KDEL-sequence are transported to the cis-Golgi apparatus, receptors bind to the retention signal and COP I vesicle-mediated backtransport to the ER is initiated.[186]

It is demonstrated that screening for crystals in different compartments can enhance the chance of a hit. So far, only *Nc*HEX-1 is known to crystallize in every tested cellular compartment. As with natively crystallizing proteins in cells[65,66,68–70,119], after recombinant gene expression there seems to be no preferential crystallization compartment. Produced proteins are guided within the secretory pathway according to their native or artificially added translocation signal.

Do recombinantly introduced proteins in insect cells preferably crystallize in a certain morphology? Needle-like protein crystals are predominantly observed in insect cells so far. The exception is *Nc*HEX-1 which crystallyzes in all tested compartments. The *Nc*HEX-1 crystal morphology depends on the crystallization compartment, and ranges from a hexagonal spindle-shaped- to a pipyramidal morphology. Crystal size is mostly protein sequence- and cell line dependant[71,181]: With outliers, each construct crystallizes up to a typical length and thickness.

To date, *in cellulo* crystallization success, comparable to conventional *in vitro* crystallization without prior knowledge of the protein's phase diagram, remains unpredictable. In additional unpublished work, four biologically relevant proteins of *Plasmodium falciparum (Pf)*, the causative agent of malaria, were channelled through the InCellCryst pipeline[181]. Furthermore, the pipeline was applied to *Bacillus subtilis* lumazinsynthase and *Arabidopsis thaliana* pyrimidindesaminase. Sf9 and High Five cells infected with rBVs at MOI 1 of all these targets. Namely *Pf* 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase, *Pf* IspD; *Pf* 4-diphosphocytidyl-2C-methyl-d-erythritol kinase, *Pf* IspE; *Pf* lyt B transcript, *Pf* IspH; and *Pf* DXP synthase were investigated by light microscopy. Cells infected with P3 rBV *Pf* IspE, *Pf* lyt B and *Pf* 1-deoxy-d-xylulose-5-phosphate synthase, *Pf* DXP synthase, showed sharp-edged structures, in a cell fraction well below one percent. Comparable structures were not observed in mock-infected control cells, which motivated that these targets were tested at the EMBL beamline P14 at DESY. No diffraction was observed during the screening beamtimes. Even after the successful identification of parameters that influence intracellular crystallization - namely primary structure, MOI, cell line and the cellular compartment - there are still very little parameters that can be optimized if no crystals are found.

SX datasets, in contrast to standard cases of single crystal rotational datasets, still require considerable expertise to process. Notably, triggered by community efforts to fight the COVID-19 pandemic, initial processing of serial diffraction data now can be done e.g. at

EMBL beamline P14-2 at DESY. Two diffraction data processing software suites, CrystFEL[36,187–189] and XDS[39,190], were applied. The number of diffraction patterns needed to collect a complete dataset for solving a protein structure strongly depends on a set of parameters: a) the signal-to-noise ratio (SNR) of the individual patterns, b) the symmetry of the crystal, c) steadiness of parameters important for diffraction from shot to shot, and d) systematic errors like an incorrect assumption of detector distances or geometry of the setup.[59]

CrystFEL was written for SFX data. SFX typically requires many thousands of crystals: Due to the *diffraction-before-destruction* principle, only still images are recorded in SFX, as the crystal does not have time to rotate during data collection. Therefore, the detector captures partial reflections, and a high redundancy is needed to capture whole Bragg reflections from multiple frames. Additionally, the crystal orientation is unknown during data collection. Consequently, the number of crystals that need to be measured has to be dramatically increased to (statistically) completely sample reciprocal space. SFX raw data are very large diffraction data sets that require high-throughput software to preprocess the raw images for subsequent analysis,[191] and if necessary for storage reasons, must discard detector frames that do not carry usable information content.

An alternative to diffraction data processing with CrystFEL is the X-ray Detector Software, short XDS.[190,192] In contrast to CrystFEL, developed for 'snapshot crystallography' images as generated with XFELs, XDS is a software suite developed for rotation data. In contrast to mosflm[193,194], XDS is optimized for thin-slized diffraction data. The group of Thomas Schneider at EMBL Hamburg have established scripts to combine small rotational datasets processed in parallel of the serial data collection to a single dataset. Sequentially, *XSCALE* is called for scaling. Data are merged in a separate step using *XDSCONV*. In the presented InCellCryst pipeline, a different custom-made script for XDS, implemented by Juliane Boger of the same working group, is applied for the determination of rotational mini-series with XDS.[181] For both approaches, the first processing step for the data analysis with XDS is performed by a program named *dozor*[195], which is incorporated within the general workflow of data collection at P14 for each X-ray mesh scan. In principle, after the first round of unrefined detection of frames containing 'diffraction' spots – respective, hits – as opposed to empty frames where no diffraction was detected within the background,[195] a diffraction score heatmap is created such as depicted in **Figure 10 B** and **Figure 12 F**. This heatmap is normalized during data collection to the best hit containing most peaks (depicted in white), and should also discriminate between protein diffraction and diffraction of salts (very small unit cell) or ice rings. Ice-rings are typically positioned at 3.89 Å, 3.669 Å, 3.441 Å, 2.671 Å, and 2.249 Å, with no spacing between single "spots", hence the naming ice-rings[196].
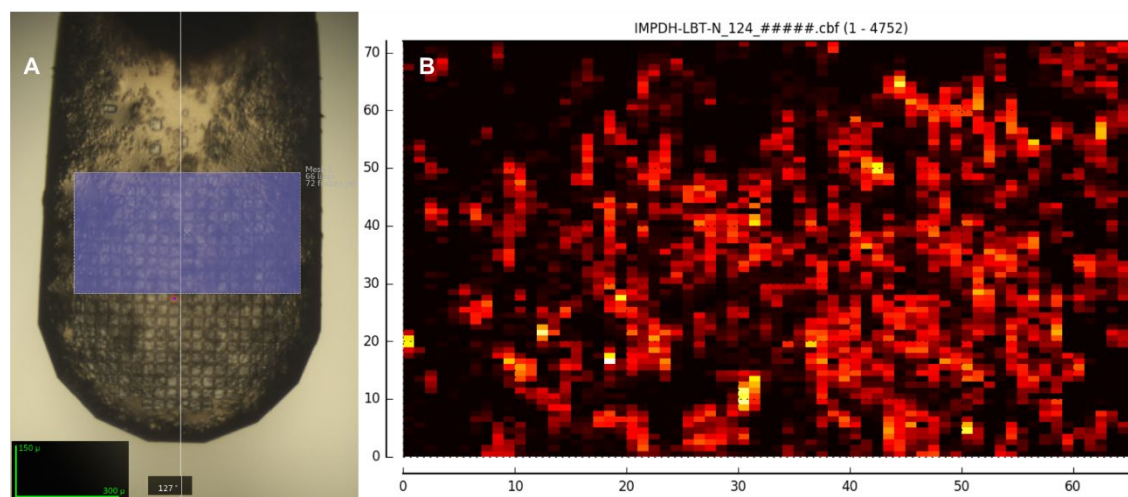
Figure 10 Loaded mesh-grid with intracellular protein crystals and associated heatmap.
A An exemplary loaded mesh-grid of IMPDH N-LBT 4 days *p.i.*. The assigned data-collection mesh-scan is depicted in blue; **B** The two-dimensional heatmap of ranked diffraction intensity created by *dozor* of the raster mesh-scan. In the heatmap, brighter colours signify higher diffraction intensity in the detector frame, respectively higher *dozor* ranking.

The idea behind the *dozor* algorithm is to think of the combined still images from multiple crystals, optimally taken in random orientations, as a (virtual) powder diffraction analysis.[59] Then, a histogram plot of difference diffraction-vector-lengths is created with normalized d-spacing on the y-axis and number of datasets with this d-spacing on the x-axis. This enables a hierarchy of sub-datasets with similar unit-cell parameters that likely belong to the same crystal. This XDS aspect is similar to the function of the *cell-parameter*[38,188] script within the CrystFEL software suite, which however identifies patterns of similar crystals rather than necessarily of the same crystal. The latter option is covered by the CrystFEL associated-script *whirligig*.

A histogram plot of difference diffraction-vector-lengths also enables detection of multi-hits within the same image, and stores positions of spots and their partial intensities.[195] Apart from structural differences, e.g. due to proteins captured in alternate conformations or crystals grown within distinct chemical environments respectively another organelle, also dehydration of crystals or microlesions from the osmotic shock induced by the use of cryo-protectant during sample-handling can influence unit-cell parameters. It is important to raise awareness to these problems before performing the experiment, and again when processing the resulting data later; Fusion of such sub-datasets blurs the electron density map.

In addition to peak detection and enabling sorting of frames into subsets with similar unit cell parameters within a chosen standard deviation, *dozor* creates a Wilson plot, the average variation of Bragg spot intensity as a function of diffraction angle. *Dozor* compares this to the ideal theoretical Wilson plot to distinguish for protein diffraction.[195]

The *resolution* of a structure gives the *d*-spacing that corresponds to the 2σ diffraction angle of the highest resolution shell (d_{min}) which is commonly assumed to represent the resolution limit of a structure. To date, a controversial question of considerable uncertainty and discussion remains the selection of the resolution cutoff of a crystallographic data set.[197] Former guidelines such as the empirical SNR I/sigma(I) > 2 in combination with a completeness in the highest resolution shell > 95 % have become blurry and also unnecessarily strict especially with local background subtraction, but also because refinement programs can handle weak data better. Some advice not to cut any data at all, as the impact on the electron density is limited and valuable information should not be discarded[197–199].

$CC_{1/2}$ is Pearson's correlation coefficient of a dataset randomly split in half:

$$CC = \sum(x - \langle x \rangle)(y - \langle y \rangle)/[\sum((x - \langle x \rangle)^2 \sum(y - \langle y \rangle)^2]^{1/2}$$

$CC^* = \sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$ is derived from $CC_{1/2}$, providing an estimate of the correlation coefficient as would be obtained between the final merged data set and the unknown true values they represent.[200] As a rule of thumb, data are usually statistically insignificant if $CC_{1/2}$ is lower than 0.15. Because it is important for comparability to have consistent thresholds, I have chosen three cut-off critera for serial diffraction data that I processed with CrystFEL:

1. $SNR \geq 0.5$
2. completeness in the highest shell > 95 % and
3. $CC^* \geq 0.5$

For XDS processing, the first criterium (minimal SNR) was sharpened to at least one, since there is no local background subtraction. In the future, additional benefit may be gained from using the program *PAIREF*.[201] This program, a part of the CCP4 Suite since 2021,[202] can determine the best resolution cutoff baised on R-factors after MR, among other parameters similar to SNR, completeness and CC*.

All in all, intracellular crystallization is a promising alternative to classical crystallization approaches, not only because it simplifies procedures by skipping protein purification and artificial enrichment. Major advantages of the BEVS as opposed to expression in bacteria is that as a eukaryotic system it can firstly provide posttranslational modifications, such as protein *N*-glycosylation[203]. This has been applied already in 2013 by Redecke and coworkers in CatB work preceding this thesis.[41] Secondly and maybe even more important, bacteria lack cellular compartments, erasing compartment screening from the limited number of parameters to enhance crystallization success.

A unique selling point for *in cellulo* crystallization, especially when combined with diffraction data collection in the viable cells, is genuine co-factor identification.[134,181]. A huge pool of potential ligands is available from the cellular environment *during in cellulo* crystallization. Unsuspected found ligands should be verified by additional experiments. This, however, may be difficult, because crystal isolation would wash out any ligands with

only moderate affinity. The *Tb*IMPDH structure example demonstrates that *in cellulo* crystallography can identify close-to-native binders in the structure and differentiate between possible candidates much better than soaking experiments *in vitro* with unphysiological ligand concentrations during cocrystallization or soaking could. In 2020, we showed that *in cellulo* crystallized IMPDH N-His (vector map **Supplemental Figure 2**) can be isolated, and solved the structure applying SFX. IMPDH harbors two canonical nucleotide binding sites. ATP and GMP were identified binding to these canonical binding sites. When comparing PDB 6RFU with results of the established *in cellulo* data collection presented in InCellCryst (PDB 8C51, 8C53 and 8CGY), it is now clear that GDP rather than GMP is the native ligand of the second canonical nucleotide binding site in the Bateman domain of *Tb*IMPDH crystallized in High Five cells. Furthermore, that the β-phosphate was hydrolized during the necessary rigorous washing procedures of crystal isolation and purification. In addition, a phosphate is identified bound to the IMP binding pocket. Since otherwise the same overall fold is observed, the differences in the bound ligands can be attributed to the data collection method rather than the omission of artificial N-terminal residues.

Artificial N-terminal residues do not influence the overall fold of *Tb*IMPDH crystals. This was confirmed by SAXS/XRPD of *Tb*IMPDH *in cellulo* crystals. All crystallized constructs shared the same SAXS/XRPD pattern. In conclusion, they crystallize in the same space group, and the indicated artificial residues are not involved in any crystal contacts.

Ligand identification may be challenging. Ligands in the first canonical binding site in the IMPDH cyto and IMPDH ori (PDB codes 8C51 and 8C53) structures were modelled as alternate conformations of ATP. Remarkably, the electron density between the gamma phosphates appears to be continuous. Therefore, it cannot be excluded that in addition to ATP, positively-charged ions like Mg^{2+} or Mn^{2+} coordinate the phosphate groups, partly neutralizing their negative charge of ATP. In fact, neither can it be excluded that there is a covalent connection of the gamma phosphates. A covalent bond here equals a dinucleoside polyphosphate with six phosphate groups, P1-(5'-Adenosyl) P6-(5'-adenosyl) hexaphosphate (AP6A), in the following referred to as the 'AP6A-model' (**Figure 11 A**). AP6A could be bound to some of the proteins in the crystals, while others could harbour the hydrolysed AP6A, respectively one ATP per polypeptide chain (**Figure 11 B**), and some might be coordinated by positively-charged ions.

Diadenosine polyphosphates (ApnA and GpnG) are a family of dinucleotides formed by two adenosines or guanosines joined by a variable number of phosphates. They were discovered in the mid-1960s,[204] and are ubiquitously present in a wide range of prokaryotic and higher eukaryotic organisms[205] including insect cells[206,207]. Already in 1965, it was shown that ApnA and GpnG were minor but surprisingly very stable end products of the ATP and GTP dismutation to pyridine.[205,206,208]
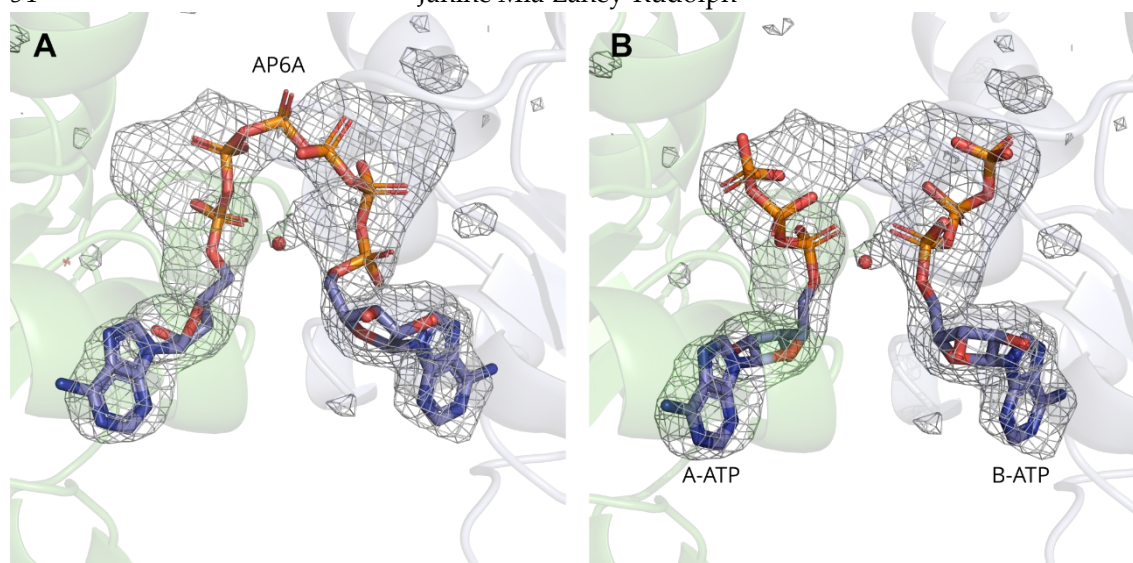
Figure 11: **Canonical nucleotide binding site 1 of the *Tb*IMPDH Bateman domain with alternate ligands**. Omit map contoured at 3 σ. **A**, AP6A-model; **B**, Two ATP are modelled in the nucleotide binding site of both the A and the B-chain, with γ-phosphates repelling each other to a distance of 10.8 Å.

Dinucleoside polyphosphates play a role in a number of cellular processes as signaling messengers, for instance in apoptosis and cell cycle arrest[205,209], calcium signaling[210], as well as DNA replication and repair[211]. Diadenosine polyphosphates with 3 to 5 phosphates have been described more frequently *in vivo*. Interestingly, the dinucleoside polyphosphate AP5G has recently been cocrystallized with the *Ashbya gossipiy* IMPDH (*Ag*IMPDH) and found to bind at the first canonical binding site for allosteric regulation in the Bateman domain. AP6A was shown to bind too, but unfortunately cocrystallization with *Ag*IMPDH was not successful. A number of dinucleotides, including AP6A, have been suggested to play a physiological regulatory role in the IMPDH allosteric switch between an open (active) and closed (inactivated) conformation of the enzyme and were suggested as physiological fine tuners of the allosteric control.[211]

Arguments against the AP6A-model are binding affinity and cellular concentration levels of ATP and AP6A in infected High Five cells. In *Ag*, the affinity of AP6A to the Bateman domain of IMPDH is two magnitudes higher than for ATP.[211] There is no literature available for High Five cells, neither rBV infected nor uninfected, but the ATP concentration can be expected to be around 300 μM[211], which is presumably more than the two orders of magnitudes higher than the native AP6A concentration. This does not prove that AP6A does *not* bind to the first canonical binding sites, and binding affinities were not determined for *Tb*IMPDH, but a low occupancy would be expected. This complicates structure modelling.

The point of this argumentation is that the huge reservoir of potential ligands and co-factors, including relatively unknown biomolecules like diadenosine polyphosphates, may provide unexpected and interesting results, but can also be difficult to interpret in moderate resolution structures.

The InCellCryst pipeline has been significantly optimized to form a more efficient and versatile method for crystallography with intracellular crystals inside insect cells. Data collection *in cellulo* and serial processing fine-tuning that try to model the physics of the diffraction experiment have a significant impact on the quality of the solved structure. For example, after thorough optimization of detector geometry, usage of the partialator offset model for serial synchrotron crystallography (SSX) experiments, and expanding integration of Bragg peaks to higher shells than recognized by peak detection (*push-res* by 2) improved the resolution cutoff on average by 0.1 Å each in IMPDH ori and IMPDH cyto processing. During application of the streamlined InCellCryst process, several problems arised and were solved. Three example challenges are elaborated in the following:

First optimization example: The adhesion of insect cells on the hydrophobic crystallization foil of crystal direct plates proved insufficient for a 90° rotation during mounting on the diffractometer. Both addition of 25% FBS and coating with poly D-lysine increased cell adhesion sufficiently that the majority of infected and crystal-containing cells did not float off the foil during 90° rotated data collection. With the use of a crystallization foil sealing the Crystal Direct plate after removal of medium, dehydration during diffraction data collection did not influence diffraction quality. However, a humidity stream should also prevent dehydration while not increasing the background with an additional foil, but this should be evaluated in future experiments.

Second optimization example: Incidentally, rotational datasets of crystals isolated from Sf9 cells were collected in loops.[31] Changing the fixed-target method to micromeshes allowed blotting away medium which improves the SNR. Data collection with *in cellulo* crystals in viable insect cells was established. Data collection *in cellulo* eliminates purification procedures while the intact surrounding cell protects also fragile crystallites.[212–214] Fishing requires a steady hand and it takes some practice to achieve a regular monolayer of cells with minimal background from medium or cryo protectant, without simultaneously acquiring ice or salt crystal formation. There is a delicate balance between maximising hitrate while limiting increased background due to multiple cells in focus. Often, loading and mounting of cells, as depiced in **Figure 12 B-E**, required much time of the precious beam-time.

This time was shifted to off-beamtime by pre-loading grids and transporting them in liquid nitrogen, which allows robot-assisted mounting with the MARVIN sample changer at the MD3 of the P14 beamline at DESY.

The contact-time with cryo-protectant should be also minimized due to unit cell microlesions that occur during an osmotic shock of the cells in contact with the cryo protectant at ambient temperatures. It takes approximately 30 s from fishing of loaded rBV infected insect cells and passing through PEG400, to entering the nitrogen gas stream (**Figure 12 D-E**).
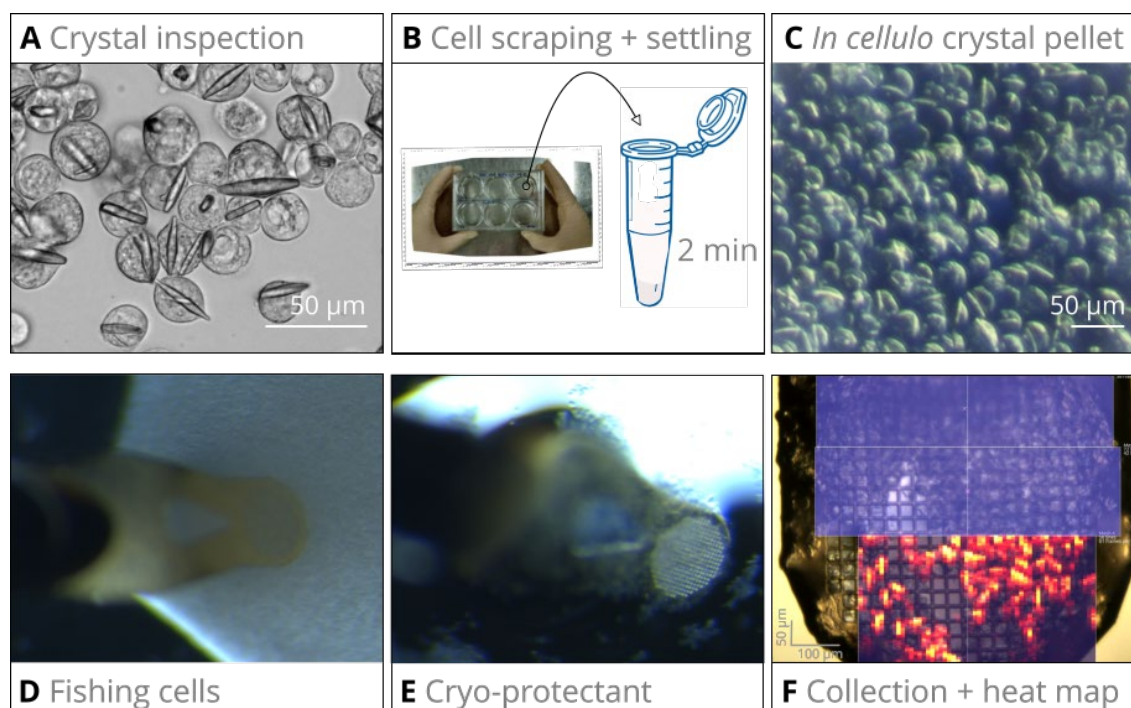
Figure 12 Manual fishing procedure of *in cellulo* crystals using MicroMeshes (MiTeGen) for SSX diffraction data collection at 100 K, and application of a helical scan diffraction data collection strategy. A Visual inspection of the infected and crystal-carrying cell culture stock(s). Manual selection of the best sample, respective best well, e.g. via light microscopy. Here, the two main crystal morphologies of HEX-1 cyto in High Five are apparent: an elongated and a smaller bipyramidal shape. Cell cultures were documented with a Zeiss Observer.Z1 inverted microscope using a 20x objective and an AxioCam MRm microscope camera. **B** Gentle flushing of adherent cells in the selected well and transfer to a 2 mL reaction tube using a 1 mL pipette. Settling of cells via gravitation takes about 2 min. Clear supernatant medium is removed without drying out the cell pellet. **C** Snapshot of a dense crystal-containing insect cell pellet; Here, the HEX-1 cyto crystals developed in rBV infected High Five cells (MOI 1). Three days after virus infection the cells were used for data collection at the EMBL P14 beamline, DESY. Imaged using a reflective-light microscope, 20x objective. **D** Cells are scooped up from the 20 µL cell pellet drop on a glass slide. Excess medium is gently blotted away. **E** The loaded mesh-mount is scooped through 37 % PEG400 for cryo-protection. **F** Diffraction data collection using a helical scan strategy. MicroMesh mounted at the goniometer as seen through the on-axis microscope. The grid in blue is indicating the illumination area and the resulting dozor score heatmap is shown below.

Mounting the MicroMesh to a goniometer base (B1A, MiTeGen) before loading the cell pellet with a pipette and applying cryo-protectant, then plunging the MicroMesh to liquid litrogen for transport to the beamline slightly reduces the contact time with cryoprotectant at RT. Unfortunately, this new procedure also introduced a new problem – salt crystal reflections in the diffraction data, presumably from crystallized medium. When such a peak overlaps with a diffraction peak from protein, the measured intensity is larger than those of symmetry-related reflections, and/or more intense than what is predicted from Wilson statistics[215,216]. These intense reflections are visualizable with the CrystFEL script *peakogram-stream*. These high intensity reflections were removed from stream files by a custom-made script, which is also removing any protein reflections of similar pixel intensity at the same

resolution range. Any salt reflection contaminations well within the intensity of protein diffraction at a given resolution range are not removed with this script. It would be more elegant to determine the symmetry of ice crystals in the dataset and then mask all predicted reflection positions, as very recently established but not yet integrated to the CrystFEL suite by Gevorkov et al.[217]

Third optimization example: Helical line scans during diffraction data collection of needle-shaped crystals with a 0.1° rotation and 10 μm stepsize allowed acquisition of few damage-free sequential images from the same crystal. However, increasing the rotational offset while reducing the translation steps reduced radiation damage, while simultanously allowing more differing consecutional images from one crystal. This is useful for mini-rotational wedge data processing with XDS.

An open question remains whether crystallization of membrane crystals is possible in insect cells. Two dimensional bacteriorhodopsin (bR) crystals have been generated in live *Halobacterium halobium* bacteria and identified by EM and SONICC.[218] A study from 2018 by Casadey and coworkers suggests that meaningful structural information can be retrieved from 2D membrane crystals of bR.[219] The proteins studied here, when not located in the cytaplasm, are membrane surrounded.[181] It seems plausible that two-dimensional crystals of membrane proteins could form. An intracellular layer of membranes would have to fold accordingly to provide the scaffold for membrane protein crystals. Future studies might show whether targeting membrane proteins to the crystae of mitochondria may result in intracellular membrane protein crystals.

How would these supposedly thin 2D crystals be detected from a medium-throughput screening experiment? Superresolution microscopes are seldom available, TEM or SEM techniques require elaborate sample preparations, also the correct layer in the cell containing the crystal would have to be cut. Such crystals might be too small for detection by bright-field microscopy, crystals may grow in a minor percentage of cells, and the morphology be unknown.

3.2 Detection of intracellular protein crystals

Rapid screening of *in cellulo* grown protein crystals via a small-angle X-ray scattering/X-ray powder diffraction synergistic approach

Janine Mia Lahey-Rudolph,[a,b] Robert Schönherr,[a,c] Cy M. Jeffries,[d] Clément E. Blanchet,[d] Juliane Boger,[a] Ana Sofia Ferreira Ramos,[a] Winnie Maria Riekehr,[a] Dimitris-Panagiotis Triandafillidis,[e] Alexandros Valmas,[e] Irene Margiolaki,[e] Dmitri Svergun[d] and Lars Redecke[a,c]*

[a]Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, Lübeck 23562, Germany, [b]Center for Free-Electron Laser Science (CFEL), Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, Hamburg 22607, Germany, [c]Photon Science, Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, Hamburg 22607, Germany, [d]European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Notkestrasse 85, Hamburg 22607, Germany, and [e]Department of Biology, Section of Genetics, Cell Biology and Development, University of Patras, Patras GR-26500, Greece. *Correspondence e-mail: redecke@biochem.uni-luebeck.de

Crystallization of recombinant proteins in living cells is an exciting new approach for structural biology that provides an alternative to the time-consuming optimization of protein purification and extensive crystal screening steps. Exploiting the potential of this approach requires a more detailed understanding of the cellular processes involved and versatile screening strategies for crystals in a cell culture. Particularly if the target protein forms crystalline structures of unknown morphology only in a small fraction of cells, their detection by applying standard visualization techniques can be time consuming and difficult owing to the environmental challenges imposed by the living cells. In this study, a high-brilliance and low-background bioSAXS beamline is employed for rapid and sensitive detection of protein microcrystals grown within insect cells. On the basis of the presence of Bragg peaks in the recorded small-angle X-ray scattering profiles, it is possible to assess within seconds whether a cell culture contains microcrystals, even in a small percentage of cells. Since such information cannot be obtained by other established detection methods in this time frame, this screening approach has the potential to overcome one of the bottlenecks of intracellular crystal detection. Moreover, the association of the Bragg peak positions in the scattering curves with the unit-cell composition of the protein crystals raises the possibility of investigating the impact of environmental conditions on the crystal structure of the intracellular protein crystals. This information provides valuable insights helping to further understand the *in cellulo* crystallization process.

1. Introduction

Nowadays, it is well established that living cells from all kingdoms of life possess an intrinsic ability to form intracellular protein crystals, denoted as 'in vivo grown crystals' or 'in cellulo crystals' (Schönherr *et al.*, 2018). The assembly of intracellular proteins into native crystalline states could provide specific advantages for the organism, mainly in terms of storage and protection. However, this phenomenon also applies to recombinant proteins produced by heterologous gene expression, as highlighted by the growing number of examples predominantly observed in mammalian and baculovirus-infected insect cells. During recent years, novel developments in serial crystallography data collection strategies on X-ray free-electron lasers (XFELs) and synchrotron
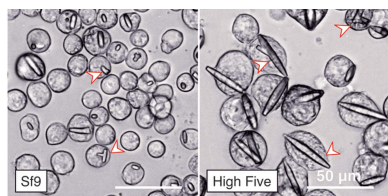
sources (Standfuss & Spence, 2017; Yamamoto *et al.*, 2017; Yabashi & Tanaka, 2017) have paved the way to use *in cellulo* crystals with dimensions in the low micrometre or even the nanometre size range as suitable targets for X-ray crystallography (Gati *et al.*, 2014; Schönherr *et al.*, 2018). High-resolution structural information on several recombinant proteins has already been obtained from diffraction of *in cellulo* crystals, *e.g.* for the coral *Dipsastraea favus* derived fluorescent protein Xpa (Tsutsui *et al.*, 2015), the metazoan-specific human kinase PAK4 in complex with Inka1 (Baskaran *et al.*, 2015) and the BinAB larvicide from *Lysinibacillus sphaericus* (Colletier *et al.*, 2016), as well as of cathepsin B (CatB; Redecke *et al.*, 2013) and IMP dehydrogenase (IMPDH; Nass *et al.*, 2020) from the parasite *Trypanosoma brucei*.

These results question the earlier opinion that the crowded environment in living cells might impact the order of the crystalline structure (Doye & Poon, 2006). Moreover, they indicate that *in cellulo* protein crystallization is able to offer exciting possibilities complementary to conventional crystallization techniques (Chayen & Saridakis, 2008). The approach is particularly important for proteins that were/are not accessible for crystallization using established *in vitro* screening strategies, as shown for *T. brucei* IMPDH (Nass *et al.*, 2020) and fully glycosylated *T. brucei* CatB (Redecke *et al.*, 2013). *In cellulo* crystallization provides an alternative to the time-consuming optimization of protein purification and extensive crystal screening steps. Additionally, the quasi-native conditions in host cells prevent crystal distortion that could arise from non-physiological conditions imposed by re-crystallization and provide the opportunity to identify native co-factors present in the highly versatile natural reservoir of compounds within living cells (Nass *et al.*, 2020). However, exploiting the tremendous potential of *in cellulo* protein crystallization requires a more detailed understanding of the cellular processes involved in crystal formation. Insights into the mechanisms that control the size and shape of crystals, and also the identification of biological parameters suitable for screening approaches, could further widen the applications of *in cellulo* crystallization.

On the basis of a detailed comparison of reported intracellular protein crystallization events, specific requirements have been proposed to favour *in cellulo* crystal growth in fruitful interplay (Koopmann *et al.*, 2012; Schönherr *et al.*, 2015, 2018; Duszenko *et al.*, 2015). This includes the intrinsic crystallization tendency of the target protein under the specific environmental conditions provided by the individual cellular compartments. Moreover, high local protein concentrations seem to be required, which might result from a preceding protein phase separation event (Hasegawa, 2019). In insect cells, crystals occurred in the endoplasmic reticulum (CatB; Koopmann *et al.*, 2012) and in peroxisomes (IMPDH, luciferase; Nass *et al.*, 2020; Schönherr *et al.*, 2015), depending on the native translocation signals harboured in the sequence of the recombinant proteins. Furthermore, a cytosolic localization of crystals was observed [calcineurin, avian reovirus nonstructural protein fused to green fluorescent protein

(GFP-μNS), IMPDH; Fan *et al.*, 1996; Schönherr *et al.*, 2015; Nass *et al.*, 2020]. Thus, different cellular environments may represent the basis for developing a more systematic *in cellulo* crystallization screening approach that would exploit living cells as crystallization factories for a large number of recombinant proteins. An initial strategy to test the crystallization capability of living insect cells has already been proposed and applied to recombinant CPV1 polyhedrin crystals (Boudes *et al.*, 2016, 2017).

The successful detection of protein crystals inside living cells represents a crucial – and somewhat challenging – task in the development of a versatile screening strategy for *in cellulo* crystallization. During recent years a variety of methods have been optimized to identify even nanometre-sized protein crystals in conventional crystallization setups and to locate these crystals after mounting at the beamline (Becker *et al.*, 2017). Unfortunately, the environmental challenges imposed by the living cells largely prevent the direct and efficient detection of *in cellulo* crystals. Most frequently, bright-field microscopy methods including contrast enhancement techniques, *e.g.* differential interference contrast (DIC) or integrated modulation contrast, are applied to visualize the intracellular crystals (Schönherr *et al.*, 2015). The main advantages of these non-invasive methods include frequently accessible equipment, the lack of elaborate sample preparation steps and the good visualization of sufficiently sized crystals. However, the limited resolution of visible-light-based approaches combined with marginal differences in refractive indices makes it difficult to reliably differentiate the ordered crystalline structures in the nanometre size range from the chaotic cellular background. For nanocrystals, transmission electron microscopy (TEM) (Stevenson *et al.*, 2014) was developed into a tool that enables the study and optimization of crystal formation processes *in vitro* (Stevenson *et al.*, 2016) and can be used to characterize *in cellulo* crystals directly within the cellular environment. A resolution in the low nanometre size range allows the visualization of the crystal structure, which can also be applied to identify *in cellulo* crystals (Schönherr *et al.*, 2018). However, since TEM requires ultrathin sectioning (usually <90 nm), a crystal-containing cell has to be selected by chance from the entire population and the crystal must be intersected by the ultrathin cut. If intracellular crystal growth is restricted to a few cells in the entire culture or only very few nanocrystals per cell are produced, this represents a significant limitation, which, together with the time-consuming sample preparation, hampers the simple and rapid detection of crystals in a cell culture. Second harmonic generation (SHG) microscopy used in combination with UV two-photon excited fluorescence, and frequently referred to as second-order nonlinear optical imaging of chiral crystals (SONICC; Kissick *et al.*, 2011; Haupert *et al.*, 2012), represents another emerging technique to rapidly achieve successful crystal formation in conventional screening setups with high sensitivity, selectivity and the potential for automatization (Becker *et al.*, 2017; Tang *et al.*, 2020). However, UV fluorescence is less helpful for intracellular crystals owing to the high protein concentration surrounding the crystal in the cellular environment, and high

crystal symmetry may reduce the crystal-specific SHG signal in practice by about two orders of magnitude (Haupert *et al.*, 2012). Together with the possibility of SHG signal generation by filaments within the cells (Campagnola & Loew, 2003) this could prevent a reliable *in cellulo* crystal detection.

A direct proof for the presence of crystallites is given by the detection of specific Bragg diffraction of electrons or X-rays from a sample. The technique of micro-electron diffraction has the potential to unravel structures of proteins and other biological molecules at 1–3 Å resolution from a few crystals in the nanometre size range, because of the strong interaction between electrons and the crystal. However, ultrathin samples are required, which are frequently obtained by milling (Shi *et al.*, 2013; Jones *et al.*, 2018). X-ray powder diffraction (XRPD) provides a fingerprint of every crystalline phase exhibiting a unique diffraction pattern, and differences between the various crystalline forms can be observed by examining the peak positions and intensities in XRPD patterns (Katrincic *et al.*, 2009). Even small changes in the form of new peaks, additional shoulders or shifts in the peak positions often imply the presence of a second polymorph (Davidovich *et al.*, 2004). Thus, information about crystalline sample composition is obtained, yielding knowledge of whether the sample consists of one or more phases. During the past decade, XRPD has moved beyond fingerprinting of microcrystalline samples by extraction of accurate lattice parameters, elucidating new structural information from biological macromolecules at low and medium resolution (Von Dreele, 2019; Karavassilia & Margiolaki, 2016; Karavassili *et al.*, 2017; Spiliopoulou *et al.*, 2020; Margiolaki, 2019). Densely packed, randomly oriented crystals produce Debye–Scherrer rings on the detector that allow the evaluation of the diffraction capabilities of the sample (Von Dreele *et al.*, 2000; Margiolaki *et al.*, 2007). Even if a relatively small number (<50) of low-angle peaks is considered to be sufficient to precisely refine the unit-cell parameters (Von Dreele, 2019), the volume of the cellular soft matter that surrounds intracellular crystals significantly restricts the crystal density. Thus, the powder diffraction intensity of intracellular crystals at synchrotron crystallography beamlines is often restricted, especially when the crystal-to-cell number ratio in the sample is low (Margiolaki & Wright, 2008).

Small-angle X-ray scattering (SAXS) is performed in solution to structurally characterize biological macromolecules under dilute conditions. SAXS instruments are optimized to minimize the scattering background to detect weak scattering signals that are often orders of magnitude smaller in intensity than diffraction peaks. SAXS profiles provide information on size, shape and oligomerization state but also about interactions between particles in solution. SAXS is extremely sensitive to the formation of crystallites, and this technique has previously been used to analyse protein nucleation (Kovalchuk *et al.*, 2016) and crystallization kinetics (Poplewska *et al.*, 2019). Furthermore, the micro- and nano-GISAX method could even significantly exceed the sensitivity of the SAXS technique for studying protein nucleation (Pechkova & Nicolini, 2017).

In this study, we exploited SAXS and XRPD for a rapid and sensitive detection of protein microcrystals grown within insect cells. We employed the high-brilliance and low-background P12 bioSAXS beamline (Blanchet *et al.*, 2015) of the EMBL at the PETRA III storage ring (DESY, Hamburg). Four test proteins were measured: *Photinus pyralis* luciferas, *T. brucei* IMPDH and CatB, and *Neurospora crassa* HEX-1. Mock-virus-infected and uninfected cells were used as a control. Combining the high sensitivity of SAXS with XRPD analysis methods, we demonstrate that it is possible to assess within seconds whether a cell culture contains microcrystalline material based on the presence of Bragg peaks in the recorded scattering profiles, even for target proteins that form crystals only in a small percentage of cells. This screening approach has the potential to overcome the methodological bottleneck of crystal detection within living cells and opens up opportunities to investigate and understand the influence of growth conditions, stress, temperature, starvation, cellular compartmentalization and the choice of cell line on the size and formation of *in cellulo* crystals.

2. Methods

2.1. Cloning

Cloning procedures for *T. brucei* IMPDH (gene bank accession number M97794) and *T. brucei* CatB (gene bank accession number AY508515) have been described previously (Nass *et al.*, 2020; Koopmann *et al.*, 2012). The genes coding for *P. pyralis* luciferase (Luc, gene bank accession number AB644228) and *N. crassa* HEX-1 (gene bank accession number XM_958614) were amplified by PCR using primers 5′-GAAGACGCCAAAAACATAAAGAA-′3 (sense) and 5-CAATTTGGACTTTCCGCCCTTC-3′ (antisense), and 5′-TACTACGACGACGACGCTCACG-′3 (sense) and 5′-GAG-GCGGGAACCGTGGACG-3′ (antisense), respectively. ALLin HiFi DNA polymerase (highQu) was used according to the manufacturer's instructions. The amplicons were ligated into a modified pFastBac1 vector (Thermo Scientific) containing the sequence 5′-ATGGGCGCCTAA-3′ between the *Bam*HI and *Hind*III restriction sites to accommodate an *Ehe*I restriction site. The vector was linearized using FastDigest *Ehe*I (Thermo Scientific) and blunt-end ligation was achieved using T4 DNA ligase (Thermo Scientific) according to the manufacturer's protocol. Plasmids were transformed into competent *Escherichia coli* DH5α cells (Stratagene) and purified (GeneJET plasmid miniprep kit, Thermo Scientific). The integrity of the cloned sequences was verified by Sanger sequencing. All generated pFastBac1 plasmids were transformed into competent *E. coli* DH10EmBacY cells (Geneva Biotech) according to the manufacturer's instructions. Recombinant bacmid DNA was purified using the GeneJET plasmid miniprep kit (Thermo Scientific) and subsequently used for PCR analysis of the transposed sequence, employing standard pUC M13 forward and reverse primers. For mock-virus generation, bacmid DNA was directly isolated from *E. coli* DH10EmBacY cells without prior transposition of a recombinant gene of interest.

research papers

2.2. Insect cell culture

Sf9 and High Five insect cells were held in suspension culture in serum-free ESF921 insect cell culture medium (Expression Systems) at 300 K on an orbital shaker at 100 r min^{-1}. Suspension culture cells were seeded at 0.5–1 × 10^6 cells ml^{-1}, in a total volume of 25 ml in an upright-standing 75 cm^2 disposable T-flask. Cell density was counted daily and cultures were split when the density reached 4 × 10^6 cells ml^{-1} for High Five or 6 × 10^6 cells ml^{-1} for Sf9 cells.

2.3. Recombinant virus generation

Recombinant bacmid DNA was used for lipofection with Sf9 insect cells grown in ESF921 serum-free medium at 300 K using Escort IV reagent (Sigma–Aldrich) according to the manufacturer's instructions. In brief, 0.45 × 10^6 Sf9 cells per well in a 12-well plate were transfected with 1 µg of bacmid DNA and 3 µl of Escort IV reagent for 18 h. After 4 days of incubation at 300 K the first supernatant (P1) was harvested by centrifugation at 21 000 relative centrifugal force (r.c.f.) for 30 s. For high-titre stock production (third passage, P3), 0.9 × 10^6 Sf9 cells per well in a six-well plate were infected with 100 µl of P1 or 20 µl of P2 viral stock and incubated for 4 days. Viral P2 and P3 stocks were harvested as described above.

2.4. Viral titre determination

A serial dilution assay was used to calculate the titre of the viral P3 stocks. In a 96-well plate, a suspension of 3 × 10^4 High Five cells in 180 µl of antibiotic-free ESF921 insect cell culture medium was added to each well and incubated for 30 min to let cells attach to the bottom. Then, a 1:10 dilution of the virus solution with medium was prepared and 20 µl portions of this solution were added to each of six wells of the first row. For each serial dilution step the medium containing the virus was mixed in the well using a multi pipette and 20 µl of the supernatant was transferred into the next row. Pipette tips were discarded after each row; eight rows were prepared per titration. After 4 days at 300 K, enhanced yellow fluorescent protein (EYFP) fluorescence indicating a successful infection was evaluated, and wells with at least two fluorescent cells were counted as positive. The virus titre was calculated using the TCID$_{50}$ (tissue culture infectious dose; Reed & Muench, 1938).

2.5. Sample preparation for X-ray measurements

In one well of a six-well cell culture plate, 8 × 10^5 Sf9 or High Five cells were plated in 2 ml of ESF921 insect cell culture medium and subsequently infected with P3 stock of the recombinant baculovirus (rBV) using a multiplicity of infection (MOI) of 1. Cells were incubated as a semi-adherent culture at 300 K for 40–96 h until needed for the diffraction experiments. The cells were then gently flushed from the well bottom with a 1000 µl pipette and centrifuged for 30 s at 270 r.c.f., and the cell pellet was resuspended in 25 µl of Tris-buffered saline (TBS; 20 mM Tris, 150 mM NaCl pH 7.0). 40–45 µl of this suspension was transferred into the sample tubes and immediately used for the X-ray scattering experiments.

For dilution series of crystal-carrying cells, High Five insect cells expressing the target gene were mixed with mock-rBV-infected cells in a 1:2 ratio. Up to seven serial dilution steps were carried out directly prior to the X-ray scattering experiments with samples prepared in TBS as previously mentioned.

2.6. Light microscopy

For cell and crystal counting, cell cultures were imaged with a Leica DM IL LED microscope equipped with a 20× objective and a Leica MC170 HD microscope camera prior to the diffraction experiment. The crystal-containing cells and those without crystals were manually counted, and their ratio was calculated. The images of the cell cultures were generated using a Zeiss Observer.Z1 inverted microscope with a 20× objective and an AxioCam MRm microscope camera.

2.7. Propidium iodide staining of infected cells

To visualize the effects of the sample preparation procedure on the cell viability, High Five insect cells were infected as described above for diffraction experiments. Four days after infection, cells were imaged within the wells on a Zeiss Observer.Z1 microscope using differential interference contrast mode and wide field fluorescence. The cells were then gently flushed from the well bottom with a 1000 µl pipette and centrifuged for 30 s at 270 r.c.f., and the cell pellet was resuspended in 25 µl of TBS containing 500 ng ml^{-1} of propidium iodide. Cells were incubated for 10 min at room temperature and then spread on a glass coverslip and imaged again as described above. All samples were prepared in triplets, imaged and manually counted.

2.8. X-ray data collection

Data were collected at the EMBL P12 beamline (PETRA III, DESY, Hamburg, Germany) (Blanchet et al., 2015). A photon energy of 10 keV (1.24 Å) was used throughout the experiments, with a photon flux of about 10^{13} ph s^{-1} at the sample position. Data [$I(s)$ versus s, where $s = 4\pi\sin(\theta)/\lambda$, 2θ is the scattering angle and λ is the X-ray wavelength] were recorded at a sample–detector distance of 3.00 m using a Pilatus 6M detector (setup 1) or a Pilatus 2M detector (setup 2), both from DECTRIS, Switzerland. 40–45 µl of the insect cell suspension prepared as described above was loaded bubble free into the reaction vessels of the SAXS setup, of which 30 µl was transferred into a temperature-controlled 1.8 mm quartz capillary using the automatic bioSAXS sample changer (Arinax) (Round et al., 2015). The high cell density prevented cell settling in the sample tube during the auto-mated loading of up to eight consecutive samples by the sample changer robot.

Using a focal spot of 0.2 × 0.12 mm (FWHM) in a fixed-flow measurement at 293 K, 40 detector frames were recorded per sample separated by 40 buffer frames, all with a single-frame exposure time of 0.045 s and a readout time of 0.005 s, result-ing in a total exposure time of 4 s per data set. For each cell sample, a single data set was collected with the corresponding buffer (TBS), enabling the buffer subtraction during data analysis.

2.9. Data processing

For each sample and corresponding buffer measurement, the 40 individual 2D-detector data frames collected during the course of exposure were summed to produce a final 2D image that was subsequently radially averaged using *im2dat* (Franke *et al.*, 2017) to generate 1D scattering profiles (data deposited with the Small Angle Scattering Biological Data Bank, SASBDB; http://www.sasbdb.org). The data measured from the TBS control were then subtracted, applying the *ATSAS* program suite (Petoukhov *et al.*, 2012;). 1D profile plots were created with *PRIMUS* (Konarev *et al.*, 2003). The data were converted from $I(s)$ to $I(2\theta)$ to facilitate indexing and profile refinement with software packages designed for the analysis of XRPD data as described in the following sections.

2.10. Data clustering and Pawley analysis

Since indexing of acquired data was not feasible owing to the paucity of diffraction peaks, information about data similarities has been evaluated via principal component analysis (PCA) on the $I(2\theta)$ data of all data sets over the 0.4–2.0° 2θ range, using *HighScore Plus* (Degen *et al.*, 2014). This program was also used to extract accurate unit-cell parameters by applying the Pawley approach (Pawley, 1981) for whole powder pattern fitting (WPPF). In the absence of indexing solutions, reasonable starting values for unit-cell parameters were retrieved from relevant PDB entries (Supplementary Table S1). Peak profiles were simulated using a pseudo-Voigt function with the standard description for FWHM and peak asymmetry variation over the 2θ range (Von Dreele, 2019). The background was initially estimated and later it was modelled after a shifted Chebyshev polynomial with varying number of terms (~10–14), depending on the data set, which were refined during Pawley analysis. Parameters were included for refinement of the polynomial background, as well as for instrumental angular offset (zero shift). In the case of highly overlapping reflections, the intensity was equipartitioned to the constituent peaks and gradually refined.

3. Results and discussion

Four test proteins were measured to evaluate the capability of the low-background SAXS beamline P12 for a reliable intracellular crystal detection in living insect cells. Of these proteins, three are known to crystallize in living insect cells infected by rBV, but they differ in crystallization efficiency, as well as in crystal volume and morphology. *T. brucei* IMPDH and CatB have previously been reported to form micrometre-sized needle-shaped crystals in most cells from populations that diffract XFEL pulses and synchrotron radiation to high resolution, enabling the elucidation of the corresponding protein structures (Koopmann *et al.*, 2012; Redecke *et al.*, 2013; Gati *et al.*, 2014; Nass *et al.*, 2020). Needle-shaped *in cellulo* crystals were also observed for firefly (*P. pyralis*) luciferase, growing up to a remarkable length of more than 180 μm, but the spontaneous disintegration after cell

membrane disruption has prevented the validation of X-ray diffraction so far (Schönherr *et al.*, 2015). Additionally, HEX-1, a natively self-assembling protein that forms the solid, crystalline core of Woronin bodies in the fungus *N. crassa* (Tenney *et al.*, 2000), assembles into regular spindle-shaped crystals with a hexagonal cross section in almost all insect cells of the culture, which has not been reported previously.

3.1. Detection of *in cellulo* crystals using SAXS and XRPD

Prior to the diffraction experiment, the previously observed intracellular crystallization tendency of the test proteins in rBV-infected High Five insect cells was verified by light microscopy at day 4 post infection (p.i.). No ordered structures have been detected in the uninfected and in the mock-rBV-infected cells, which served as controls for the subsequent diffraction experiments. The percentage of crystal-containing cells within the entire culture, subsequently denoted as 'crystallization efficiency', was estimated to be around 70–80% for Luc, 40–60% for IMPDH, 50–90% for CatB and more than 90% for HEX-1, slightly varying depending on the individual culture (Fig. 1). Immediately before X-ray experiments at P12, dense cell suspensions were prepared in TBS. At this stage, 60–80% of the rBV-infected cells are still vital in all samples, as confirmed by propidium iodide staining (Fig. 2). Thus, neither virus infection and intracellular crystal growth nor the sample preparation procedures affected the integrity of the predominant fraction of the High Five cells. Moreover, the percentage of crystal-containing cells remained almost constant during sample preparation. Only for luciferase-producing cells was the proportion of crystal-containing cells significantly reduced, from approximately 50 to 30% [Fig. 2(*b*)], which can be attributed to the instability of *in cellulo* grown luciferase crystals outside the living cell (Schönherr *et al.*, 2015).
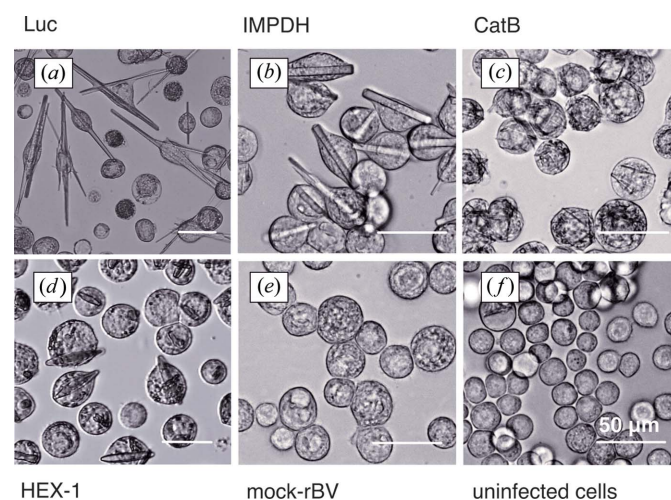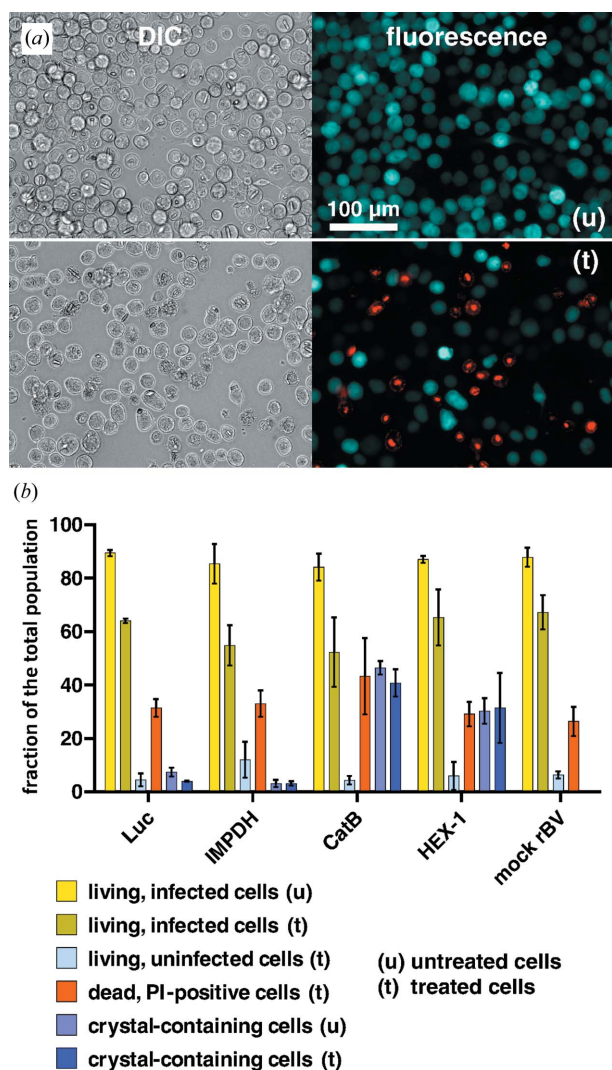

Figure 1
Detection of protein crystals in living insect cells. Light-microscopic images of High Five insect cells 4 days after infection with recombinant baculoviruses (MOI = 1) encoding (*a*) *P. pyralis* Luc, (*b*) *T. brucei* IMPDH, (*c*) *T. brucei* CatB and (*d*) *N. crassa* HEX-1. (*e*) Mock-rBV-infected cells; (*f*) uninfected cells. The intracellular crystals of the individual test proteins differ in size and morphology.

The samples were automatically loaded with a robotic sample changer into the quartz capillary for X-ray diffraction [Fig. 3(a)]. The short exposure time of 0.045 s and readout time of 0.005 s per frame in the steady-state mode resulted in a measurement time per sample of 4 s, since 40 detector frames have individually been recorded for each sample, followed by 40 frames of buffer irradiation. Additionally considering the time required for the automated sample loading and removal as well as capillary cleaning, eight consecutive diffraction data sets were collected within 24 min without opening the hutch, representing the optimal agreement between efficient data collection and settling and survival of the insect cells in TBS. If the previously scored crystal-containing cells had been irradiated, summation of the detector frames consistently revealed the presence of Debye–Scherrer rings [Fig. 3(b)] resulting from the orientational average of the Bragg reflections from the many small crystals randomly oriented in the cell suspension, a typical observation during XRPD

Figure 2
Effects of sample preparation procedures on the viability of infected High Five insect cells. (a) High Five cells imaged 4 days after infection with rBV HEX-1. Upper panel: differential interference contrast light microscopy and fluorescence microscopy of insect cells prior to sample preparation (u, untreated cells). EYFP fluorescence labelling of living, baculovirus-infected cells is shown in cyan. Lower panel: DIC and fluorescence microscopy of cells after sample preparation and propidium iodide (PI) staining (t, treated cells). PI fluorescence labelling of dead cells is shown in red. The scale bar applies to all panels. (b) Analysis of fractions of living, dead, uninfected and crystal-containing cells prior to and after sample preparation procedures within different High Five cell cultures. More than 60% of rBV-infected High Five cells are still vital after sample preparation and thus at the beginning of the diffraction experiment. The reduction of luciferase-crystal-containing cells is due to the instability of luciferase crystals outside the living cell.
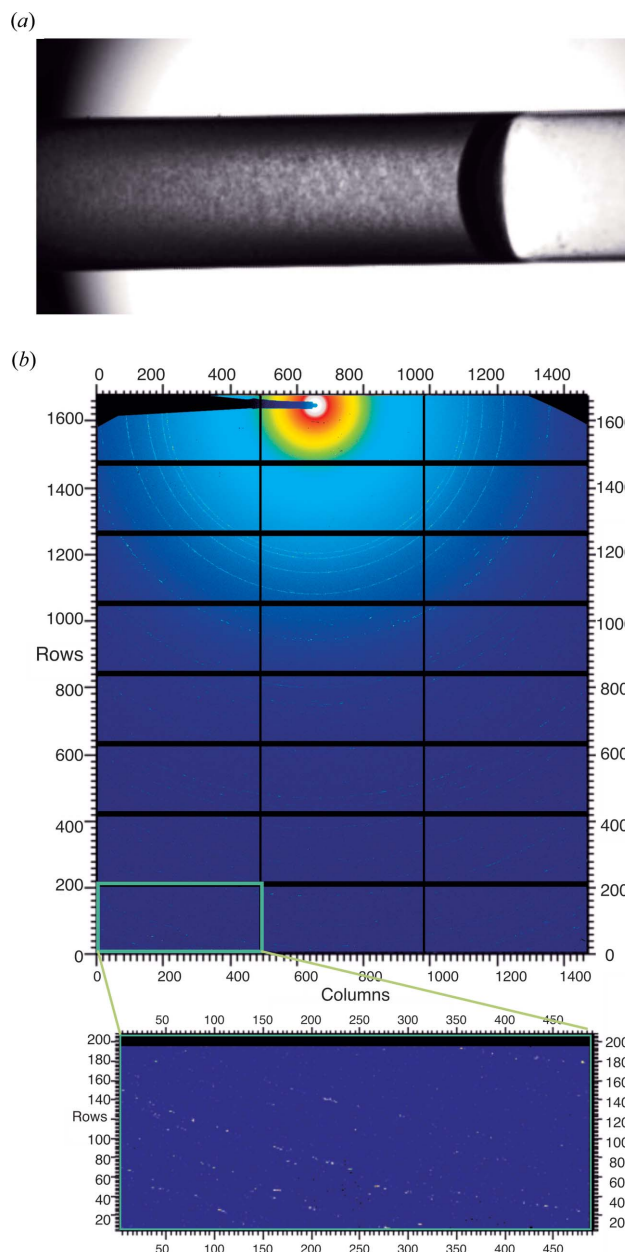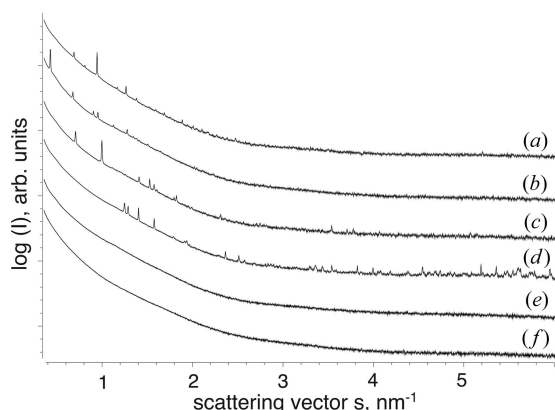

Figure 3
(a) Snapshot of the quartz capillary filled with the dense crystal-containing cell suspension during X-ray diffraction data collection. (b) Summed detector image (40 frames) of HEX-1 crystals irradiated in living High Five insect cells collected using the Pilatus 2M photon-counting detector (DECTRIS) at the EMBL SAXS beamline P12 (PETRA III, DESY, Hamburg, Germany). Granular Debye–Scherrer rings can be detected up to the edge of the detector image (inset). The image was generated using the program *FIT2D* (Hammersley, 2016).

measurements. No rings were observed for the control samples.

Subtraction of the buffer signal and radial averaging resulted in 1D plots representing the intensity versus momentum transfer s (Fig. 4). The corresponding real-space distances are determined as $d = 2\pi/s$. The scattering curves of the crystal-containing cell samples exhibit clear peaks at defined s values, representing the Debeye–Scherrer rings [Fig. 4, curves (a)–(d)]. Depending on its unit cell, each crystal type produced a distinct XRPD profile that can act as a fingerprint of the crystallite.

The intensity of the peaks depends on the overall scattering capability of the irradiated part of the sample. Using an X-ray beam of 0.20×0.12 mm and a 1.8 mm quartz capillary, a volume of 0.043 mm^3 is irradiated, which could incorporate several thousand cells, estimating a diameter of approximately 0.030 mm per cell. Comparable magnitudes of the scattering intensity can be recorded in the case when just a few relatively large crystals are present within the irradiated volume, or when a large number of small crystals are illuminated – it is important that the total number of crystallographic unit cells is above the detection limit defined by the photon flux of the X-ray beam. Thus, the comparatively low crystallization efficiency of IMPDH (40–60%) and Luc (70–80%) is compensated by the significantly increased scattering volume of these long needle-shaped crystals (Fig. 1), resulting in a comparable intensity of the dominant scattering peaks observed for the more abundant but smaller crystals of HEX-1 (>90% efficiency) and CatB (up to 90% efficiency). Consequently, the presence of specific peaks in the scattering curve reliably indicates the presence of crystalline structures with the scattering volume suitable for detection at the given experimental conditions, but the peak intensity on its own does not repre-

sent a suitable measure to compare the number and/or size of different crystallites in the living cells (Fig. 3).

3.2. Extraction of refined unit-cell parameters

The low-angle region of XRPD data usually allows for a precise refinement of the unit-cell parameters of the diffracting crystals, if pure and highly dense microcrystalline suspensions are used in conventional powder diffraction experiments (Margiolaki, 2019). In our samples, a significant volume is occupied by the solvent and the soft matter of the cells, limiting the accessible crystal density and thus the intensity of the Bragg scattering patterns. Only a few significant peaks at low s values can be detected in the scattering curves of intracellular Luc, IMPDH, CatB and HEX-1 crystals (Fig. 4), preventing *ab initio* indexing.

It has been demonstrated in earlier studies (Norrman *et al.*, 2006; Fili *et al.*, 2015; Valmas *et al.*, 2015) that information about data similarities can be evaluated via PCA. PCA reduces the dimensionality of data sets by projecting them to distinct principal component (PC) axes, which are planes in the multidimensional space (Hotelling, 1933). By definition, the first PC is the plane where data exhibit the largest variance when projected along it. Subsequent PCs must be orthogonal to the first one. Once the required number of PCs is identified (typically two or three), data are projected into a new coordinate system defined by these PCs. The position of each observation in the PC coordinate system and its distance to other observations is indicative of the similarities between the observations. Analysis performed on the $I(2\theta)$ data over the $0.4–2.0°$ 2θ range produced four distinct clusters for the samples under study, each containing one of the four different phases observed in our experiments (Supplementary Fig. S1). Clustering not only allowed us to detect the existence of four well separated crystalline phases in our data (marked A–D in Supplementary Fig. S1), even before their identification, but also enhanced the rapidity of the analysis.

Even when only a few peaks are present, accurate unit-cell parameters can be extracted from XRPD data sets using WPPF procedures (Karavassilia & Margiolaki, 2016; Margiolaki, 2019). On the basis of the starting lattice parameters, Pawley analysis (Pawley, 1981) theoretically simulates the experimental profiles in terms of peak shapes and background and, most importantly, allows for their refinement. Here, a structural model is not required, since peak intensities are considered as refinable parameters, contrary to Rietveld refinement (Rietveld, 1969). Using the reported unit-cell dimensions and space groups determined by X-ray crystallographic structure elucidation of *T. brucei* IMPDH and CatB (using *in cellulo* grown crystals), as well as of *P. pyralis* Luc and *N. crassa* HEX-1 (using crystals grown by microbatch and vapour diffusion techniques *in vitro*), as reasonable starting values (Supplementary Table S1), accurate lattice parameters were extracted for each data set (Fig. 5). A complete list of the refined reflections and their position in 2θ, d spacing and momentum transfer is presented in Supplementary Tables S2–S5.

For the Luc data set, the refined lattice parameters revealed a significant increase in the length of the unit-cell axes a and b

Figure 4
1D radially averaged X-ray scattering data of the insect cells containing intracellular crystals of the target proteins (a) Luc (SASBDB: SASDHY5), (b) IMPDH (SASBDB: SASDHZ5), (c) CatB (SASBDB: SASDH26) and (d) HEX-1 (SASBDB: SASDH36). (e) Mock-rBV-infected cells (SASBDB: SASDH46); (f) uninfected cells (SASBDB: SASDH56). If crystalline structures are present within the cells, distinct Bragg diffraction peaks are detected in the scattering curves at scattering vectors characteristic for the unit-cell dimensions of the protein crystals. All scattering curves were acquired with detector setup 1.

by approximately 10 Å, compared with the expected values extracted from PDB entry 1lci (Conti *et al.*, 1996) (Supplementary Table S1). The unit cell of Luc *in cellulo* crystals has not been determined so far, but these differences indicate that the intracellular crystal growth affects the unit-cell geometry of Luc crystals. On the other hand, the detection of specific Bragg reflections from the intracellular Luc structures represents the first proof of the crystalline character. This result confirms our hypothesis that the intact cell protects the crystals from deterioration induced by environmental changes, *e.g.* during cell lysis and crystal isolation (Schönherr *et al.*, 2015).

For the other three data sets, Pawley analysis resulted in reasonable agreement of the refined and the expected unit-cell parameters (Supplementary Table S1). At least for IMPDH and CatB, this was expected, since the starting parameters have been obtained from the corresponding X-ray structures elucidated using these *in cellulo* grown crystals [IMPDH, PDB code 6rfu (Nass *et al.*, 2020); CatB, PDB 4hwy/4n4z (Redecke *et al.*, 2013; Gati *et al.*, 2014)]. However, the intracellular

environment obviously did not change the unit-cell geometry of the HEX-1 crystals, as shown by the agreement with the parameters of crystals grown by applying the sitting drop vapour diffusion method (PDB code 1khi; Yuan *et al.*, 2003).

### 3.3. Sensitivity of *in cellulo* crystal detection

One of the major obstacles in intracellular protein crystallization is the observation that the proportion of crystal-containing cells within the entire culture can be very low. By applying light microscopy, one sometimes detects well ordered structures of a recombinant target protein in only 1% (or even less) of cells, rendering the proof of successful *in cellulo* crystallization a laborious and time-consuming effort. We have therefore further assessed the sensitivity of the scattering-based detection method by irradiation of a dilution series of High Five insect cells containing intracellular crystals of CatB and HEX-1. Starting from 100% infected cells, infected cells were diluted in a 1:2 ratio with mock-virus-infected cells. The intensity of the distinct diffraction peaks in the scattering
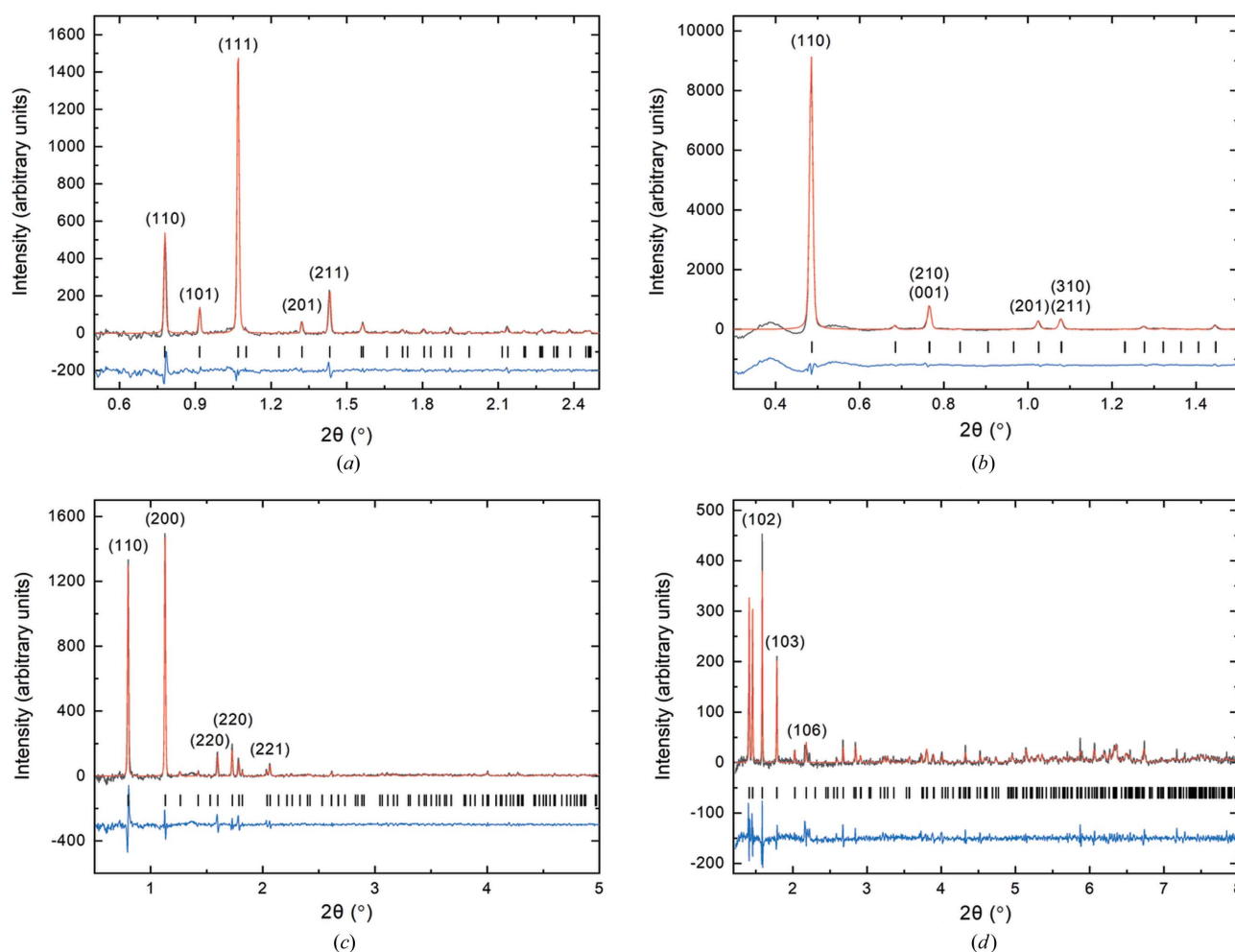
Figure 5
Pawley fits of XRPD data of High Five insect cells containing intracellular crystals of the target proteins (*a*) Luc (SASBDB: SASDHY5), (*b*) IMPDH (SASBDB: SASDHZ5, (*c*) CatB (SASBDB: SASDH26) and (*d*) HEX-1 (SASBDB: SASDH36). Background has been subtracted for clarity. The black, red and blue lines represent the experimental data, the calculated profile, and the difference between experimental and calculated patterns, respectively. A few of the major peaks have been annotated. The vertical black lines correspond to Bragg reflections compatible with the respective refined unit cells: (*a*) $P4_12_12$, $a = b = 129.13 (6)$ Å, $c = 97.1 (1)$ Å, $R_{wp} = 0.80$, $\chi^2 = 1.76$, (*b*) $P42_12$, $a = b = 209.3 (1)$ Å, $c = 93.44 (2)$ Å, $R_{wp} = 0.90$, $\chi^2 = 3.79$, (*c*) $P4_22_12$, $a = b = 125.69 (1)$ Å, $c = 54.408 (7)$ Å, $R_{wp} = 1.57$, $\chi^2 = 1.73$, (*d*) $P6_522$, $a = b = 58.01 (2)$ Å, $c = 195.2 (7)$ Å, $R_{wp} = 2.63$, $\chi^2 = 1.81$.

curves consistently drops with each dilution step owing to the reduced number of crystals in the irradiated sample volume (Fig. 6). However, the overall course of the scattering curve was not affected. At a dilution of 16-fold, corresponding to 0.34 and 5.68% of cells in the sample that contain *in cellulo* CatB and HEX-1 crystals, respectively, even the originally most intense peaks can barely be distinguished from the background scattering from the cell suspensions. Progressive dilution yields scattering curves superimposable to that of the mock-virus-infected cells, defining the detection limit of the crystalline material in the irradiated volume at the specific conditions defined by this experimental setup. Considering the uncertainties in the determination of the detection limit, *e.g.* a slight volume increase of the insect cells after baculovirus infection (Schopf *et al.*, 1990) and individually varying cell sizes, this scattering approach enables the rapid detection of
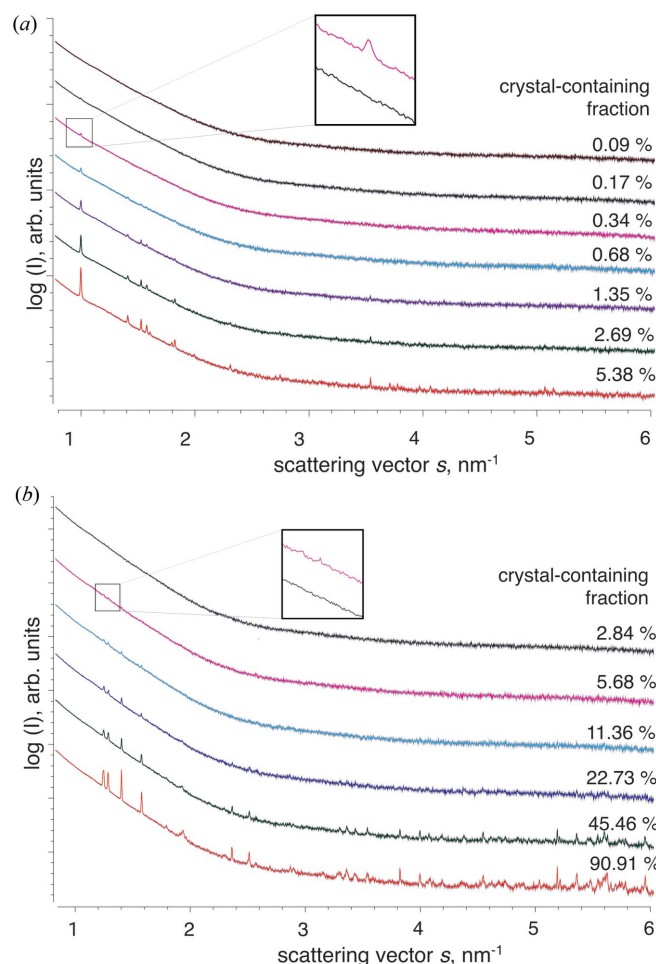
Figure 6
1D radially averaged X-ray scattering data of High Five insect cells containing intracellular crystals of the target proteins CatB (*a*) and HEX-1 (*b*) (SASBDB IDs: SASDH76 and SASDH66). The percentage of crystal-containing cells within the entire culture of each sample, as determined by light microscopy, is presented next to the scattering curves. The detection limit for *in cellulo* crystals using X-ray scattering at the P12 beamline setup 1 was determined to range between 0.3 and 6% of a crystal-containing cell fraction, depending on the respective protein. The insets show the scattering curve of the 16-fold diluted sample compared with that of mock-infected cells.

intracellular crystals of CatB and HEX-1 if present in at least 0.3–6% of all cells in the culture, depending on the individual protein. A comparable detection limit was determined for IMPDH *in cellulo* crystals in High Five cells (Supplementary Fig. S2).

3.4. Impact of the insect cell line

It was previously reported that the crystallization efficiency of recombinant target proteins in living insect cells varies depending on the individual cell line (Fan *et al.*, 1996). In High Five cell cultures, a larger proportion of cells produced intracellular crystals of the heterodimeric calcineurin complex, compared with Sf9 cell cultures. Our study clearly confirms this correlation. A significant drop in crystallization efficiency, ranging between 45 and 84%, was observed after infection of Sf9 cells with the same MOI of recombinant rBV stocks encoding CatB and HEX-1 (Fig. 7). Expectedly, the reduced crystalline scattering volume of the Sf9 cell samples leads to a decreased intensity of the distinct Bragg peaks in the scattering curves [Figs. 7(*c*) and 7(*f*)]. The peak positions, however, which directly depend on the symmetry and the unit-cell parameters of the irradiated crystals, did not change. Next to the important proof of the presence of crystalline material, the peak fingerprint obtained from the scattering data represents a precise and highly sensitive marker for the crystal architecture, which is not affected by the insect cell line, at least for the IMPDH, CatB and HEX-1 (Supplementary Fig. S2) proteins analysed in this study. This marker is much more reliable than the visual inspection of the crystals by light microscopy, which basically confirmed the needle-shaped tetragonal morphology of the IMPDH and CatB crystals and the elongated spindle-shaped hexagonal morphology of HEX-1 crystals, if grown in Sf9 cells (Fig. 7).

3.5. Timeline of intracellular crystal growth

The timing of the X-ray measurements represents another parameter that essentially affects a reliable scoring of an *in cellulo* crystallization experiment. In the applied baculovirus expression vector system (BEVS; Smith *et al.*, 1983) recombinant target gene expression is controlled by the *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) polyhedrin promotor. Owing to its activation late in the infection cycle (Chambers *et al.*, 2018), target protein production starts approximately 24 h after rBV infection of the insect cells. First indications of intracellular crystal formation can be detected by light microscopy at least 72 h (3 days) p.i., as previously shown by real-time investigation of the spontaneous crystallization processes of *P. pyralis* Luc and GFP-μNS from avian reovirus (Schönherr *et al.*, 2015), as well as of *T. brucei* IMPDH (Nass *et al.*, 2020). Crystal growth usually continued up to day 5 p.i., when the majority of cells started to gradually lyse, triggered by the ongoing viral proliferation process. The associated environmental change can significantly affect the integrity and thus the X-ray diffraction capacity of *in cellulo* crystals (Schönherr *et al.*, 2015), defining the optimal time slot for intracellular crystal detection as between 24 and 120 h p.i.

However, intracellular Luc crystals showed an unexpected dynamic degradation and reassembly within the same living cell over the entire growth period (Schönherr *et al.*, 2015), which turns the definition of the optimal time for detection into a more complicated task.

Samples with different offsets between the insect cell infection and the X-ray diffraction experiment have been prepared to monitor the time-dependent powder diffraction of High Five cells infected with rBVs encoding all four test proteins used in this study. On the basis of the results mentioned above, offsets ranging between 40 and 94.5 h were tested. In the scattering curves of cells producing Luc, CatB and HEX-1, Bragg diffraction peaks clearly distinguishable from the background scattering of the cells consistently appeared at approximately 51 h p.i. (Table 1 and Supplementary Fig. S4). Subsequently, the peak intensities increased up to approximately 81 h p.i. and remained constant. The Bragg peak intensities of IMPDH-producing cells exhibited a comparable trend, but the onset of crystal detection was delayed by 10 h, starting approximately at 61 h p.i. However, after 64 h p.i. the ratio of quantity, volume and intrinsic order of the crystalline material formed by all test proteins in the infected insect cells was consistently sufficient for detectable Bragg scattering, even if the single parameters were strongly dependent on the individual protein crystallization process. Our data indicate that the high brilliance and low background

Table 1
Timeline of intracellular crystal growth in High Five cells.

X-ray diffraction experiments using setup 2 have been performed at indicated time-points after insect cell infection with rBVs encoding Luc, IMPDH, CatB and HEX-1. The earliest time-point where a Bragg diffraction peak was detected is indicated with a plus sign (+). Arrows illustrate changes in the intensity of the Bragg peaks compared with the previous time-point. The consistent decrease of the signal intensity at the 64 h time-point is most likely attributable to a cell culture problem rather than to an effect of the intracellular crystallization process.

Protein	40 h	51.5 h	57.5 h	61.25 h	64 h	72.5 h	81.5 h	94.5 h
Luc	−	+	↗	↗	↘	↗	↗	→
IMPDH	−	−	−	+	↘	↗	↗	→
CatB	−	+	↗	↗	↘	↗	↗	→
HEX-1	−	+	→	→	↘	↗	↗	→

afforded by the SAXS instrument setup (*e.g.* all-in-vacuum beam path) enables a reliable scoring of *in cellulo* crystallization trials.

Obtaining insights into the kinetics of intracellular protein crystallization represents another reason to monitor the time dependence of *in cellulo* crystal growth. The associated molecular mechanisms are difficult to determine in the cellular context, preventing a comprehensive understanding so far. Initial insights have been obtained by live cell imaging techniques, but only after the size of the tracked crystal exceeded the detection limit of DIC light or fluorescence microscopy
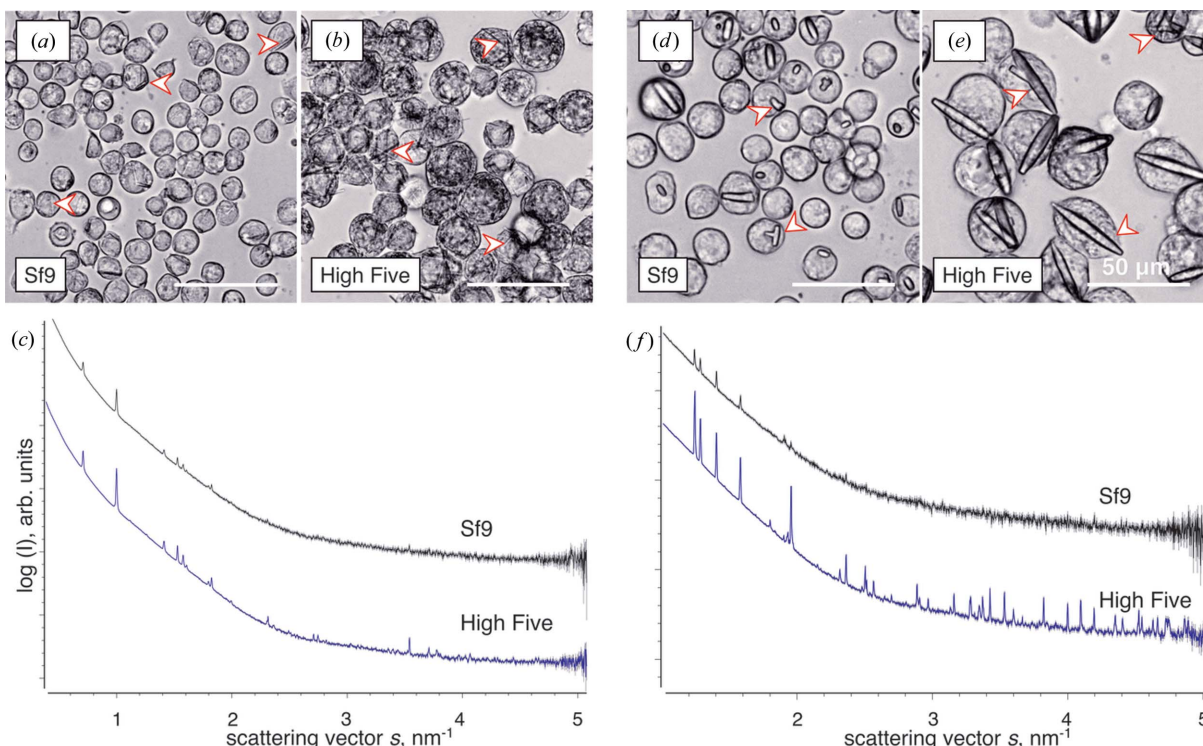
Figure 7
The unit-cell parameters of intracellular crystals do not depend on the insect cell line used for protein crystallization. Differential interference contrast light microscopy of intracellular CatB (*a*), (*b*) and HEX-1 (*d*), (*e*) crystals grown in Sf9 insect cells shows a highly comparable morphology compared with the growth in High Five cells. Red arrowheads highlight selected intracellular crystals. Corresponding X-ray scattering data show identical positions of the Bragg diffraction peaks in both cell lines for either CatB (SASBDB: SASDH96, SASDH86) (*c*) or HEX-1 (SASBDB: SASDHB6, SASDHA6) (*f*), indicating identical unit-cell parameters independent of the cell line used for crystal growth. The peak intensity is reduced in Sf9 cells owing to the observable drop in the crystallization efficiency. The standard deviation of each data point is presented as grey bars.

(Schönherr *et al.*, 2015), which is far beyond the nucleation event and the initial growth phase. Microscopy-based techniques are particularly problematic for proteins that exhibit a low *in cellulo* crystallization efficiency. These techniques focus on a few individual cells in the culture that have been selected by chance without guarantee that crystals will form inside these cells. Kinetic analysis by SONICC strongly depends on the orientation and symmetry of the growing crystals, which affects the signal intensity (Haupert *et al.*, 2012) and thus also the crystal detection. However, the cell selection problem is overcome by monitoring a large and representative fraction of all cells in the culture at the same time, as performed in X-ray powder-diffraction-based approaches. On the other hand, probing all cells at the same time without spatial resolution prevents the elucidation of the growth kinetics of a single crystal, since the total crystalline volume hit by the comparatively large X-ray beam contributes to the diffraction signal. Consequently, the low-background diffraction approach using high-brilliance X-ray beams will not provide more detailed insights into the crystallization process of individual crystals. However, it is able to provide information on the timing of the formation of detectable crystalline structures and a good estimate for overall crystal production inside the living cells, which is important to choose the optimal time point of further diffraction data collection at a synchrotron or XFEL for elucidation of the respective protein structure.

4. Conclusion

Detection of intracellular crystals in cell cultures can be a time-consuming and challenging task, particularly if the target protein forms crystalline structures of unknown morphology only in a small fraction of cells. Furthermore, light-microscopy-based detection of well ordered structures yields a promising indication, but not a proof, of crystallinity. The presented SAXS–XRPD screening approach has the potential to overcome these major limitations of *in cellulo* crystallization. Owing to the automated robot-assisted sample handling, the flow-through setup, the short irradiation time and an exceptionally low background scattering of the SAXS beamline setup, this approach allows one within seconds to prove if diffracting crystalline structures of any order and morphology are present in at least a low percentage of cells within a culture. Such information cannot be obtained by other established detection methods in this time frame. Applying light microscopy, a comparable result would usually require several hours of tedious screening. Since the intensity of the X-rays determines the minimum diffractive volume that is required for reliable detection, a further increase in peak brilliance will allow the detection of smaller crystals or even a smaller percentage of crystal-containing cells, *e.g.* using fourth-generation synchrotrons or XFELs in the future. High-throughput SAXS–XRPD screening of potentially crystal-containing samples can be directly linked to subsequent serial diffraction data collection at a macromolecular crystal-lography beamline to streamline the structure determination.

Moreover, since the Bragg peak positions in the 1D scattering curves depend on the unit-cell composition of the protein crystals, this approach also provides the possibility to investigate the impact of environmental conditions, *e.g.* the cellular compartment, cellular stress or the cell line itself, on the size and the composition of the intracellular protein crystals. This information could contribute to more detailed insights into the understanding of the *in cellulo* crystallization process.

References

Baskaran, Y., Ang, K. C., Anekal, P. V., Chan, W. L., Grimes, J. M., Manser, E. & Robinson, R. C. (2015). *Nat. Commun.* **6**, 8681.

Becker, M., Kissick, D. J. & Ogata, C. M. (2017). *Methods Mol. Biol.* **1607**, 143–164.

Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M. & Svergun, D. I. (2015). *J. Appl. Cryst.* **48**, 431–443.

Boudes, M., Garriga, D. & Coulibaly, F. (2017). *J. Vis. Exp.* **21**, e55793.

Boudes, M., Garriga, D., Fryga, A., Caradoc-Davies, T. & Coulibaly, F. (2016). *Acta Cryst.* D**72**, 576–585.

Campagnola, P. J. & Loew, L. M. (2003). *Nat. Biotechnol.* **21**, 1356–1360.

Chambers, A. C., Aksular, M., Graves, L. P., Irons, S. L., Possee, R. D. & King, L. A. (2018). *Curr. Protoc. Protein Sci.* **91**, 5.

Chayen, N. E. & Saridakis, E. (2008). *Nat. Methods*, **5**, 147–153.

Colletier, J.-P., Sawaya, M. R., Gingery, M., Rodriguez, J. A., Cascio, D., Brewster, A. S., Michels-Clark, T., Hice, R. H., Coquelle, N., Boutet, S., Williams, G. J., Messerschmidt, M., DePonte, D. P., Sierra, R. G., Laksmono, H., Koglin, J. E., Hunter, M. S., Park, H.-W., Uervirojnangkoorn, M., Bideshi, D. K., Brunger, A. T., Federici, B. A., Sauter, N. K. & Eisenberg, D. S. (2016). *Nature*, **539**, 43–47.

Conti, E., Franks, N. P. & Brick, P. (1996). *Structure*, **4**, 287–298.

Davidovich, M., Gougoutas, J. Z., Scaringe, R. P., Vitez, I. & Yin, S. (2004). *Am. Pharm. Rev.* **7**, 10–17.

Degen, T., Sadki, M., Bron, E., König, U. & Nénert, G. (2014). *Powder Diffr.* **29**, S13–S18.

Doye, J. P. K. & Poon, W. C. K. (2006). *Curr. Opin. Colloid Interface Sci.* **11**, 40–46.

Duszenko, M., Redecke, L., Mudogo, C. N., Sommer, B. P., Mogk, S., Oberthuer, D. & Betzel, C. (2015). *Acta Cryst.* F**71**, 929–937.

Fan, G. Y., Maldonado, F., Zhang, Y., Kincaid, R., Ellisman, M. H. & Gastinel, L. N. (1996). *Microsc. Res. Tech.* **34**, 77–86.

Fili, S., Valmas, A., Norrman, M., Schluckebier, G., Beckers, D., Degen, T., Wright, J., Fitch, A., Gozzo, F., Giannopoulou, A. E., Karavassili, F. & Margiolaki, I. (2015). *IUCrJ*, **2**, 534–544.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M. & Svergun, D. I. (2017). *J. Appl. Cryst.* **50**, 1212–1225.

Gati, C., Bourenkov, G., Klinge, M., Rehders, D., Stellato, F., Oberthür, D., Yefanov, O., Sommer, B. P., Mogk, S., Duszenko, M., Betzel, C., Schneider, T. R., Chapman, H. N. & Redecke, L. (2014). *IUCrJ*, **1**, 87–94.

Hammersley, A. P. (2016). *J. Appl. Cryst.* **49**, 646–652.

Hasegawa, H. (2019). *Exp. Cell Res.* **379**, 92–109.

Haupert, L. M., DeWalt, E. L. & Simpson, G. J. (2012). *Acta Cryst.* D**68**, 1513–1521.

Hotelling, H. (1933). *J. Educ. Psychol.* **24**, 498–520.

Jones, C. G., Martynowycz, M. W., Hattne, J., Fulton, T. J., Stoltz, B. M., Rodriguez, J. A., Nelson, H. M. & Gonen, T. (2018). *ACS Cent. Sci.* **4**, 1587–1592.

Karavassili, F., Valmas, A., Fili, S., Georgiou, C. & Margiolaki, I. (2017). *Biomolecules*, **7**, 63.

Karavassilia, F. & Margiolaki, I. (2016). *Protein Pept. Lett.* **23**, 232–241.

Katrincic, L. M., Sun, Y. T., Carlton, R. A., Diederich, A. M., Mueller, R. L. & Vogt, F. G. (2009). *Int. J. Pharm.* **366**, 1–13.

Kissick, D. J., Wanapun, D. & Simpson, G. J. (2011). *Annu. Rev. Anal. Chem.* **4**, 419–437.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). *J. Appl. Cryst.* **36**, 1277–1282.

Koopmann, R., Cupelli, K., Redecke, L., Nass, K., DePonte, D. P., White, T. A., Stellato, F., Rehders, D., Liang, M., Andreasson, J., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Boutet, S., Bozek, J. D., Caleman, C., Coppola, N., Davidsson, J., Doak, R. B., Ekeberg, T., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Hunter, M. S., Kassemeyer, S., Kirian, R. A., Lomb, L., Maia, F. R. N. C., Kimmel, N., Martin, A. V., Messerschmidt, M., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R. G., Soltau, H., Stern, S., Strüder, L., Timneanu, N., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Williams, G. J., Wunderer, C. B., Fromme, P., Spence, J. C. H., Stehle, T., Chapman, H. N., Betzel, C. & Duszenko, M. (2012). *Nat. Methods*, **9**, 259–262.

Kovalchuk, M. V., Blagov, A. E., Dyakova, Y. A., Gruzinov, A. Y., Marchenkova, M. A., Peters, G. S., Pisarevsky, Y. V., Timofeev, V. I. & Volkov, V. V. (2016). *Cryst. Growth Des.* **16**, 1792–1797.

Margiolaki, I. (2019). *International Tables for Crystallography*, Vol. H, *Powder Diffraction*, edited by C. Gilmore, J. A. Kaduk & H. U. Schenk, pp. 718–736. Chichester: Wiley.

Margiolaki, I. & Wright, J. P. (2008). *Acta Cryst.* A**64**, 169–180.

Margiolaki, I., Wright, J. P., Fitch, A. N., Fox, G. C., Labrador, A., Von Dreele, R. B., Miura, K., Gozzo, F., Schiltz, M., Besnard, C., Camus, F., Pattison, P., Beckers, D. & Degen, T. (2007). *Z. Kristallogr. Suppl.* **26**, 1–13.

Nass, K., Redecke, L., Perbandt, M., Yefanov, O., Klinge, M., Koopmann, R., Stellato, F., Gabdulkhakov, A., Schönherr, R., Rehders, D., Lahey-Rudolph, J. M., Aquila, A., Barty, A., Basu, S., Doak, R. B., Duden, R., Frank, M., Fromme, R., Kassemeyer, S., Katona, G., Kirian, R., Liu, H., Majoul, I., Martin-Garcia, J. M., Messerschmidt, M., Shoeman, R. L., Weierstall, U., Westenhoff, S., White, T. A., Williams, G. J., Yoon, C. H., Zatsepin, N., Fromme, P., Duszenko, M., Chapman, H. N. & Betzel, C. (2020). *Nat. Commun.* **11**, 620.

Norrman, M., Ståhl, K., Schluckebier, G. & Al-Karadaghi, S. (2006). *J. Appl. Cryst.* **39**, 391–400.

Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.

Pechkova, E. & Nicolini, C. (2017). *Nat. Protoc.* **12**, 2570–2589.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.

Poplewska, I., Łyskowski, A., Kołodziej, M., Szałański, P., Piątkowski, W. & Antos, D. (2019). *Chem. Eng. Res. Des.* **141**, 580–591.

Redecke, L., Nass, K., DePonte, D. P., White, T. A., Rehders, D., Barty, A., Stellato, F., Liang, M., Barends, T. R. M., Boutet, S., Williams, G. J., Messerschmidt, M., Seibert, M. M., Aquila, A., Arnlund, D., Bajt, S., Barth, T., Bogan, M. J., Caleman, C., Chao, T.-C., Doak, R. B., Fleckenstein, H., Frank, M., Fromme, R., Galli, L., Grotjohann, I., Hunter, M. S., Johansson, L. C., Kassemeyer, S., Katona, G., Kirian, R. A., Koopmann, R., Kupitz, C., Lomb, L., Martin, A. V., Mogk, S., Neutze, R., Shoeman, R. L., Steinbrener, J., Timneanu, N., Wang, D., Weierstall, U., Zatsepin, N. A., Spence, J. C. H., Fromme, P., Schlichting, I., Duszenko, M., Betzel, C. & Chapman, H. N. (2013). *Science*, **339**, 227–230.

Reed, L. J. & Muench, H. (1938). *Am. J. Epidemiol.* **27**, 493–497.

Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.

Round, A., Felisaz, F., Fodinger, L., Gobbo, A., Huet, J., Villard, C., Blanchet, C. E., Pernot, P., McSweeney, S., Roessle, M., Svergun, D. I. & Cipriani, F. (2015). *Acta Cryst.* D**71**, 67–75.

Schönherr, R., Klinge, M., Rudolph, J. M., Fita, K., Rehders, D., Lübber, F., Schneegans, S., Majoul, I. V., Duszenko, M., Betzel, C., Brandariz-Nuñez, A., Martinez-Costas, J., Duden, R. & Redecke, L. (2015). *Struct. Dyn.* **2**, 041712.

Schönherr, R., Rudolph, J. M. & Redecke, L. (2018). *Biol. Chem.* **399**, 751–772.

Schopf, B., Howaldt, M. W. & Bailey, J. E. (1990). *J. Biotechnol.* **15**, 169–185.

Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *eLife*, **2**, e01345.

Smith, G. E., Summers, M. D. & Fraser, M. J. (1983). *Mol. Cell. Biol.* **3**, 2156–2165.

Spiliopoulou, M., Valmas, A., Triandafillidis, D.-P., Kosinas, C., Fitch, A., Karavassili, F. & Margiolaki, I. (2020). *Crystals*, **10**, 54.

Standfuss, J. & Spence, J. (2017). *IUCrJ*, **4**, 100–101.

Stevenson, H. P., Lin, G., Barnes, C. O., Sutkeviciute, I., Krzysiak, T., Weiss, S. C., Reynolds, S., Wu, Y., Nagarajan, V., Makhov, A. M., Lawrence, R., Lamm, E., Clark, L., Gardella, T. J., Hogue, B. G., Ogata, C. M., Ahn, J., Gronenborn, A. M., Conway, J. F., Vilardaga, J.-P., Cohen, A. E. & Calero, G. (2016). *Acta Cryst.* D**72**, 603–615.

Stevenson, H. P., Makhov, A. M., Calero, M., Edwards, A. L., Zeldin, O. B., Mathews, I. I., Lin, G., Barnes, C. O., Santamaria, H., Ross, T. M., Soltis, S. M., Khosla, C., Nagarajan, V., Conway, J. F., Cohen, A. E. & Calero, G. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 8470–8475.

Tang, Y., Saul, J., Nagaratnam, N., Martin-Garcia, J. M., Fromme, P., Qiu, J. & LaBaer, J. (2020). *Sci Rep.* **10**, 13323.

Tenney, K., Hunt, I., Sweigard, J., Pounder, J. I., McClain, C., Bowman, E. J. & Bowman, B. J. (2000). *Fungal Genet. Biol.* **31**, 205–217.

Tsutsui, H., Jinno, Y., Shoda, K., Tomita, A., Matsuda, M., Yamashita, E., Katayama, H., Nakagawa, A. & Miyawaki, A. (2015). *Mol. Cell*, **58**, 186–193.

Valmas, A., Magiouf, K., Fili, S., Norrman, M., Schluckebier, G., Beckers, D., Degen, T., Wright, J., Fitch, A., Gozzo, F., Giannopoulou, A. E., Karavassili, F. & Margiolaki, I. (2015). *Acta Cryst.* D**71**, 819–828.

Von Dreele, R. B. (2019). *International Tables for Crystallography*, Vol. H, *Powder Diffraction*, edited by C. Gilmore, J. A. Kaduk & H. U. Schenk, pp. 263–269. Chichester: Wiley.

Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. (2000). *Acta Cryst.* D**56**, 1549–1553.

Yabashi, M. & Tanaka, H. (2017). *Nat. Photon.* **11**, 12–14.

Yamamoto, M., Hirata, K., Yamashita, K., Hasegawa, K., Ueno, G., Ago, H. & Kumasaka, T. (2017). *IUCrJ*, **4**, 529–539.

Yuan, P., Jedd, G., Kumaran, D., Swaminathan, S., Shio, H., Hewitt, D., Chua, N.-H. & Swaminathan, K. (2003). *Nat. Struct. Mol. Biol.* **10**, 264–270.

3.2.1 Synopsis

Crystal detection in cells is one of the bottlenecks of *in cellulo* crystallography. They can be difficult to identify by light microscopy (LM) techniques, owing to crystallite sizes between nanometers to several hundred micrometers, among other reasons.[71] Finding the cells carrying intracellular protein crystals in a given cell culture is even more challenging for targets whose nucleation events and subsequent crystal growth of unknown morphology occur only in a low fraction of cells. The ultimate proof for the presence of crystalline states in a sample is, of course, the detection of Bragg diffraction after irradiation, e.g. the detection of X-ray powder diffraction patterns. A rapid screening approach of insect cell cultures for *in cellulo* crystals has been developed and published 2020 in the Journal of Applied Crystallography. The investigated model systems were *in cellulo* crystals of firefly luciferase, IMPDH N-His, CatB of the pathogen *T. brucei*, and fungal HEX-1 cyto. Experiments were controlled with mock-virus and uninfected infected Sf9 and High Five cells. This small angle X-ray/ X-ray powder diffraction (SAXS-XRPD) synergistic approach is based on the sensitive detection of Debye-Scherrer rings from a crystal-containing insect cell culture at a low-background SAXS beamline. This method can be used to reliably detect protein crystals, also if they are only present in a minor percentage of cells, and this within seconds. The sensitivity of the method for CatB, HEX-1 cyto and IMPDH N-His was measured as followed: First, the crystallization efficiency (percentage of crystal-containing cells in a cell culture batch) was determined for each batch by manual counting using bright-field microscopy, this ranged between 40 to 90 %. Then, immediately before X-ray irradiation, dense cell suspensions were prepared in TBS buffer and diluted with mock-infected cells up to 32-fold. 1D radially averaged X-ray scattering data of the irradiated High Five cells for each dilution were compared to mock-infected High Five cells. Depending on the target protein, crystal-containing fractions between 0.34 % and 5.68 % were still clearly identifieable.

In additional (unpublished) experiments, both High Five and Sf9 cells were diluted with uninfected insect cells of the same cell line, otherwise the same procedure was followed. **Table 3** gives an overview of bright-field LM-observed percentages of cells with at least one crystal and the associated minimal crystal concentration required to confirm crystallinity in the SAXS/XRPD experiment. Interestingly, there is no clear trend for whether the minimal crystal concentration in High Five cells or in Sf9 cells is higher. Considering the larger crystal dimensions crystals grow to in the former, resulting in higher diffraction volumes per crystal/ cell-carrying crystal(s), the starting hypothesis was that High Five cells would have a lower detection limit.

In total, the minimal crystal concentration that allows crystal detection determined this way ranges between <0.1 and 1.5 %, which is even lower than determined in the published manuscript.

However, this is artificially low, since the virus infection (even without a target protein sequence after the polyhedrin promotor like for mock rBV) increases the cell diameter and cell volume[220,221]. For High Five cells, the cell biovolume increases on average 1.9-fold due to the rBV infection.[221]

Table 3 Initial crystallization efficiency and determined detection limit of *in cellulo* crystals in High Five cells and Sf9 cells via dilution series with uninfected insect cells.
The detection limit was determined via a SAXS-XRPD synergistic approach[182] in dilution series with uninfected High Five cells or respectively Sf9 cells in (1:2) cell number quantity steps. The proportion of cells containing crystals was manually determined using four light microscopy images with at least 400 counted cells. C-HA/ N-HA – hemagglutinin tag at the C-or the N-terminus; n/d – no data; N-LBT – N-terminal lanthanide binding tag.

	% of cells with protein crystals	Minimal crystal-containing cell concentration for detection
High Five cells		
CatB	92.45 %	0.73 %
HEX-1 cyto	91.20 %	0.72 %
HEX-1 C-HA	88.00 %	1.38 %
IMPDH cyto	60.59 %	0.95 %
IMPDH N-LBT	35.02 %	0.55 %
IMPDH N-HA	75.42 %	0.58 %
Sf9 cells		
CatB	8.41 %	0.26 %
HEX-1 cyto	23.03 %	1.44 %
HEX-1 C-HA	23.51 %	n/d
IMPDH cyto	15.65 %	0.49 %
IMPDH N-LBT	1.1 %	n/d
IMPDH N-HA	6.13 %	0.095 %

Small changes in a XRPD pattern like new peaks, shifted peak positions or additional shoulders often imply that an additional polymorph is present[222] - a useful feature of the SAXS/XRPD approach if the impact of varied screening parameters is to be analysed. Medium- or high-throupput screening of samples under different conditions allows a direct observation of even minor structure modifications and the exploitation of radiation- and sample-induced effects.[223,224] Dataset similarities were investigates by principal component analysis with the program *HighScore Plus*[225] and support by the group of Irene Margiolaki (University of Patras, Greece) on all $I(2\theta)$ data ranging from 0.2 to 2.0°.
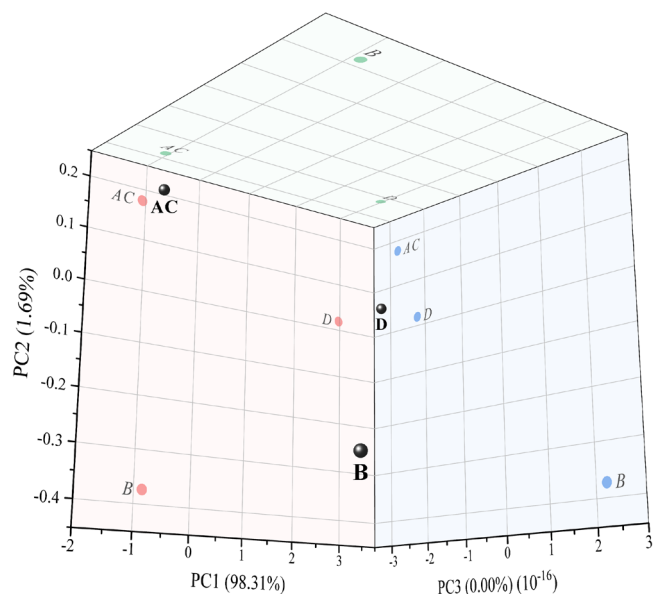
Figure 13 Principal component analysis of SAXS/XRPD data shows different clusters in all three principal component planes. Datasets with subtracted background were used, in the range of $0.4 - 2.0°\ 2\theta$. The clustering results are shown in the three main principal component (PC) axes, as well as their projections to the three orthogonal planes. All datasets are far apart in all three PC axes, thus indicating significant differences between them. Datasets A and C are clustered together, which could be explained by strong peaks at similar positions. Results were presented in the JAC (2020) publication. A – luciferase, B – IMPDH N-His, C – CatB, D – HEX-1 cyto.

The Pawley approach that was applied to *HighScore Plus* takes background estimates into consideration - delivering exact unit cell parameters of the intracellular crystals, and crystal symmetry. The method was used to study the dynamics of protein crystal growth and decline within a cell batch. First indications of intracellular crystal formation can be detected by bright-field LM at least three days *p.i.,* much after first nucleation events and the initial growth phase*.* Using the SAXS/XRPD approach, Bragg diffraction peaks evolve from the background scattering of the cells at ~ 51 h after infection, with the exception of cell cultures infected with rBV IMPDH N-His. Here, first diffraction signals appear ten hours later. As a batch method, this SAXS/XRPD approach cannot give single-crystal or single-cell kinetics. On the bright side, the approach enables the elucidation of the optimal diffraction data collection time point for SX, at 81 h after infection, from where on there is no further signal intensity growth.

All *in cellulo* powder diffraction experiments were performed at RT. An osmotic shock induced by cryo-protectant would broaden Bragg-peaks at cryogenic temperature, due to micro distortions in the unit cell, which broadens Debye-Scherrer rings and therefore would lead to a higher uncertainty of determined d-spacings.[226] The downside of RT diffraction experiments is more radiation damage induced by the same dose. Therefore, less photons can be used for irradiating the same sample volume and the signal can only be detected to lower resolution as compared to measurements that would be performed at cryo conditions. This limits the minial crystal volume in the measured unit.

A low crystallisation efficiency rate between 0.2 and 6 % of cells[182] is a serious disadvantage for every imaging method applied, in particular for small crystallites. This can be partly compensated for by enriching the proportion of cells carrying protein crystals before imaging[111], e.g. indirectly by FACS gating for a high YFP expression which is also under the polyhedrin promotor control.[181] Potentially, artificial intellicence (AI) could help overcome this crystallization efficiency bottleneck for imaging techniques, when combined with automated high-throughput image data collection. If properly implemented, AI classification on microscopy images[227] could narrow down the number of images that are potentially crystal containing for manual inspection by the trained eye, effectively accelerating sample screening for intracellular crystal growth in favourable cases.

Densely packed, randomly oriented crystals loaded into a borosilicate or Kapton capillary, or, for very low amount of material, into a double silicon-nitride window[228], produce intense Debye-Scherrer rings on the detector that allow the evaluation of the diffraction capabilities of the sample.[228–230] Therefore, the SAXS/XRPD *in cellulo* approach could be modified to SAXS/XRPD analysis of isolated *in cellulo* crystals. Isolation and purification would allow for a denser crystal packing, yielding crystal symmetry, precise lattice parameters ideally without prior information, and in cases of high-quality data medium resolution structures. MR has worked in many protein cases with powder data of conventionally produced crystalline powder.[223,230,231] Structure refinements are now a routine with algorithms like Rietfield refinements, which can compensate for overlapping peaks, a main problem in powder diffraction data. However, in these cases SX remains the superior technique to solve the 3D structure for crystals inside living cells.

Assuming a low crystallization efficiency, optimized data collection strategies are important. Crystal volumes that are insufficient for diffraction data collection using synchrotron radiation due to lack of diffraction volume, can be considered for studies of fast dynamics (ps or fs) or of particular biological relevance can be considered for SFX using XFEL.

3.3 **Data collection at Free-Electron Lasers** *in cellulo*

To circumvent potentially harmful isolation of the crystals from the cells, data collection strategies were implemented for *in cellulo* data collection in this thesis. This is possible with synchrotron radiation - implemented procedures are presented in the InCellCryst publication (chapter 0) - and it is suitable for XFEL radiation, as presented in the following.

What are the most important factors for a successful SFX experiment, in particular with regard to protein crystals grown in living cells?

1. Good sample quality, meaning well-ordered, homogenous and sufficiently large crystals, is the everything of any X-ray scattering experiment.[200] Unfortunately, the cellular mechanisms of intracellular protein crystallization are too poorly understood to enforce excellent diffraction properties of crystals grown in living cells by controlling certain experimental conditions. Diminishing *in cellulo* crystal-quality, on the other hand, is easy. Since isolation from the cellular environment can reduce crystal quality, as demonstrated for polyhedra by Boudes and coworkers in 2017[232], data collection still within intact organelles of the cell or in whole viable cells is desirable[125,126,181,233]. This holds also true for XFEL experiments[75] and motivates the establishment of *in cellulo* SFX data collection for insect cells.

2. Since the intense XFEL pulse of about 10^{12} photons in tens of femtoseconds will inevitably result in the destruction of any exposed protein crystal, serial data collection is mandatory.[56,234] With a narrow bandwidth and short X-ray irradiation that is preventing crystal rotation during the exposure, mostly partial reflections are recorded. The still reflection patterns from several tens to several hundreds of thousands of crystals must be merged to derive a complete set of structure factors usable for structure determination. [56,187,235] Recent enhancements to the basic Monte Carlo integration method enable the retrieval of more accurate data from fewer diffraction patterns by modelling the physics of the underlying experiment as best as possible.[236] It still follows that a large number of reasonably similar protein crystals of the target protein must be available for a successful SFX experiment.

   Insect cells are well suited as protein crystallization factories for intracellular crystals in the nm- to μm-range. Upscaling sample production from millions of infected cells (ml scale) to trillions of cells (L scale) was established and achieved within a few months. If the infrastructure and the necessary cell culture and rBV quantities are available, upscaling for *in cellulo* SFX can be done within one infection round in only four days time, and with minimal volume-due increase in preparation time. However, upscaling is redundant if a crystal delivery method like fixed-target SFX data collection in living cells is available.

3. Destruction of the crystal after only one shot – resulting in random or rather unknown orientation of the vaporized crystals – is a major reason, why a high multiplicity enhances structure quality solved from the SX experiment.[44,200,237] In fact, a random orientation of crystals during data collection would statistically ensure most efficient sampling of the reciprocal space, whereas preferred orientation of crystals during data collection, as resulting e.g. by needle-like crystals in a liquid stream, may prove a challenge.[183]

In the 2021 IUCrJ publication, it could be shown that with data collection in cells, preferred orientation is not a problem.[183] It can be presumed that the roundish cell adhers in random orientation on the micro-patterned silicone chips used for FT-SFX, uninfluenced by the orientation of protein crystal(s) within the cell.

4. Ideally, irradiated crystals should be naked to achieve the best possible SNR.[238–240] Why should SFX data collection *in cellulo* then be aspired? In practice, a minimal hydration shell is necessary to prevent diffraction quality loss due to dehydration.[241] Intracellular crystals stay hydrated for a longer period of time than isolated *in cellulo* crystals due to the protective shell.[100] Gentle changes or a conservative environment reduces chances of crystal lattice disruptions and thereby enhances the achievable resolution.[71,75,109,111,112]

An advantage of the utilized micropatterned silicon chips[239,242] for the FT-SFX *in cellulo* approach is the simple removal of mother liqor/medium that results in low background scattering during the diffraction experiment while still benefiting from dehydration-preventing cellular surroundings.

5. Time is money - both working lifetime and beamtime are valuable goods. Where possible whithout compromising on output quality, since XFEL beamtime is exceptionally rare and precious, experimental effort should be minimalized. This may be trivial, but is a clear incentive for SFX in viable crystal-containing cells, as *in cellulo* diffraction data collection diminishes the need for protein and crystal isolation and purification[71,109,126,183,232,233].

6. In the ideal SFX experiment, every pulse hits one crystal. Assuming a crystal slurry with a Gaussean distribution of crystal concentration sometimes multiple patterns would be produced in the same diffraction image, this could result in indexing rates approaching 100 % and more indexed patterns than detector frames. There are challenges: for example, the EuXFEL pulse pattern with a current maximal frequency of 4.5 MHz (220 ns) per pulse within a pulse train is up to $1 \cdot 10^5$ times faster than the Roadrunner II fast-scanning goniometer can currently manage with the FT-SFX used micro-patterned silicon chips. This calculation is considering a translation speed of up to 100 mm/s and 100 µm spacings between sample-containable pores[182,240].

Concerning maximal achievable hitrates, SFX *in cellulo* is in disadvantage compared to the ideal isolated, purified, robust crystal batch. The crystallization efficiency is a bottleneck to achieveable hitrates during a diffraction experiment with viable insect cells. The hitrate for diffraction data collection *in cellulo* is also largely dependant on the crystallization efficiency, the percentage of cells carrying at least one crystal of the target protein, which differs considerably between targets.[182] Increasing crystallization efficiency is possible[243] by optimizing expression system, MOI or screening cell compartments, but to date only up to a target-dependant limit.[181] Also, the position inside the cell not necessarily overlaps with the focused X-ray beam. As long as it is not possible to gently extract cells with crystals from a given culture, the achievable crystal density during the diffraction experiment remains limited. Without such a filtering-method, most efficient data collection methods are crucial to minimize the needed (XFEL) beamtime per solved structure.[244,245]

7. SFX requires a reliable sample delivery method that brings many protein microcrystals with equal properties one after the other to the focus point of the XFEL beam. There are two principal approaches for SFX sample delivery currently available: Either, the sample can be streamed across the beam using a special injector device (like a liquid jet, **Figure 14 A**), or, the sample can be fixed on a solid support, which is then rastered by the beam (fixed target, FT, **Figure 14 C**). Hybrid methods like a tapedrive[44,246,247] or a crystal extractor have also been developed (**Figure 14 B**).[244]
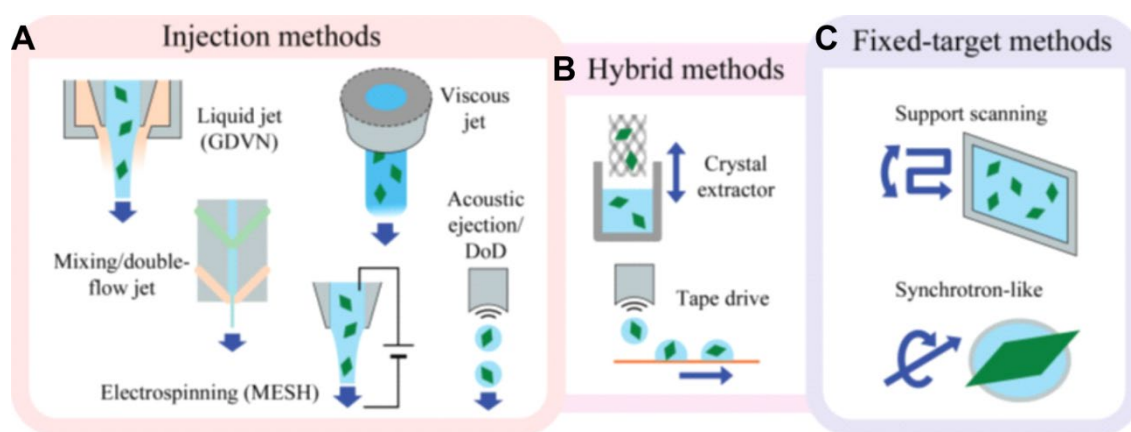
Figure 14: **Sample delivery methods for SFX.**
A Injection methods, **B** hybrid methods and **C** fixed-target methods. Reproduced from Martiel, Müller-Werkmeister and Cohen (2019)[244].

To suit the properties of the different sample types like size, stability or viscosity of mother liquor, and make best use of the experimental setups, many method variations have been developed and published in the last years, as reviewed e.g. by Grünbein and Nass Kovacs in 2018.[234] Both general strategies (liquid jet, chapter 3.3.1 and a published FT-SFX approach, chapter 3.3.2) have been tested for delivery of intracellular protein crystals into XFEL beams within the scope of this thesis.

3.3.1 SFX experiments with *in cellulo* crystals using liquid jets

The European XFEL (EuXFEL) started early user operation in September 2017. Samples of intracellular protein crystals, produced in the scope of this thesis and embedded in the group proposal *XFEL2012* entitled 'SFX at MHz using a liquid jet setup' were contributed to the first experiments at the SPB/SFX hutch. The Nct ECD SKL was selected as a target for high-resolution structure elucidation by MHz SFX (vector map **Supplemental Figure 4**).

A high-resolution structure of Nct is of high medicinal scientific relevance, and the diffraction volume of the needle-shaped crystals obtainable in insect cells is too small for using 3rd generation synchrotron radiation. Nct ECD SKL crystals that formed in rBV-infected High Five cells 3 to 6 days *p.i.* are also too small to be detected in intact insect cells with available LM techniques[248]. From a previous test diffraction experiment at the LCLS CXI beamline it was known that such crystals, if crudely isolated from Sf9 cells,[248] diffracted to up to 2.3 Å. Due to impurities this only allowed a hit-rate of approximately 1 %. Sample preparation of intracellular Nct ECD SKL crystals for this first EuXFEL beamtime with a sample delivery by liquid jet required a significant upscaling of insect cell culture[249]; since the experiment was perfomed when safety regulations at the EuXFEL did not allow S1 experiments yet, isolation of protein crystals from insect cells was required.

Therefore, the first objective was to establish diffraction data collection on isolated intracellular crystals as groundwork for *in cellulo*-data collection using liquid jets. This required production of a crystal pellet of 1 mL volume with an ideal crystal density of 1×10^{11} crystals per mL in order to match the one MHz repetition rate (at the time) of the XFEL pulse. Assuming that every cell contains one Nct ECD SKL crystal, starting with 1×10^{11} High Five cells, and assuming loss-free isolation, 50 L ESF921 medium with 2×10^6 insect cells per mL were needed to upscale cell culture accordingly for rBV infection and subsequent isolation crystals from the insect cells.

Imaging and jetting of crystals isolated from rBV infected insect cells

To confirm the presence of Nct ECD SKL crystals, cell isolates were characterized by scanning electron microscopy (SEM) at CFEL, with the help of Dr. Mauro Prasciolu. Due to time limitations before the start of the EuXFEL beamtime, samples were not spladdered with gold. Visible charging effects appear in effect. Needle-sharped structures with clear sharp edges can be identified in the SEM images (**Figure 15 A+B**). PBS in the isolate also crystallized upon drying. The structures with approximate sizes of $1.9 \times 1.9 \times 5.5 \ \mu m^3$ (**Figure 15 C**) potentially represent the crystalline target protein.

These structures show crystal fractures (**Figure 15 B**) from repeated shearing stress during pipetting and high-speed centrifugation to separate cell debris from protein crystals.
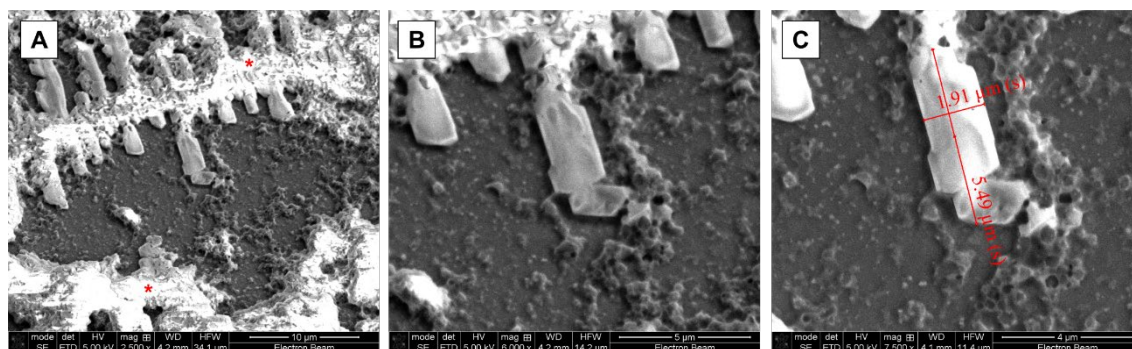
Figure 15 Isolated Nct ECD SKL isolates 5 days after infection of High Five cells. Isolated crystal fragments in PBS, dried and imaged with SEM, 5 kV high voltage electron energy on a Zeiss EVO MA25, with $1x\,10^5$ mbar. **A** Dried isolates contain elongated crystalline structures that are arranged orderly and are covered by a crystalline salt mesh from dried solvent. Charging effects are marked with asterisk. **B** Close-up of a relevant region with a damaged crystalline sample. **C** Size estimation of individual broken sharp-edged pieces of elongated size sticking together.

The principle of a gas dynamic virtual nozzle (GDVN) was chosen for initial SFX experiments performed at the European XFEL (EuXFEL)[250,251]: a liquid crystal slurry stream is accelerated to high speed and focussed by a sourrounding helium gas. The sample gets pushed from the inside of a sample reservoir to injector lines through a syringe by a high-pressure liquid chromatography pump (Shimadzu) driven plunger. The exact GDVN settings can be found in the group publication from 2018.[252] Isolated but hydrated Nct ECD SKL crystals were used for high-speed GDVN liquid jet tests at CFEL, in cooperation with Max Wiedorn. A stable jet for 30 min was achieved with 300 pounds per square inch, using an isolate concentration of 35 % in PBS. In oder to prevent nozzle clogging, all samples were filtered (Valco) using pore sizes down to two µm directly before experiments.

During the first BAG beamtime at the EuXFEL, the prepared Nct ECD SKL isolated structures were not subjected to the FEL beam. Due to technical limitations only a comparatively large beam focus of around 10 x 10 µm² was available, instead of the previously proposed 1 x 1 µm² focus. Thus, larger lysozyme crystals provided by other collaborators have been more suitable and were used for diffraction data collection to perform the required validation of the SFX technique at the EuXFEL.

Jetting tests of crystals in paraformaldehyde fixated insect cells

A compromise to test the injection of crystal-containing cells in the context of a subsequent recommissioning EuXFEL beamtime following the valid safety regulations was the use of inactivated virus, 'dead' but intact cells with intracellular crystals. Human Nct ECD SKL *in cellulo* crystals were again selected as "high-impact" sample, and additionally, *Nc*HEX-1 as a well-behaving *in cellulo* crystallized model protein. Since a trend towards a higher crystallization efficiency in rBV infected adherent cell culture was observed, cells were grown and rBV-infected in cell discs (Bio One), with 4000 cm² per disc harboring $5x\,10^8$ cells. The developed procedure is described in the methods section.

Virus was inactivated after infection and incubation by fixating crystal-containing cells with the cross-linking agent paraformaldehyde (PFA), which covalently links the protein (crystals) and any other cellular component while largely preserving the structure and integrity of the treated insect cells.

The possible influence of PFA fixation on crystal morphology was investigated. Isolated and fixated (PFA concentration of 0.01 %) samples from High Five cells infected with rBV Nct ECD SKL were subsequently desalted and imaged by SEM at the Institute of Anatomy at the University of Lübeck (**Figure 16**). This was unnecessary for *Nc*HEX-1 crystals, as these are readily observable by LM and appear largely intact (**Figure 19 A**).
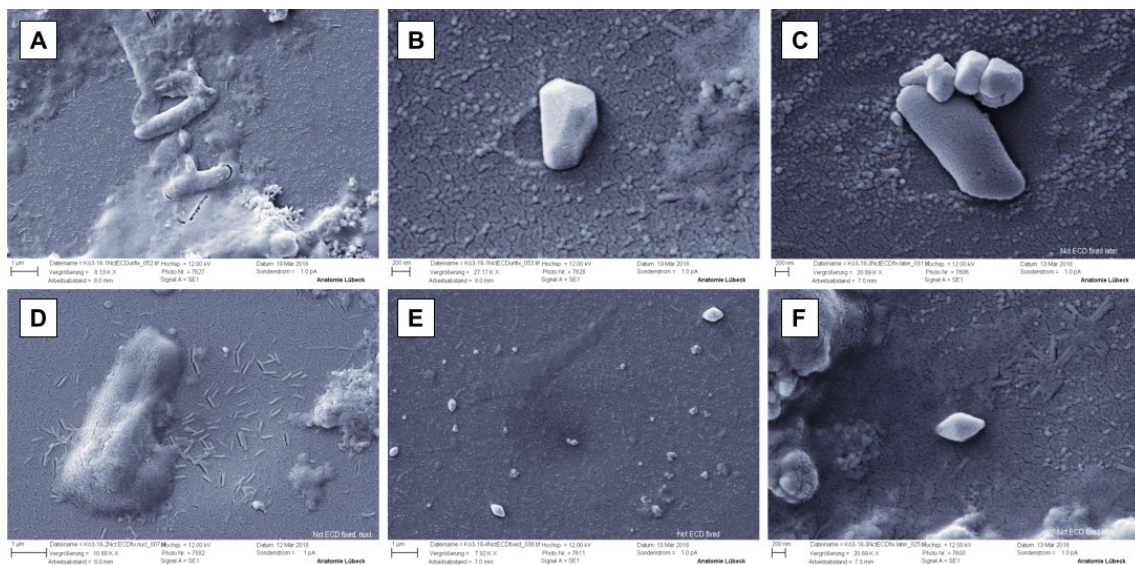
Figure 16 SEM of 0.01 % PFA fixated isolates from High Five cells, infected with Nct ECD SKL.
High Five cells were subjected to radio immuno precipitation assay (RIPA) buffer 4 days *p.i.*, step-wise centrifuged to separate the white crystal pellet from cellular debris, filtered and desalted. Electron energy was set to 12 kV. **A-D** show the first of two observed morphologies of crystal-like structures: an elongated one comparable in shape to the structures visible in unfixed and un-desalted SEM micrographs (see Figure 15) with an estimated size of 2- 4 x 1 x 1 μm^3 in size. Smaller, ~ 200 nm long rod-like structures visible in the **A**, **D** and **F** represent the fixated baculovirus. **E-F** The second morphology is characterized by a bipyramidal shape and a size of ~ 400 x 300 x 300 nm^3.

Harry Manfeldt prepared samples for SEM, starting from filtered and desalted isolates. Objects characterized by an elongated shape with clear edges and a bipyramidal shape of below one μm^3 were observed in the isolates. The presence of sharp-edged structures in the PFA-fixated isolates indicates that 0.01 % PFA fixation does not destroy Nct ECD SKL crystal integrity.

To speed up sample preparation, experiments were significantly nested. Parallel running tests showed only a minimal PFA concentration of 0.1 % fixation, incubation for 90 min and two washing steps with PBS prohibited viral spread adequately: In fluorescence microscopy (ZEISS) of cells infected with the supernatant of intermediate and final steps of rBV infected High Five cells no YFP fluorescence was detectable 4 days *p.i.*.

Since experiments were time-pressing and the SEM machine broken at the scheduled date, SEM imaging could not be repeated with crystal-containing cells fixated with higher PFA concentrations than 0.01 %. Initial jetting tests using 0.1 % PFA fixated rBV Nicastrin ECD and rBV HEX-1 cyto infected High Five cells after gravity flow filtering were performed at CFEL in cooperation with Max Wiedorn and Luigi Adriano (**Figure 17 A**). Nozzle sizes (20 μm, 30 μm and 40 μm) were tested at different jetting speeds. Different cell densities were jetted to determine the highest possible density producing a sufficiently stable jet.

Best results were obtained with a sample consisting of 20 % fixated cell pellet and 80 % PBS, continuously jetting for > 1h. In order to prevent clogging of the jet nozzle due to aggregated cells, samples were filtered, starting with a 100-μm filter, then a 50-μm filter, and a 30-μm filter (MACS SmartStrainers). Directly before the experiment, samples were filtered with a 10-μm filter and a two-μm filter (Valco).

## SFX using GDVN jetting of fixated crystal carrying insect cells

Jetting of the fixated High Five cells were performed at the Single Particles, Clusters, and Biomolecules & Serial Femtosecond Crystallography (SPB/SFX) instrument of the EuXFEL in April 2018. It was planned to first establish injection and data collection with the HEX-1 cyto as *in cellulo* model crystals to then, if successful, to then switch to Nct ECD SKL crystals. The jet ran unstably (**Figure 17 B**) for 20 min with a 40 μm 3D-printed GDVN.[250,251]
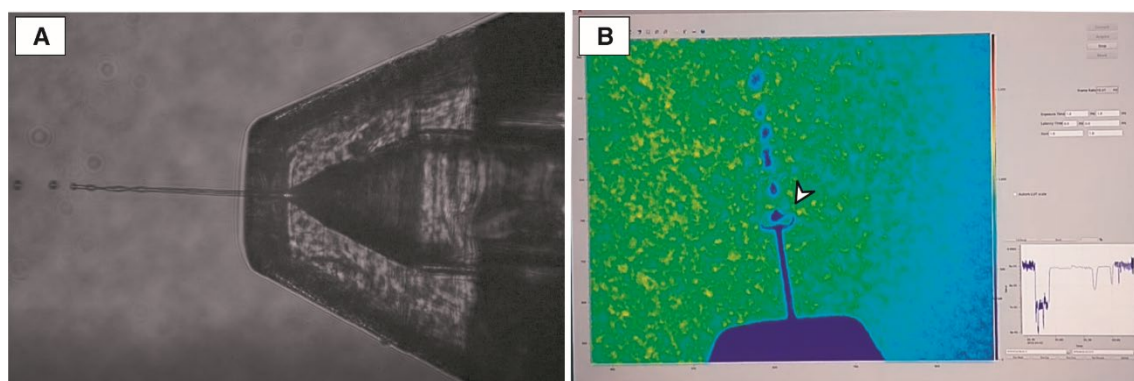
Figure 17 GDVN jetting of fixated crystal carrying cells.
A A 40-μm 3D-printed GDVN nozzle with a liqid jet of fixated rBV-infected High Five cells with HEX-1 cyto crystals during jetting tests at the CFEL. The image was captured using a FASTCAM SA4 model 500K-M1. **B** The liquid jet of fixated rBV-infected High Five cells with HEX-1 cyto crystals at the EuXFEL is depicted. A jet explosion (arrow) is visible after the fixated insect cell crosses the intersection of the X-ray beam.

After 20 min, the pressure started rising in an uncontrolled manner, probably due to cells clogging the nozzle tip. At this point, the injection was terminated and the sample was changed to prevent damaging the AGIPD 1M detector[253]. The hit rate for HEX-1 cyto crystals was below one per cent. However, when a cell with a crystal got hit by the beam, diffraction was recorded of up to 2 Å resolution (**Figure 18**).
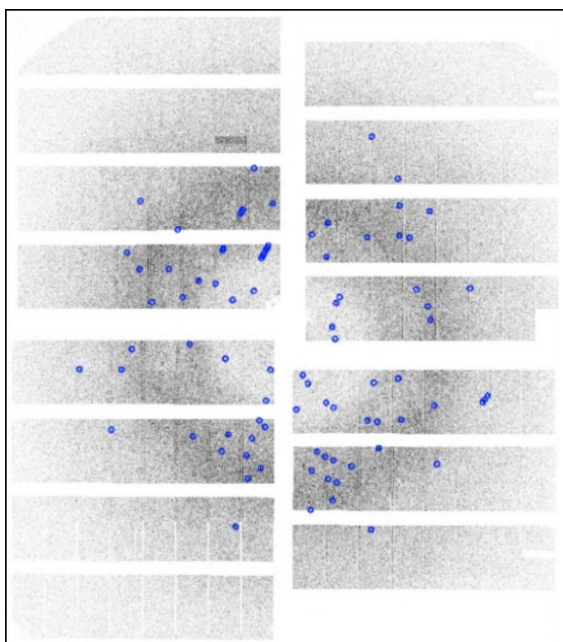
Figure 18 Exemplary AGIPD diffraction pattern of HEX-1 cyto in 0.1 % PFA fixated High Five cells.
Generated with a 15 μm beam focus at the EuXFEL. Bragg spot up to 2 Å identified with Cheetah are
encircled in blue. Deficient modules appear white-masked.

After removal of the sample reservoir, the sample was extracted with a syringe from the
upper and lower solution remaining in the reservoir (**Figure 19**). Despite continuous shaking
of fixated cells in PBS buffer[254], the majority of cells had quickly settled at the bottom of the
sample reservoir, preventing an injection into the XFEL beam. The reservoir rotation had
previously been esablished for isolated *in cellulo* CatB crystals to prevent settling of crystals
in the sample reservoir.[41] Since crosslinking with PFA influences the cellular mass density
due to permeabilization of membranes and osmotic pressure[255], the fixation procedure might
have accelerated settling speed.

Unfortunately, fixated Nct ECD SKL crystals in High Five cells caused immense clogging
within one minute, therefore more easily jettable samples were prioritized. The *in cellulo*
structure of human Nct ECD SKL remains to be solved in future experiments.
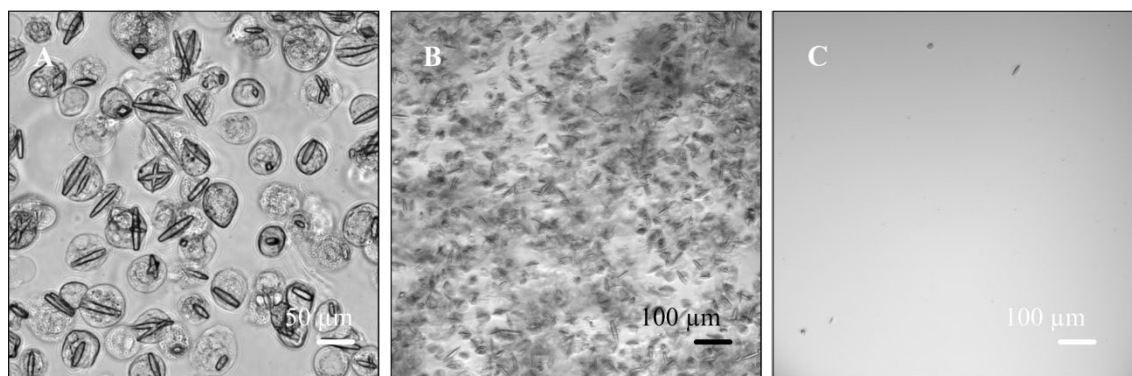
Figure 19 Settling of fixated insect cells in the sample reservoir despite sample rotator.
A Fixated High Five cells containing HEX-1 cyto crystals prior to jetting at the EuXFEL. **B** Dense cell-crystal pelett at the bottom of the sample reservoir. **C** The top of the sample reservoir mainly consists of buffer with few fixated cells and free-floating crystals. Despite continuos shaking both in the sample reservoir holder during the experiment and during the transport to the experimental hutch, fixated insect cells settled quickly during the experiment. Light microscopy images were taken with differential interference contrast using a *Nikon Ti-E* microscope.

In summary, proof-of-principle that jetting fixated, crystal-containing cells can produce diffraction was demonstrated. But, until a solution for the cell settling problem is found, the resulting hit-rate will prevent sufficient diffraction data collection in reasonable beamtime.

3.3.2 Fixed target *in cellulo* SFX on micro-patterned chip

Fixed-target serial femtosecond crystallography using *in cellulo* grown microcrystals

J. Mia Lahey-Rudolph, Robert Schönherr, Miriam Barthelmess, Pontus Fischer, Carolin Seuring, Armin Wagner, Alke Meents and Lars Redecke

IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

IUCrJ

Fixed-target serial femtosecond crystallography using *in cellulo* grown microcrystals

J. Mia Lahey-Rudolph,[a,b]‡ Robert Schönherr,[a,c] Miriam Barthelmess,[b] Pontus Fischer,[b] Carolin Seuring,[b,d] Armin Wagner,[e] Alke Meents[b,c] and Lars Redecke[a,c]*

[a]Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany, [b]Center for Free-Electron Laser Science (CFEL), Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany, [c]Photon Science, Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, 22607 Hamburg, Germany, [d]The Hamburg Center for Ultrafast Imaging, 22671 Hamburg, Germany, and [e]Diamond Light Source, Diamond House DH2-52, Chilton, Didcot OX11 0DE, United Kingdom. *Correspondence e-mail: redecke@biochem.uni-luebeck.de

The crystallization of recombinant proteins in living cells is an exciting new approach in structural biology. Recent success has highlighted the need for fast and efficient diffraction data collection, optimally directly exposing intact crystal-containing cells to the X-ray beam, thus protecting the *in cellulo* crystals from environmental challenges. Serial femtosecond crystallography (SFX) at free-electron lasers (XFELs) allows the collection of detectable diffraction even from tiny protein crystals, but requires very fast sample exchange to utilize each XFEL pulse. Here, an efficient approach is presented for high-resolution structure elucidation using serial femtosecond *in cellulo* diffraction of micometre-sized crystals of the protein HEX-1 from the fungus *Neurospora crassa* on a fixed target. Employing the fast and highly accurate Roadrunner II translation-stage system allowed efficient raster scanning of the pores of micro-patterned, single-crystalline silicon chips loaded with living, crystal-containing insect cells. Compared with liquid-jet and LCP injection systems, the increased hit rates of up to 30% and reduced background scattering enabled elucidation of the HEX-1 structure. Using diffraction data from only a single chip collected within 12 min at the Linac Coherent Light Source, a 1.8 Å resolution structure was obtained with significantly reduced sample consumption compared with previous SFX experiments using liquid-jet injection. This HEX-1 structure is almost superimposable with that previously determined using synchrotron radiation from single HEX-1 crystals grown by sitting-drop vapour diffusion, validating the approach. This study demonstrates that fixed-target SFX using micro-patterned silicon chips is ideally suited for efficient *in cellulo* diffraction data collection using living, crystal-containing cells, and offers huge potential for the straightforward structure elucidation of proteins that form intracellular crystals at both XFELs and synchrotron sources.

1. Introduction

During the past two decades, macromolecular X-ray crystallography (MX) has been revolutionized by the establishment of high-brilliance radiation sources and novel developments in serial diffraction data-collection strategies (Standfuss & Spence, 2017; Yamamoto *et al.*, 2017; Yabashi & Tanaka, 2017; Spence, 2020). The extremely short femtosecond pulses generated by X-ray free-electron lasers (XFELs) outrun radiation-damage effects, which limit the usable dose at synchrotron sources, enabling almost damage-free diffraction data collection at room temperature following the 'diffraction-before-destruction' approach (Neutze *et al.*, 2000; Chapman *et al.*, 2011). Immediately after a single high-energy XFEL pulse hits the sample, it is destroyed, essentially requiring continuous sample replenishment at the interaction region matching
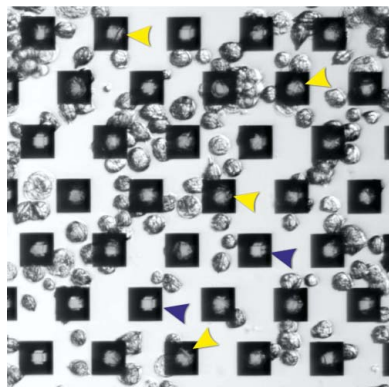
the pulse-repetition rate of the XFEL. This major challenge in XFEL diffraction experiments has been overcome in recent years by advancements in crystal-delivery strategies (Martiel *et al.*, 2019). In particular, advanced injection methods include, for example, liquid jets (DePonte *et al.*, 2008, 2011) and their developments (Oberthuer *et al.*, 2017; Knoška *et al.*, 2020), lipidic cubic phase (LCP; Weierstall *et al.*, 2014) and aerosol injectors (Bogan *et al.*, 2008; Hantke *et al.*, 2014), as well as hybrid methods, for example crystal extractors (Kovácsová *et al.*, 2017; Sugahara *et al.*, 2015; Park *et al.*, 2019; Conrad *et al.*, 2015) and tape drives (Roessler *et al.*, 2013; Beyerlein *et al.*, 2017), goniometer-based scanning using fixed-target supports (Zarrine-Afsar *et al.*, 2012; Hunter *et al.*, 2015; Roedig *et al.*, 2015, 2016; Murray *et al.*, 2015; Mueller *et al.*, 2015; Mehrabi *et al.*, 2020; Schulz *et al.*, 2018), and many more. The sequential collection of single-pulse diffraction images from thousands of individual crystals in random orientations subsequently merged into a three-dimensional data set enables the structure elucidation of biological macromolecules (Chapman *et al.*, 2011; Boutet *et al.*, 2012; Ayyer *et al.*, 2016). Based on the ultrashort X-ray pulses that diffract at high resolution even from crystals with dimensions in the submicrometre size range, this serial femtosecond X-ray crystallography (SFX) approach provides new opportunities for structural biology (Johansson *et al.*, 2017), including time-resolved X-ray diffraction to visualize the molecular dynamics of macromolecules by pump–probe experiments (Kupitz *et al.*, 2014; Tenboer *et al.*, 2014; Shimada *et al.*, 2017; Nango *et al.*, 2016). Sequential crystal-delivery strategies have successfully been adapted for serial synchrotron data collection to reduce dose accumulation in a single crystal, advancing the application of serial synchrotron crystallography (SSX) on microcrystals even at room temperature (Gati *et al.*, 2014; Stellato *et al.*, 2014; Roedig *et al.*, 2016; Owen *et al.*, 2017; Weinert *et al.*, 2017; Schulz *et al.*, 2018; Mehrabi *et al.*, 2019).

As a consequence of the increased X-ray intensities and serial data-collection strategies, small crystals formed within living cells, denoted '*in cellulo* crystals', also became suitable targets for X-ray crystallography (reviewed in Schönherr *et al.*, 2018). Depending on the specific demands of the target crystals, diverse diffraction data-collection strategies have been reported for the elucidation of structural information from recombinant proteins that form intracellular crystals. At state-of-the-art synchrotron sources, serial mesh-screening approaches using microcrystals randomly positioned and frozen in a loop (Gati *et al.*, 2014) or on a mesh mount (Coulibaly *et al.*, 2007; Boudes *et al.*, 2016) have been applied. In 2012, the capability of SFX to generate new bioinformation was initially confirmed by the elucidation of the native *Trypanosoma brucei* procathepsin B (TbCatB) structure at 2.1 Å resolution using *in cellulo* grown and extracted crystals (Redecke *et al.*, 2013), followed by a number of additional examples highlighting the synergistic combination of SFX and intracellular protein crystallization (Gallat *et al.*, 2014; Ginn *et al.*, 2015; Gati *et al.*, 2017; Nass *et al.*, 2020). Motivated by the extraordinary intrinsic stability of the crystals that were used as the first diffraction targets, such as insect virus polyhedra

(Coulibaly *et al.*, 2007), *in cellulo* grown crystals were usually isolated and purified from the cells prior to diffraction data collection to enable liquid-jet delivery. Moreover, isolation should avoid diffuse background scattering from the surrounding cellular material, which has been suggested to limit the recordable diffraction from the microcrystals (Duszenko *et al.*, 2015).

For the delivery of isolated *in cellulo* grown crystals into the XFEL beam, liquid-jet injectors that utilize gas dynamic virtual nozzles (GDVNs) have consistently been used, and this is the most commonly used method for SFX crystal injection to date (DePonte *et al.*, 2008). GDVNs inject at rapid flow rates (Schlichting, 2015), which is a benefit if fast reactions are being probed (Calvey *et al.*, 2016) or if ultrafast repetition rates need to be matched, for example at the European XFEL (Wiedorn, Awel *et al.*, 2018; Wiedorn, Oberthür *et al.*, 2018). On the other hand, rapid flow causes high sample consumption, as large quantities of crystals pass unexposed between subsequent X-ray pulses, significantly affecting the proportion of crystals hit by XFEL pulses. Typically, hundreds of thousands of small crystals are required to completely sample reciprocal space if the hit rate is less than 10%, as observed during GDVN injection of isolated *in cellulo* grown TbCatB (7.4%) and *T. brucei* IMP dehydrogenase (TbIMPDH; 2.3%) crystals (Redecke *et al.*, 2013; Nass *et al.*, 2020), which limits application to more challenging targets. Moreover, GDVN injectors tend to produce unstable jets due to clogging of the nozzle with crystals, and the aqueous environment usually results in an increased X-ray background scattering, which varies from frame to frame (Martiel *et al.*, 2019).

Aside from complications from the injector setup itself, the integrity of the protein samples can be affected by small environmental changes, as also observed for *in cellulo* crystals. Firefly luciferase and Xpa crystals grown in insect and mammalian cells, respectively, dissolve quickly after disruption of the cellular membrane (Schönherr *et al.*, 2015; Tsutsui *et al.*, 2015). On the other hand, high-quality diffraction data can be collected from *in cellulo* grown crystals that remain within the living cells without any impact from the cellular environment, as recently confirmed using synchrotron radiation (Tsutsui *et al.*, 2015; Baskaran *et al.*, 2015) and XFEL pulses (Sawaya *et al.*, 2014). Thus, it appears to be beneficial to exploit the protective effects of the viable cells on the intracellular crystals for diffraction data collection, instead of performing detrimental isolation and purification procedures. However, liquid-jet injection of intact crystal-containing cells at an XFEL is challenged by the comparatively large cell diameter (between 15 and 20 µm for insect cells), provoking increased settling and aggregation that might lead to rapid clogging of the jet nozzle.

In this study, we present a method for *in cellulo* SFX diffraction data collection that exploits the benefits of fixed-target sample delivery. To establish this approach, the selected test protein needs to fulfil specific requirements. Since the expected hit rate in the fixed-target SFX (FT-SFX) experiment directly correlates with the number of crystal-containing cells, crystals within every cell of the culture would be optimal. Only

one crystal per cell would reduce the chance of multiple patterns per detector frame. Therefore, the protein should form intracellular crystals with dimensions in the low-micrometre size range in living cells reproducibly and with high efficiency to allow estimation of the crystallization efficiency by light microscopy. Finally, a reference structure should be available to validate the *in cellulo* FT-SFX approach. Considering these requirements, we have chosen the protein HEX-1 from the filamentous fungus *Neurospora crassa*, which represents a prominent example of native *in cellulo* protein crystallization (Schönherr *et al.*, 2018), as the crystallization target, employing a baculovirus-based approach for gene expression in insect cells.

In ascomycetes, HEX-1 self-assembles into a hexagonal crystalline core after import into peroxisome-derived compartments, named Woronin bodies, located in the vicinity of the septal pore (Markham & Collinge, 1987). The fungi exploit the stiffness and mechanical stability of the crystalline lattice to withstand the high intracellular turgor pressure to seal pores after damage to the organism, preventing cytoplasmic bleeding (Tenney *et al.*, 2000; Yuan *et al.*, 2003). Elucidation of the structure of HEX-1 recombinantly produced in in *Escherichia coli* and crystallized applying the sitting-drop vapour-diffusion method provided insights into the native assembly process (Yuan *et al.*, 2003). This HEX-1 structure determined from synchrotron diffraction of single crystals using the rotation method at cryogenic temperatures is subsequently denoted the HEX-1 reference structure. We recently reported that spontaneous self-assembly of HEX-1 into intracellular crystals is not restricted to the native environment of the fungal cells. On infection with a recombinant baculovirus encoding the HEX-1 gene, living insect cells also form regular, micrometre-sized hexagonal crystals (Lahey-Rudolph *et al.*, 2020), supporting a strong intrinsic crystallization tendency of HEX-1 that might have evolved along with its physiological function.

Sample supports for fast, reliable and efficient crystal delivery in FT-SFX and SSX have recently been described, particularly with regard to their increased hit rates and reduced background scattering compared with liquid-jet and LCP injection systems (Meents *et al.*, 2017; Lieske *et al.*, 2019; Tolstikova *et al.*, 2019; Mehrabi *et al.*, 2020). Diverse solid supports have already been designed based on silicon or polymers, including microgrids (Zarrine-Afsar *et al.*, 2012; Feld *et al.*, 2015; Baxter *et al.*, 2016) and silicon chips (Hunter *et al.*, 2015; Murray *et al.*, 2015; Roedig *et al.*, 2015, 2016, 2017; Schulz *et al.*, 2018; Karpik *et al.*, 2020; Shelby *et al.*, 2020; Frank *et al.*, 2014; Mehrabi *et al.*, 2020), and have successfully been established for isolated, conventionally grown protein crystals. We performed FT-SFX experiments at the Macromolecular Femtosecond Crystallography (MFX) end station of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, Menlo Park, California, USA (Sierra *et al.*, 2019) using crystal-containing insect cells loaded into the pores of micro-patterned, single-crystalline silicon chips in combination with the fast and accurate Roadrunner II translation-stage system (Lieske *et al.*, 2019). The location of the HEX-1 crystal-containing cells at predefined positions on the chip allowed fast and efficient raster scanning through the X-ray pulses, which were spatially aligned with the micro-pores of the chip by synchronization with the 120 Hz pulse-repetition rate of the LCLS (Roedig *et al.*, 2017). High hit rates of up to 30% (detector frames containing diffraction) resulted in short data-collection times and significantly reduced the sample consumption and the background scattering caused by non-crystalline material around the crystal compared with liquid-jet SFX. The overall agreement of the elucidated HEX-1 structure with that previously determined using synchrotron radiation and single HEX-1 crystals grown by sitting-drop vapour diffusion (PDB entry 1khi; Yuan *et al.*, 2003) validated our approach and demonstrated that FT-SFX is ideally suited for efficient *in cellulo* diffraction data collection using living, crystal-containing cells.

2. Materials and methods

2.1. Insect-cell expression of *N. crassa* HEX-1 and *in cellulo* crystallization

The cloning procedure and recombinant bacmid generation for HEX-1 from the filamentous fungus *N. crassa* (GenBank accession No. XM_958614) have previously been described (Lahey-Rudolph *et al.*, 2020). In brief, the HEX-1 coding sequence was amplified using primers 5′-TACTACGACGAC GACGCTCACG-3′ (sense) and 5′-GAGGCGGGAACCG TGGACG-3' (antisense) and blunt-end ligated into an EheI-restricted pFastBac1 vector containing the sequence 5′-ATGGGCGCCTAA-3′ between the BamHI and HindIII restriction sites. For recombinant baculovirus (rBV) production, recombinant bacmid DNA was generated in *E. coli* DH10EmBacY cells (Geneva Biotech) and used for lipofection of *Spodoptera frugiperda* Sf9 insect cells with Escort IV reagent (Sigma–Aldrich) according to the manufacturer's instructions. The virus titre of the third-passage (P3) stock was calculated using the TCID$_{50}$ (tissue-culture infectious dose; Reed & Muench, 1938) in a serial dilution assay as described previously (Lahey-Rudolph *et al.*, 2020).

For intracellular crystallization, 0.8×10^6 Sf9 insect cells were plated in 2 ml serum- and protein-free ESF921 insect-cell culture medium (Expression Systems) per six-well cell-culture plate and subsequently infected with recombinant P3 baculovirus as a vector for HEX-1 with a multiplicity of infection (MOI) of 1. The cells were incubated at 27°C for 96 h. *In cellulo* crystal formation was verified and crystallization efficiency was determined by light-microscopic cell imaging using a Leica DM IL LED microscope equipped with a 40× objective employing integrated modulation contrast (Fig. 1).

2.2. Chip loading

For diffraction experiments, single-crystalline silicon microchips designed for the Roadrunner II goniometer (Tolstikova *et al.*, 2019) were used. Each 36 × 13 mm frame comprised a 300 µm thick chip with three 10 × 10 mm silicon windows thinned to 25 µm. These windows were patterned

with a triangular grid of 12 µm pores spaced 90 and 100 µm apart, yielding approximately 44 000 pores that could hold crystal samples (Fig. 2). Prior to loading, the chips were glued onto an aluminium support.

Sf9 cells containing intracellular HEX-1 crystals were infected and incubated as described above before being transported to the MFX end station at room temperature. For loading onto one micro-patterned silicon chip, the cells were rinsed gently from the bottom of each well with a 1 ml pipette. After combining cells from three wells of a six-well culture plate into a reaction vessel, the cells were centrifuged for 1 min at 300g. The supernatant was discarded, and the cell pellet was gently resuspended in 200 µl ESF921 cell-culture medium and pipetted onto the chip surface. Utilizing a wedge cut from Whatman filter paper (grade 1), the medium was blotted off the back of the chip, pulling the cells into the 12 µm pores of the silicon support membrane. The cell coverage of the pores was confirmed by optical light microscopy of the chip using an Olympus SZX16 microscope illuminated by an Olympus KL 1600 LED (Fig. 2). Until just before loading the sample, the two cell-loaded chips were protected from dehy-

dration using a 'humidor', a small movable cover that provides a humid environment (Lieske et al., 2019).

2.3. FT-SFX data collection

FT-SFX experiments were carried out at the MFX end station at LCLS (experiment No. P09215). X-ray pulses with a photon energy of 9.45 keV and a length of 27.5 fs were focused to a 1 × 1 µm spot. The pulse energy was attenuated to 20% of the full flux (4.5 mJ per pulse). Chips loaded with the crystal-containing cells were mounted on the high-precision Roadrunner II fast-scanning goniometer (Tolstikova et al., 2019) in a humidified helium atmosphere to prevent dehydration of the cells during diffraction data collection. After manual alignment, data were collected at room temperature with a Cornell–SLAC pixel-array detector (CSPAD; Blaj et al., 2015) by raster-scanning the chips through the X-ray focus row by row. The scanning speed was precisely controlled to synchronize the arrival of each X-ray pulse at the 120 Hz repetition rate of the LCLS with the spatial alignment of a micro-pore on the chip (Roedig et al., 2017; Lieske et al., 2019).
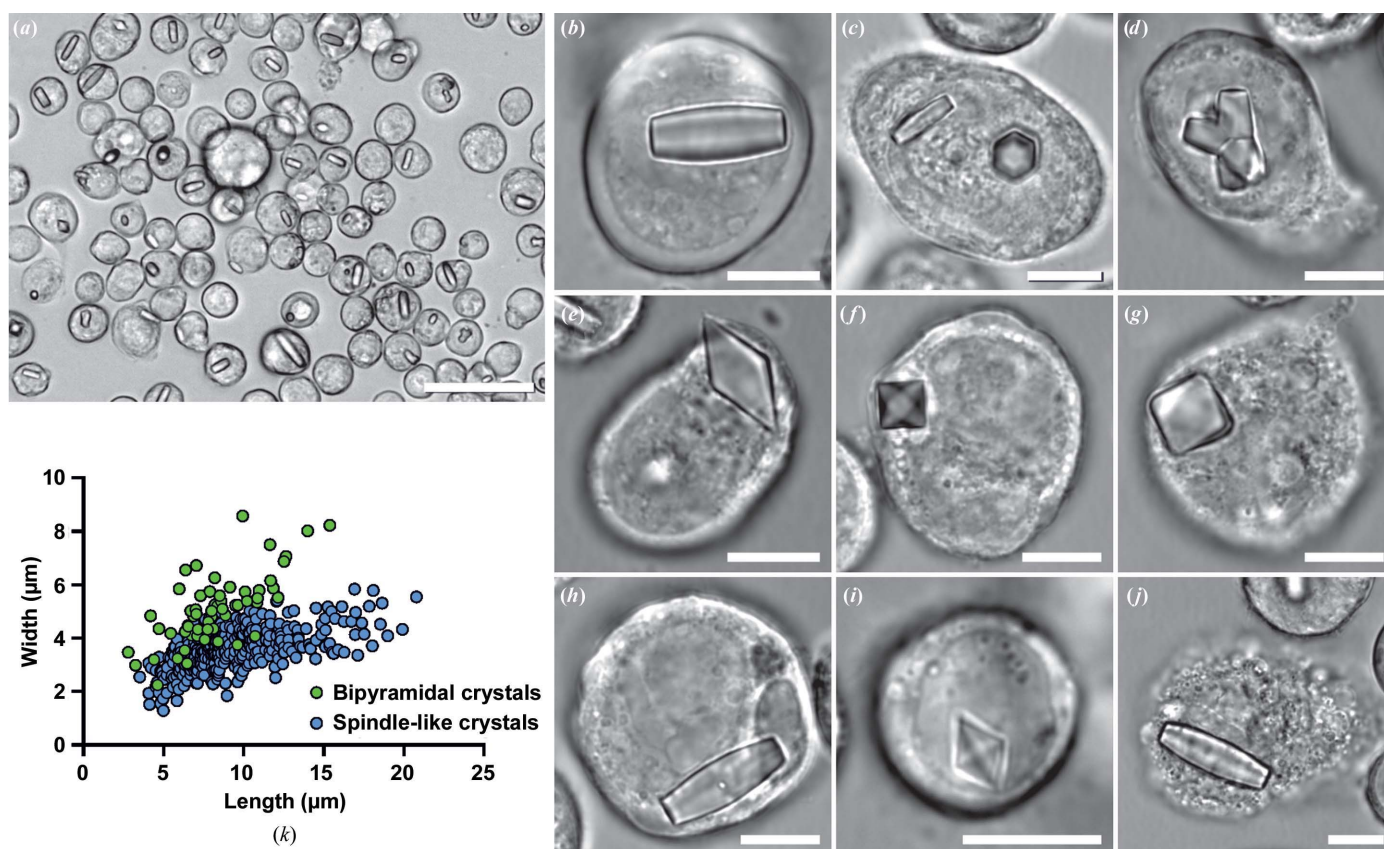

Figure 1
Crystallization of HEX-1 from *N. crassa* in living insect cells. Light-microscopic images of Sf9 cells four days after infection with a recombinant baculovirus encoding the target protein (MOI = 1). Most cells produce a single crystal per cell (a). Two different crystal morphologies can appear. The first is a spindle-like morphology with flat ends and a slightly increased diameter in the middle (b) showing a hexagonal cross section (c). These crystals sometimes grow in a star-like manner, conjoined in the middle (d). The second morphology is bipyramidal (e) with pointed tips and a square base (f). These crystals can sometimes form somewhat irregular forms with rounded tips (g). Both morphologies can grow cytoplasmatically as well as within the cell nucleus [(b) versus (h) and (e) versus (i)]. After cell lysis the crystals show high stability in the cell-culture medium, while mostly sticking to cell remnants (j). Bipyramidal crystals constitute about 13% of the total crystal population. The size distribution shows that the bipyramidal crystals form shorter but broader crystals overall compared with the spindle-like crystals (k). Size bars: 50 µm (a), 10 µm (b)–(j).

By enclosing the direct beam shortly before and after the sample in thin-walled tantalum capillaries and by continuously flushing the Roadrunner humidity chamber with humidified helium gas, air scattering was further reduced to a minimum (Meents *et al.*, 2017). The sample-to-detector distance was set to 116 mm, corresponding to a maximum resolution of 2.0 Å at the detector edge. Higher resolution diffraction data were collected in the detector corners. Data collection was monitored using *OnDa* for immediate feedback in terms of the hit rate (Mariani *et al.*, 2016).

### 2.4. Data processing and structure refinement

Identification of hits, indexing, integration and data reduction were performed using *indexamajig* as implemented in *CrystFEL* version 0.6.3 (White *et al.*, 2016), involving the application of the fast Fourier transform-based autoindexing algorithms *MOSFLM* (Leslie, 2006), *DirAx* (Duisenberg, 1992) and *XDS* (Kabsch, 2010). Concentric rings to determine background, buffer and peak estimation regions were set to 6.4, 5 and 3.4, respectively. To improve the data quality, the geometry of the CSPAD was refined using *geoptimiser* (Yefanov *et al.*, 2015). Frames containing more than three diffraction patterns of individual crystals were sorted out. The resulting stream file was subjected to post-refinement (scaling) and merged using *partialator* version 0.6.3, excluding saturated peaks above 14 300 ADU (analogue-to-digital units). Figures of merit were calculated using *compare_hkl* (R_{split}, $CC_{1/2}$ and CC*) and *check_hkl* (signal-to-noise ratio, multiplicity and completeness) from the *CrystFEL* suite (White *et al.*, 2016). Merged intensities obtained by *CrystFEL* were converted to MTZ format for further processing. Molecular replacement was performed with *Phaser* (McCoy *et al.*, 2007) using the

HEX-1 structure previously obtained from crystals formed by conventional vapour diffusion as a search model (PDB entry 1khi; Yuan *et al.*, 2003), followed by iterative circles of refinement using *phenix.refine* (Afonine *et al.*, 2012) and manual model building with *Coot* (Emsley *et al.*, 2010). *Polygon* (Urzhumtseva *et al.*, 2009) and *MolProbity* (Chen *et al.*, 2010) were used for final model validation. The HEX-1 coordinates have been deposited in the PDB (accession codes 7asx and 7asi). Structural C^{α}-atom overlays of residues 31–171 with the equivalent range of PDB entry 1khi were performed using *SUPERPOSE* from the *CCP*4 suite (Winn *et al.*, 2011) and were visualized in *PyMOL* (Schrödinger).

3. Results and discussion

3.1. Intracellular crystallization of HEX-1

In this study, our established *in cellulo* crystallization protocol was applied to produce quantities of intracellular HEX-1 crystals sufficient to elucidate its high-resolution structure by FT-SFX *in cellulo* diffraction. At day four post infection (p.i.) with the HEX-1 rBV, light-microscopy investigation of the Sf9 insect cells showed that 85% of the cells used for diffraction contained HEX-1 crystals, with a minor percentage of these cells harbouring multiple crystals [Fig. 1(*a*)]. Two crystal morphologies can be observed in infected insect cells: an elongated, spindle-shaped morphology with a hexagonal cross section [Figs. 1(*b*) and 1(*c*)] and a less often appearing bipyramidal shape with a square base [Figs. 1(*e*) and 1(*f*)]. Both formed in the cytoplasm or the nucleus [Figs. 1(*b*), 1(*e*), 1(*h*) and 1(*i*)]. The crystals exhibited average dimensions of 9.1 ± 3.2 µm in length and 3.5 ± 0.7 µm in width in the case of the spindle-like shape and 8.3 ± 2.5 µm

in length and 5.1 ± 1.3 µm in width in the case of the bipyramidal crystals [Fig. 1(*k*)], not exceeding the normal dimensions of infected Sf9 cells (17.8 ± 2.6 µm) or affecting cell viability. In rare cases, spindle-shaped crystals grew up to a maximum length of 20 µm and a maximum width of 5.8 µm [Fig. 1(*k*)]. If cell lysis occurred due to the ongoing viral replication process, individual HEX-1 crystals floated in the medium or stayed attached to cell remnants, indicating significant crystal stability in the cell-culture medium [Fig. 1(*j*)]. However, free-floating HEX-1 crystals were rarely detected, in contrast to previous observations during the crystallization of TbIMPDH and TbCatB in living insect cells (Nass *et al.*, 2020; Redecke *et al.*, 2013), which might be attributed
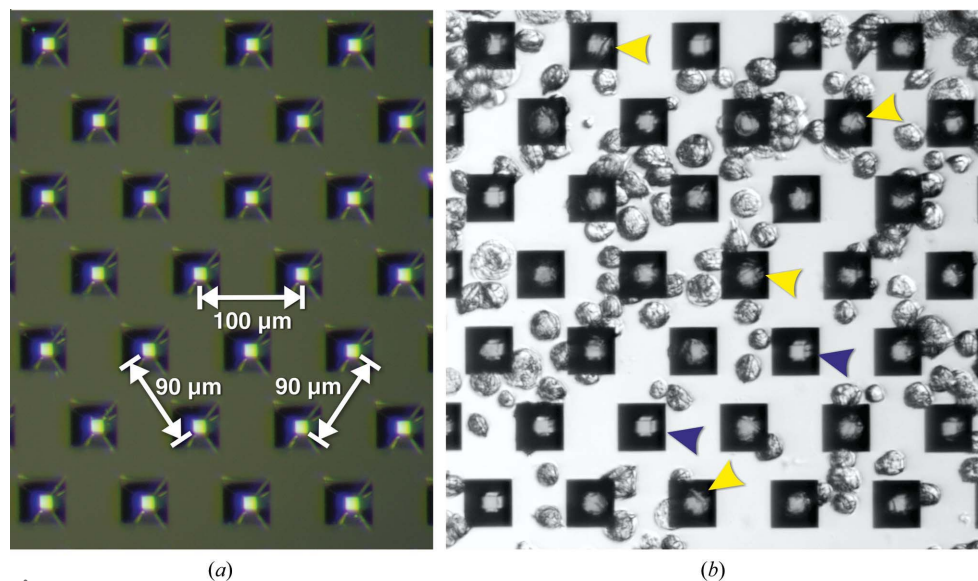
Figure 2
(*a*) An empty micro-patterned silicon chip designed for the Roadrunner II goniometer. The chip contains 12 µm sized pores that are spaced 90 and 100 µm apart, allowing well defined positioning of crystal-containing cells. (*b*) Sf9 insect cells containing *in cellulo* grown HEX-1 crystals were sucked into the chip pores. Visible intracellular crystals positioned in the pores are marked with yellow arrows; blue arrows point into empty pores. The images were recorded using reflected light on an Olympus SZX16 microscope.

to the used bacmid. The TbIMPDH and TbCatB rBVs were generated applying the Bac-to-Bac system, while the HEX-1 gene was transposed into the EmBacY bacmid (Trowitzsch *et al.*, 2010). Due to the deletion of the viral cathepsin and chitinase genes, the generated baculovirus caused a reduced degree of cell lysis four days p.i.. The high intracellular crystallization efficiency and the comparatively narrow size range of the obtained *in cellulo* crystals, as well as the available reference structure, indeed qualify *N. crassa* HEX-1 as a suitable test protein to adapt FT-SFX to *in cellulo* diffraction data collection. In future studies, this intracellular diffraction data-collection approach may enable the structure solution of more delicate systems that lack the high intrinsic stability of HEX-1 crystals outside the protective cellular environment.

3.2. Sample loading onto micro-patterned silicon chips

In terms of sample loading, crystal-containing insect cells behaved like isolated protein crystals, except for the increased dimensions of the cells (15–20 µm) that need to be considered during chip design. We used silicon chips designed for the Roadrunner II fast-scanning goniometer (Lieske *et al.*, 2019; Tolstikova *et al.*, 2019) with micro-pores of 12 µm in diameter to harbour the crystal-containing cells for pre-positioning. A spacing of 90 µm between the pores optimized the cell density to a single cell per pore on average. Since the surrounding liquor was efficiently removed by applying capillary forces from below by touching the bottom chip surface with a filter paper, the living insect cells were directly loaded onto the chip surface by carefully pipetting the cells suspended in culture medium. This loading procedure made washing steps or buffer exchange to fit the specific requirements of the diffraction experiment obsolete. Combining three wells of a six-well plate, each containing approximately 0.8×10^6 Sf9 cells, resulted in the loading of 2.4×10^6 cells per chip. This cell density optimized the cell coverage of the 44 000 pores of the chip after removal of the medium, as revealed by reflective light microscopy (Fig. 2). Almost all pores were filled with a single, crystal-containing cell in a monolayer, limiting the expected multiple crystal hits by a single XFEL pulse during diffraction data collection.

Sample dehydration after the removal of mother liquor through the pores of a micro-patterned chip has been considered as an obstacle in FT-SFX experiments, and various technical solutions have been designed to keep isolated crystals hydrated during chip loading, transport to the beamline and data collection (Lieske *et al.*, 2019; Shelby *et al.*, 2020; Martiel *et al.*, 2019). Protein crystals formed in a living cell are to a certain degree intrinsically protected from environmental stress if the cell membrane is intact. However, due to the presence of aquaporins in the cellular membrane, strong evaporation pressure on the cell surface will be compensated by water efflux from inside the cell, damaging the contained crystals. To prevent the dehydration of cells and crystals, loading of the crystal-containing cells onto the chip surface was performed in a humidified environment. The 'humidor', a specific humidifying feature of the Roadrunner II chips

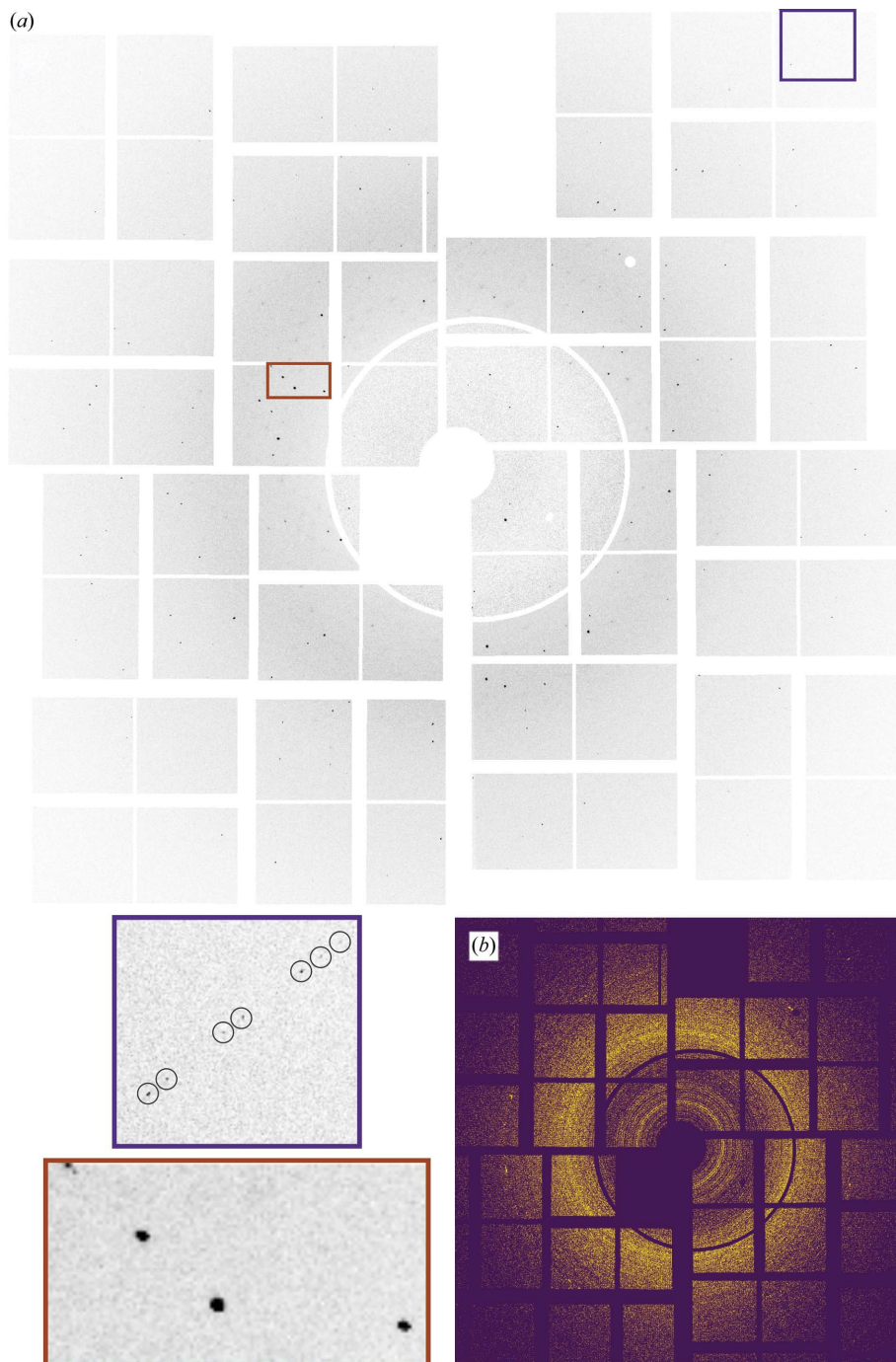(Lieske *et al.*, 2019), was slid over the prepared chips during transport and mounting.

3.3. FT-SFX structure determination

In the humidified helium environment at the MFX end station, detector images from a total of 87 450 X-ray pulses were recorded by the CSPAD detector from raster-scanning chip 1 (43 890 frames) and chip 2 (43 560 frames) within 25 min. Chips 1 and 2 resulted in 13 247 and 6429 frames containing detectable Bragg spots, respectively, yielding hit rates of approximately 30% (chip 1) and 15% (chip 2). Since the chip pores are specifically targeted with the XFEL beam, almost full pore coverage with cells that form HEX-1 crystals with 85% efficiency should result in higher hit rates. However, this is limited by diffraction from empty holes, holes with empty cells, holes with intracellular crystals positioned outside the pore centre and holes with multiple crystals grown either in a single cell or in multiple cells. The recorded diffraction patterns showed an average maximum resolution of 2.3 Å (Supplementary Fig. S1). Individual patterns contained diffraction even at the corners of the detector (Fig. 3), corresponding to a maximum per-frame resolution of 1.70 Å. This clearly indicates a higher diffraction capacity of the intracellular HEX-1 crystals within the cellular environment than that recorded using the available experimental setup. However, this FT-SFX experiment was part of a protein crystal screening beamtime that immediately followed another beamtime using the Roadrunner II goniometer setup at the MFX end station. Due to the short time for switching between these two shifts there was no time to optimize the goniometer–detector geometry specifically for this experiment, which somewhat limited the achievable resolution. No indications of crystal damage during diffraction data collection were obtained. Calculating the maximum and average per-pattern resolution for each quarter of a split data set resulted in comparable distributions (Supplementary Fig. S1). Only low background contribution was observed in the majority of the diffraction patterns, as exemplarily shown in Fig. 3(*a*), confirming the minor diffuse scattering from the soft matter of the insect cell embedding the HEX-1 crystals. The efficient removal of cell-culture medium during sample loading was revealed by the absence of significant water scattering in most patterns, which is typically detected around 3.1 Å resolution in room-temperature liquid-jet SFX experiments (Chapman *et al.*, 2011; Boutet *et al.*, 2012). Flow alignment of rod-shaped crystals represents another problem that frequently occurs during serial data collection of rod-shaped crystals, if injected into the XFEL beam by a liquid jet. This was previously observed for *in cellulo* grown, but isolated TbCatB crystals (Koopmann *et al.*, 2012) and for human aquaporin 2 (hAQP2) crystals (Lieske *et al.*, 2019. The preferred crystal orientation requires an increased number of hits to record a complete diffraction data set, and thus more beamtime. In FT-SFX a similar problem occurs when non-isometric, such as plate-like or needle-shaped, crystals attach in a preferred orientation to the chip surface (Lieske *et al.*, 2019). Apparently, in the

present case the roundish cells prevented such a preferred orientation of the elongated HEX-1 crystals during chip loading, as revealed by a virtual powder pattern exhibiting a uniform orientation distribution of the HEX-1 crystals in the intact cells loaded onto the silicon chip surface [Fig. 3(b)].

Figure 3
(a) An individual diffraction pattern of an intracellular HEX-1 crystal in a Sf9 cell, recorded at room temperature on a Roadrunner II chip with a CSPAD detector. The maximum resolution was limited by the setup to 2.0 Å at the detector edges. Higher resolution peaks of up to 1.70 Å could be detected in the detector corners (blue inset). The shape of the peaks indicates low mosaicity (red inset). (b) Virtual powder diffraction pattern of HEX-1 crystals recorded *in cellulo* shows a homogeneous distribution of crystal orientations. The sum of detected Bragg peaks detected by *indexamajig* from the *CrystFEL* suite in all images resulting from the data collection from chips 1 and 2 is depicted.

Applying the parameters of a primitive hexagonal unit cell, 15 224 of the totally recorded diffraction patterns were successfully indexed (77% indexing rate), including 15% successfully indexed multi-hits from either two or three crystals. Initially, two populations were found, differing in the length of the c axis by 0.93 Å (Supplementary Fig. S3). Following indexing with *MOSFLM*, *ASDF*, *XDS* and *stream-grep*, 12 180 patterns originating from a single crystal population with unit-cell parameters $a = b = 58.73 \pm 0.08$, $c = 192.83 \pm 0.24$ Å, $\alpha = \beta = 90 \pm 0.06°$, $\gamma = 120 \pm 0.07°$ were combined. Within the error of the respective measurements, the c axis of the observed unit cell is slightly shortened compared with that of the HEX-1 reference structure, while the a and b axes are slightly elongated ($a = b = 57.43$, $c = 196.98$ Å; PDB entry 1khi; Yuan *et al.*, 2003). The refined unit-cell parameters extracted from X-ray powder diffraction tests of HEX-1 crystal-containing insect cells on a synchrotron beamline ($a = b = 58.01$, $c = 195.27$ Å; Lahey-Rudolph *et al.*, 2020) had already indicated a similar composition of the *in vitro* and *in cellulo* grown HEX-1 crystals, both belonging to space group $P6_522$ and containing one HEX-1 molecule in the asymmetric unit. Thus, the unit-cell shortening of c, together with a slight expansion of a and b, are more likely to be attributable to an impact of the different buffers in which the cells were floating (ESF 921 insect-cell culture medium versus Tris-buffered saline) during diffraction data collection than to an effect of the intracellular environment of the living insect cells during crystal growth.

It was possible to assemble a complete high-quality data set to elucidate the FT-SFX structure of HEX-1 using only diffraction data obtained from chip 1, collected within 12 min. 8579 indexed crystal diffraction snapshots filtered using the *stream-grep* script were merged into the resulting stream file with *indexamajig*. Molecular replacement using the coordinates of the *N. crassa* HEX-1 reference structure (PDB entry 1khi) as a search model and subsequent refinement resulted in a structural model of *in cellulo* crystallized HEX-1 at 1.80 Å resolution. The resolution is comparable with that

obtained by synchrotron diffraction of a single native $0.2 \times 0.2 \times 0.2$ mm HEX-1 crystal at 100 K (1.8 Å; Yuan *et al.*, 2003). However, considering the previously mentioned limits imposed by the goniometer–detector geometry on the maximum recordable resolution in this study, a better diffraction capability of the HEX-1 *in cellulo* crystals cannot be excluded. As expected, increasing the number of indexed and *stream-grep* filtered patterns to 12 180 (from 19 676 recorded hits) by combination of diffraction data from chips 1 and 2 slightly improved the quality parameters of the data set, for example the signal-to-noise ratio (5.17 to 5.80), CC* (0.987 to 0.991) and $CC_{1/2}$ (0.951 to 0.964), as a consequence of the increased multiplicity of the individual Bragg peaks (107.66 to 130.14). The maximum resolution of the data set improved by 0.1 Å to 1.70 Å (Table 1).

3.4. HEX-1 structure

The 19 kDa protein HEX-1 forms a two-domain structure consisting of mutually perpendicular antiparallel β-barrels. The N-terminal domain is formed by six antiparallel β-strands and a single short helix, while the C-terminal domain consists of a five-stranded β-barrel and two helices [Fig. 4(*a*)]. The resulting high-quality electron density obtained by FT-SFX provides a high level of detail, as shown for part of the HEX-1 polypeptide chain in Fig. 4(*b*). No interpretable electron density is observed for the N-terminal (residues 1–30) and C-terminal (residues 171–176) HEX-1 residues and for residue His160, indicating high flexibility of these parts of the structure. Likewise, the side chains of residues Arg41, Gln78, Gln112, Asp113, Asn139, Lys143, Glu146, Ser147, Asp159 and Arg162 are poorly defined or undefined. Two alternate conformations were found for Val157. Comparison of the HEX-1 structures obtained by crystallization and X-ray diffraction in living insect cells with the reference structure revealed no major differences in the overall conformation of the HEX-1 monomer. In general, the structures are similar, with overall root-mean-square deviations (r.m.s.d.s) of 0.466 Å (chip 1 structure) and 0.467 Å (chips 1 and 2 structure) for 141 C^α atoms. Small deviations are restricted to the loop regions as well as the linker region connecting the N- and C-terminal domains, which are known to be characterized by increased flexibility [Fig. 4(*c*)]. To rule out phase bias imposed by phase determination applying the MR approach with the similar reference structure as the search model, an OMIT map was calculated for residues 62–65 located in the most deviating loop, which defines the atomic positions of the HEX-1 *in cellulo* model well [Fig. 4(*d*)].

Three different patterns of strong intermolecular interactions have been described to stabilize the crystal lattice of HEX-1, representing the structural basis of its intrinsic tendency for self-assembly (Yuan *et al.*, 2003). All residues previously identified to establish these salt bridges and hydrogen bonds are well defined in the FT-SFX electron density, revealing conservation of the lattice interactions in the *in cellulo* grown HEX-1 crystals (Supplementary Table S1, Supplementary Fig. S4). Independent of the environmental

Table 1
FT-SFX data-collection and refinement statistics for *N. crassa* HEX-1 loaded onto Roadrunner II chips.

Values in parentheses are for the highest resolution shell.

	Chip 1	Chips 1 and 2
Data collection		
PDB code	7asx	7asi
No. of frames collected	43840	87450
No. of *Cheetah* hits [hit rate, %]	13247 [30.2]	19676 [22.5]
No. of indexed patterns [indexing rate, %]	11279 [85.1]	15224 [77.4]
No. of merged patterns	8579	12180
Resolution range (Å)	29.37–1.80 (1.86–1.80)	39.90–1.70 (1.744–1.704)
Space group	$P6_522$	
a, *b*, *c* (Å)	58.75, 58.75, 192.83	
α, β, γ (°)	90, 90, 120	
Mean $I/\sigma(I)$	5.17 (1.42)	5.42 (0.70)
$CC_{1/2}$	0.951 (0.340)	0.964 (0.160)
CC*	0.9873 (0.740)	0.991 (0.525)
R_{split}† (%)	17.22 (89.08)	16.24 (189.26)
Completeness (%)	99.95 (100.00)	99.99 (99.86)
Multiplicity	107.66 (29.10)	124.12 (16.04)
Refinement		
Resolution (Å)	2.37–1.80 (1.864–1.800)	39.90–1.704 (1.744–1.704)
Total reflections	2079122	2809415
Unique reflections	19213 (1859)	22635 (1469)
Wilson *B* factor (Å²)	17.33	19.11
Reflections used in refinement	19211 (1860)	24698 (1430)
Reflections used for R_{free}	1541 (149)	1963 (114)
R_{work}	0.226 (0.362)	0.219 (0.481)
R_{free}	0.273 (0.366)	0.264 (0.396)
No. of non-H atoms		
Total	1239	1210
Macromolecules	1111	1116
Solvent	128	94
No. of protein residues	141	141
R.m.s.d., bonds (Å)	0.002	0.012
R.m.s.d., angles (°)	0.53	1.316
Ramachandran favoured (%)	97.12	97.20
Ramachandran allowed (%)	2.88	2.82
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.00	0.00
Clashscore	1.34	2.28
Average *B* factor (Å²)		
Overall	24.00	25.30
Macromolecules	20.87	21.19
Solvent	36.48	39.09

† R_{split} as defined in White *et al.* (2012).

conditions in artificial crystallization chambers (Yuan *et al.*, 2003) or within the cytosol and the nucleus of a living insect cell, HEX-1 forms a polymeric helical spiral with 12 monomers per full turn [Supplementary Fig. S5(*a*)]. Interaction of a spiral with six identical neighbours results in the sixfold symmetry of the crystals that is consistently observed in native Woronin bodies (Markham & Collinge, 1987) [Supplementary Fig. S5(*b*)]. Consequently, the intrinsic crystallization tendency of HEX-1 appears to be maintained as long as the environmental pH does not prevent the formation of the intermolecular interaction patterns.

4. Conclusion

In this study, we exploited the advantages of FT-SFX to establish a novel approach for high-resolution structure
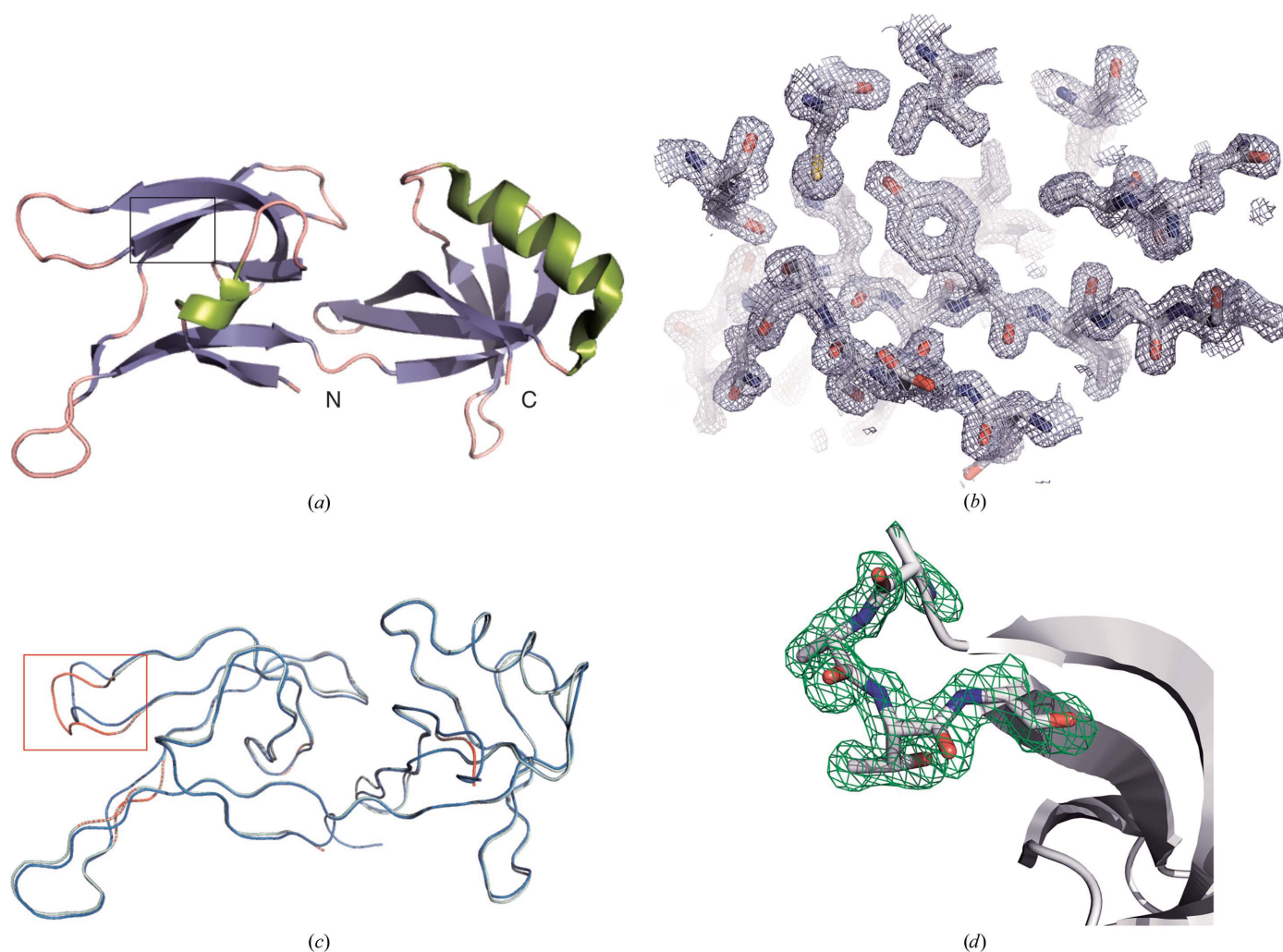
Figure 4
Details of the HEX-1 structure. (*a*) Overall structure of HEX-1 in cartoon representation. (*b*) Representative region of the electron-density map of HEX-1 obtained by FT-SFX of intracellular crystals at 1.8 Å resolution from a single chip (PDB entry 7asx). The detailed region is marked by a black box in (*a*). (*c*) Structural homology of HEX-1 crystallized in insect cells and HEX-1 purified from *E. coli* and crystallized by sitting-drop vapour diffusion. A backbone representation of the HEX-1 structure (green) obtained from *in cellulo* diffraction is superimposed with that of the HEX-1 reference structure (PDB code 1khi; blue). The average r.m.s.d. is 0.47 Å for equivalent C^α atoms. The only region showing major structural differences with r.m.s.d.s above 0.6 Å is highlighted by the red box. (*d*) OMIT map of residues 62–65 of the FT-SFX HEX-1 structure with the same residues in stick representation. The $F_o - F_c$ map (green) of the region of largest r.m.s.d.s with the reference structure, highlighted in (*c*), is contoured at 3σ.

elucidation by *in cellulo* diffraction of micrometre-sized protein crystals within living insect cells. Within 12 min of LCLS beamtime, sufficient diffraction data to elucidate the HEX-1 structure at 1.8 Å resolution were recorded, supported by the sixfold symmetry of the HEX-1 crystals. Pre-positioning of the crystal-containing cells on the silicon chips increased the hit rate by up to 30%, allowing a much more efficient use of the instrument and the sample compared with previous GDVN injection of isolated *in cellulo* grown crystals (Nass *et al.*, 2020; Redecke *et al.*, 2013). The direct use of intact, crystal-containing insect cells avoids the time-consuming isolation and enrichment of the crystalline material, restricting the sample-preparation time to a minimum. Problems associated with the usage of intact cells for liquid-jet injection, particularly settling effects of the large cells in the sample reservoir and instable jets due to clogging of GDVN nozzles by the soft matter of the cells, are prevented. Using crystals in intact cells

for diffraction protects them from environmental stress during transport, chip loading and diffraction, overcoming some of the challenges reported for the handling of pure protein crystals, for example shearing forces during pipetting (Lieske *et al.*, 2019), crystal clustering during removal of the surrounding liquid through the chip pores (Soares *et al.*, 2014) or rapid dehydration after chip loading (Martiel *et al.*, 2019). The soft matter and the large amounts of water within the living cells irradiated by the XFEL pulse did not impact crystal diffraction. Background levels were sufficiently low to recover diffraction signals to high resolution (1.71 Å), as previously observed (Tsutsui *et al.*, 2015; Baskaran *et al.*, 2015; Sawaya *et al.*, 2014). The reason for the occurrence of two dominating populations of hexagonal unit cells in the recorded diffraction patterns of the HEX-1 *in cellulo* crystals that differ only in the length of the *c* axis needs to be further investigated. However, an impact of the different crystal morphologies that have been

observed or the environmental conditions in the cytosol and in the nucleus, where the HEX-1 crystals grow, appears to be reasonable.

Our approach can easily be adopted for other serial diffraction experiments at XFELs for crystalline systems with sufficient diffraction signal at room temperature. The main advantages of fixed-target crystal delivery, lower sample consumption and higher hit rates, have also successfully been exploited for serial diffraction data collection at synchrotron sources, also enabling time-resolved data collection (Roedig *et al.*, 2016; Meents *et al.*, 2017; Owen *et al.*, 2017; Weinert *et al.*, 2017; Schulz *et al.*, 2018; Mehrabi *et al.*, 2019, 2020). Thus, the fixed-target *in cellulo* diffraction approach presented in this study also provides the possibility of improving SSX experiments in terms of sample consumption and data-collection time. This is important as beamtime at an XFEL is very expensive and difficult to obtain. The application of micro-patterned silicon chips for serial collection of diffraction data from crystals directly in living cells offers huge potential for the straightforward structure elucidation of proteins that form intracellular crystals. This includes soluble proteins in general, even those containing post-translational modifications or in complex with small molecules (Redecke *et al.*, 2013; Nass *et al.*, 2020), for which no general limitations have so far been ascertained. Except for the coral fluorescent protein Xpa (Tsutsui *et al.*, 2015) and HEX-1, all recombinant proteins that produced intracellular crystals were identified by accident and were not predicted to crystallize. In contrast, membrane proteins that need to stay embedded in or attached to their respective membranes are suggested to be less accessible for *in cellulo* crystallization, which needs to be studied in detail in the future. Since the living cell represents an environment that is very close to physiological conditions for most proteins, insights into protein dynamics by time-resolved X-ray crystallography (TRX; Orville, 2020; Pearson & Mehrabi, 2020) using intracellular crystals could provide new insights, for example for drug-screening efforts. However, any compounds delivered by photoactivatable caged compounds to induce dynamics need to overcome the cell membrane, representing a significant bottleneck, in addition to other limitations. Protein dynamics directly triggered by light appear to be more accessible if the crystal contacts remain stable during the reaction. Thus, the establishment and validation of TRX using *in cellulo* grown crystals represents a challenging but important task in the future, along with a more detailed investigation of the intracellular crystallization process itself.

5. Related literature

The following reference is cited in the supporting information for this article: Krissinel & Henrick (2007).

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Ayyer, K., Yefanov, O. M., Oberthür, D., Roy-Chowdhury, S., Galli, L., Mariani, V., Basu, S., Coe, J., Conrad, C. E., Fromme, R., Schaffer, A., Dörner, K., James, D., Kupitz, C., Metz, M., Nelson, G., Xavier, P. L., Beyerlein, K. R., Schmidt, M., Sarrou, I., Spence, J. C. H., Weierstall, U., White, T. A., Yang, J.-H., Zhao, Y., Liang, M., Aquila, A., Hunter, M. S., Robinson, J. S., Koglin, J. E., Boutet, S., Fromme, P., Barty, A. & Chapman, H. N. (2016). *Nature*, **530**, 202–206.

Baskaran, Y., Ang, K. C., Anekal, P. V., Chan, W. L., Grimes, J. M., Manser, E. & Robinson, R. C. (2015). *Nat. Commun.* **6**, 8681.

Baxter, E. L., Aguila, L., Alonso-Mori, R., Barnes, C. O., Bonagura, C. A., Brehmer, W., Brunger, A. T., Calero, G., Caradoc-Davies, T. T., Chatterjee, R., Degrado, W. F., Fraser, J. M., Ibrahim, M., Kern, J., Kobilka, B. K., Kruse, A. C., Larsson, K. M., Lemke, H. T., Lyubimov, A. Y., Manglik, A., McPhillips, S. E., Norgren, E., Pang, S. S., Soltis, S. M., Song, J., Thomaston, J., Tsai, Y., Weis, W. I., Woldeyes, R. A., Yachandra, V., Yano, J., Zouni, A. & Cohen, A. E. (2016). *Acta Cryst.* D**72**, 2–11.

Beyerlein, K. R., Dierksmeyer, D., Mariani, V., Kuhn, M., Sarrou, I., Ottaviano, A., Awel, S., Knoska, J., Fuglerud, S., Jönsson, O., Stern, S., Wiedorn, M. O., Yefanov, O., Adriano, L., Bean, R., Burkhardt, A., Fischer, P., Heymann, M., Horke, D. A., Jungnickel, K. E. J., Kovaleva, E., Lorbeer, O., Metz, M., Meyer, J., Morgan, A., Pande, K., Panneerselvam, S., Seuring, C., Tolstikova, A., Lieske, J., Aplin, S., Roessle, M., White, T. A., Chapman, H. N., Meents, A. & Oberthuer, D. (2017). *IUCrJ*, **4**, 769–777.

Blaj, G., Caragiulo, P., Carini, G., Carron, S., Dragone, A., Freytag, D., Haller, G., Hart, P., Hasi, J., Herbst, R., Herrmann, S., Kenney, C., Markovic, B., Nishimura, K., Osier, S., Pines, J., Reese, B., Segal, J., Tomada, A. & Weaver, M. (2015). *J. Synchrotron Rad.* **22**, 577–583.

Bogan, M. J., Benner, W. H., Boutet, S., Rohner, U., Frank, M., Barty, A., Seibert, M. M., Maia, F., Marchesini, S., Bajt, S., Woods, B., Riot, V., Hau-Riege, S. P., Svenda, M., Marklund, E., Spiller, E., Hajdu, J. & Chapman, H. N. (2008). *Nano Lett.* **8**, 310–316.

Boudes, M., Garriga, D., Fryga, A., Caradoc-Davies, T. & Coulibaly, F. (2016). *Acta Cryst.* D**72**, 576–585.

Boutet, S., Lomb, L., Williams, G. J., Barends, T. R. M., Aquila, A., Doak, R. B., Weierstall, U., DePonte, D. P., Steinbrener, J., Shoeman, R. L., Messerschmidt, M., Barty, A., White, T. A., Kassemeyer, S., Kirian, R. A., Seibert, M. M., Montanez, P. A., Kenney, C., Herbst, R., Hart, P., Pines, J., Haller, G., Gruner, S. M.,

Philipp, H. T., Tate, M. W., Hromalik, M., Koerner, L. J., van Bakel, N., Morse, J., Ghonsalves, W., Arnlund, D., Bogan, M. J., Caleman, C., Fromme, R., Hampton, C. Y., Hunter, M. S., Johansson, L. C., Katona, G., Kupitz, C., Liang, M., Martin, A. V., Nass, K., Redecke, L., Stellato, F., Timneanu, N., Wang, D., Zatsepin, N. A., Schafer, D., Defever, J., Neutze, R., Fromme, P., Spence, J. C. H., Chapman, H. N. & Schlichting, I. (2012). *Science*, **337**, 362–364.

Calvey, G. D., Katz, A. M., Schaffer, C. B. & Pollack, L. (2016). *Struct. Dyn.* **3**, 054301.

Chapman, H. N., Caleman, C. & Timneanu, N. (2011). *Philos. Trans. R. Soc. B*, **369**, 20130313.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Conrad, C. E., Basu, S., James, D., Wang, D., Schaffer, A., Roy-Chowdhury, S., Zatsepin, N. A., Aquila, A., Coe, J., Gati, C., Hunter, M. S., Koglin, J. E., Kupitz, C., Nelson, G., Subramanian, G., White, T. A., Zhao, Y., Zook, J., Boutet, S., Cherezov, V., Spence, J. C. H., Fromme, R., Weierstall, U. & Fromme, P. (2015). *IUCrJ*, **2**, 421–430.

Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature*, **446**, 97–101.

DePonte, D. P., Mckeown, J. T., Weierstall, U., Doak, R. B. & Spence, J. C. H. (2011). *Ultramicroscopy*, **111**, 824–827.

DePonte, D. P., Weierstall, U., Schmidt, K., Warner, J., Starodub, D., Spence, J. C. H. & Doak, R. B. (2008). *J. Phys. D Appl. Phys.* **41**, 195505.

Duisenberg, A. J. M. (1992). *J. Appl. Cryst.* **25**, 92–96.

Duszenko, M., Redecke, L., Mudogo, C. N., Sommer, B. P., Mogk, S., Oberthuer, D. & Betzel, C. (2015). *Acta Cryst.* F**71**, 929–937.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Feld, G. K., Heymann, M., Benner, W. H., Pardini, T., Tsai, C.-J., Boutet, S., Coleman, M. A., Hunter, M. S., Li, X., Messerschmidt, M., Opathalage, A., Pedrini, B., Williams, G. J., Krantz, B. A., Fraden, S., Hau-Riege, S., Evans, J. E., Segelke, B. W. & Frank, M. (2015). *J. Appl. Cryst.* **48**, 1072–1079.

Frank, M., Carlson, D. B., Hunter, M. S., Williams, G. J., Messerschmidt, M., Zatsepin, N. A., Barty, A., Benner, W. H., Chu, K., Graf, A. T., Hau-Riege, S. P., Kirian, R. A., Padeste, C., Pardini, T., Pedrini, B., Segelke, B., Seibert, M. M., Spence, J. C. H., Tsai, C.-J., Lane, S. M., Li, X.-D., Schertler, G., Boutet, S., Coleman, M. & Evans, J. E. (2014). *IUCrJ*, **1**, 95–100.

Gallat, F.-X., Matsugaki, A., Coussens, N. P., Yagi, K. J., Boudes, M., Higashi, T., Tsuji, D., Tatano, Y., Suzuki, M., Mizohata, E., Tono, K., Joti, Y., Kameshima, T., Park, J., Song, C., Hatsui, T., Yabashi, M., Nango, E., Itoh, K., Coulibaly, F., Tobe, S., Ramaswamy, S., Stay, B., Iwata, S. & Chavas, L. M. G. (2014). *Philos. Trans. R. Soc. B*, **369**, 20130497.

Gati, C., Bourenkov, G., Klinge, M., Rehders, D., Stellato, F., Oberthür, D., Yefanov, O., Sommer, B. P., Mogk, S., Duszenko, M., Betzel, C., Schneider, T. R., Chapman, H. N. & Redecke, L. (2014). *IUCrJ*, **1**, 87–94.

Gati, C., Oberthuer, D., Yefanov, O., Bunker, R. D., Stellato, F., Chiu, E., Yeh, S.-M., Aquila, A., Basu, S., Bean, R., Beyerlein, K. R., Botha, S., Boutet, S., DePonte, D. P., Doak, R. B., Fromme, R., Galli, L., Grotjohann, I., James, D. R., Kupitz, C., Lomb, L., Messerschmidt, M., Nass, K., Rendek, K., Shoeman, R. L., Wang, D., Weierstall, U., White, T. A., Williams, G. J., Zatsepin, N. A., Fromme, P., Spence, J. C. H., Goldie, K. N., Jehle, J. A., Metcalf, P., Barty, A. & Chapman, H. N. (2017). *Proc. Natl Acad. Sci. USA*, **114**, 2247–2252.

Ginn, H. M., Messerschmidt, M., Ji, X., Zhang, H., Axford, D., Gildea, R. J., Winter, G., Brewster, A. S., Hattne, J., Wagner, A., Grimes, J. M., Evans, G., Sauter, N. K., Sutton, G. & Stuart, D. I. (2015). *Nat. Commun.* **6**, 6435.

Hantke, M. F., Hasse, D., Maia, F. R. N. C., Ekeberg, T., John, K., Svenda, M., Loh, N. D., Martin, A. V., Timneanu, N., Larsson,

D. S. D., van der Schot, G., Carlsson, G. H., Ingelman, M., Andreasson, J., Westphal, D., Liang, M., Stellato, F., DePonte, D. P., Hartmann, R., Kimmel, N., Kirian, R. A., Seibert, M. M., Mühlig, K., Schorb, S., Ferguson, K., Bostedt, C., Carron, S., Bozek, J. D., Rolles, D., Rudenko, A., Epp, S., Chapman, H. N., Barty, A., Hajdu, J. & Andersson, I. (2014). *Nat. Photon.* **8**, 943–949.

Hunter, M. S., Segelke, B., Messerschmidt, M., Williams, G. J., Zatsepin, N. A., Barty, A., Benner, W. H., Carlson, D. B., Coleman, M., Graf, A., Hau-Riege, S. P., Pardini, T., Seibert, M. M., Evans, J., Boutet, S. & Frank, M. (2015). *Sci. Rep.* **4**, 6026.

Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. (2017). *Trends Biochem. Sci.* **42**, 749–762.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Karpik, A., Martiel, I., Kristiansen, P. M. & Padeste, C. (2020). *Micro Nano Eng.* **7**, 100053.

Knoška, J., Adriano, L., Awel, S., Beyerlein, K. R., Yefanov, O., Oberthuer, D., Peña Murillo, G. E., Roth, N., Sarrou, I., Villanueva-Perez, P., Wiedorn, M. O., Wilde, F., Bajt, S., Chapman, H. N. & Heymann, M. (2020). *Nat. Commun.* **11**, 657.

Koopmann, R., Cupelli, K., Redecke, L., Nass, K., DePonte, D. P., White, T. A., Stellato, F., Rehders, D., Liang, M., Andreasson, J., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Boutet, S., Bozek, J. D., Caleman, C., Coppola, N., Davidsson, J., Doak, R. B., Ekeberg, T., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Hunter, M. S., Kassemeyer, S., Kirian, R. A., Lomb, L., Maia, F. R. N. C., Kimmel, N., Martin, A. V., Messerschmidt, M., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R. G., Soltau, H., Stern, S., Strüder, L., Timneanu, N., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Williams, G. J., Wunderer, C. B., Fromme, P., Spence, J. C. H., Stehle, T., Chapman, H. N., Betzel, C. & Duszenko, M. (2012). *Nat. Methods*, **9**, 259–262.

Kovácsová, G., Grünbein, M. L., Kloos, M., Barends, T. R. M., Schlesinger, R., Heberle, J., Kabsch, W., Shoeman, R. L., Doak, R. B. & Schlichting, I. (2017). *IUCrJ*, **4**, 400–410.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N. A., Rendek, K. N., Hunter, M. S., Shoeman, R. L., White, T. A., Wang, D., James, D., Yang, J. H., Cobb, D. E., Reeder, B., Sierra, R. G., Liu, H., Barty, A., Aquila, A. L., DePonte, D., Kirian, R. A., Bari, S., Bergkamp, J. J., Beyerlein, K. R., Bogan, M. J., Caleman, C., Chao, T. C., Conrad, C. E., Davis, K. M., Fleckenstein, H., Galli, L., Hau-Riege, S. P., Kassemeyer, S., Laksmono, H., Liang, M., Lomb, L., Marchesini, S., Martin, A. V., Messerschmidt, M., Milathianaki, D., Nass, K., Ros, A., Roy-Chowdhury, S., Schmidt, K., Seibert, M., Steinbrener, J., Stellato, F., Yan, L., Yoon, C., Moore, T. A., Moore, A. L., Pushkar, Y., Williams, G. J., Boutet, S., Doak, R. B., Weierstall, U., Frank, M., Chapman, H. N., Spence, J. C. H. & Fromme, P. (2014). *Nature*, **513**, 261–265.

Lahey-Rudolph, J. M., Schönherr, R., Jeffries, C. M., Blanchet, C. E., Boger, J., Ferreira Ramos, A. S., Riekehr, W. M., Triandafillidis, D.-P., Valmas, A., Margiolaki, I., Svergun, D. & Redecke, L. (2020). *J. Appl. Cryst.* **53**, 1169–1180.

Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.

Lieske, J., Cerv, M., Kreida, S., Komadina, D., Fischer, J., Barthelmess, M., Fischer, P., Pakendorf, T., Yefanov, O., Mariani, V., Seine, T., Ross, B. H., Crosas, E., Lorbeer, O., Burkhardt, A., Lane, T. J., Guenther, S., Bergtholdt, J., Schoen, S., Törnroth-Horsefield, S., Chapman, H. N. & Meents, A. (2019). *IUCrJ*, **6**, 714–728.

Mariani, V., Morgan, A., Yoon, C. H., Lane, T. J., White, T. A., O'Grady, C., Kuhn, M., Aplin, S., Koglin, J., Barty, A. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 1073–1080.

Markham, P. & Collinge, A. J. (1987). *FEMS Microbiol. Rev.* **46**, 1–11.

Martiel, I., Müller-Werkmeister, H. M. & Cohen, A. E. (2019). *Acta Cryst.* D**75**, 160–177.

J. Mia Lahey-Rudolph *et al.* • Fixed-target SFX using *in cellulo* grown microcrystals **675**

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Meents, A., Wiedorn, M. O., Srajer, V., Henning, R., Sarrou, I., Bergtholdt, J., Barthelmess, M., Reinke, P. Y. A., Dierksmeyer, D., Tolstikova, A., Schaible, S., Messerschmidt, M., Ogata, C. M., Kissick, D. J., Taft, M. H., Manstein, D. J., Lieske, J., Oberthuer, D., Fischetti, R. F. & Chapman, H. N. (2017). *Nat. Commun.* **8**, 1281.

Mehrabi, P., Müller-Werkmeister, H. M., Leimkohl, J.-P., Schikora, H., Ninkovic, J., Krivokuca, S., Andriček, L., Epp, S. W., Sherrell, D., Owen, R. L., Pearson, A. R., Tellkamp, F., Schulz, E. C. & Miller, R. J. D. (2020). *J. Synchrotron Rad.* **27**, 360–370.

Mehrabi, P., Schulz, E. C., Dsouza, R., Müller-Werkmeister, H. M., Tellkamp, F., Miller, R. J. D. & Pai, E. F. (2019). *Science*, **365**, 1167–1170.

Mueller, C., Marx, A., Epp, S. W., Zhong, Y., Kuo, A., Balo, A. R., Soman, J., Schotte, F., Lemke, H. T., Owen, R. L., Pai, E. F., Pearson, A. R., Olson, J. S., Anfinrud, P. A., Ernst, O. P. & Miller, R. J. D. (2015). *Struct. Dyn.* **2**, 054302.

Murray, T. D., Lyubimov, A. Y., Ogata, C. M., Vo, H., Uervirojnang-koorn, M., Brunger, A. T. & Berger, J. M. (2015). *Acta Cryst.* D**71**, 1987–1997.

Nango, E., Royant, A., Kubo, M., Nakane, T., Wickstrand, C., Kimura, T., Tanaka, T., Tono, K., Song, C., Tanaka, R., Arima, T., Yamashita, A., Kobayashi, J., Hosaka, T., Mizohata, E., Nogly, P., Sugahara, M., Nam, D., Nomura, T., Shimamura, T., Im, D., Fujiwara, T., Yamanaka, Y., Jeon, B., Nishizawa, T., Oda, K., Fukuda, M., Andersson, R., Båth, P., Dods, R., Davidsson, J., Matsuoka, S., Kawatake, S., Murata, M., Nureki, O., Owada, S., Kameshima, T., Hatsui, T., Joti, Y., Schertler, G., Yabashi, M., Bondar, A.-N., Standfuss, J., Neutze, R. & Iwata, S. (2016). *Science*, **354**, 1552–1557.

Nass, K., Redecke, L., Perbandt, M., Yefanov, O., Klinge, M., Koopmann, R., Stellato, F., Gabdulkhakov, A., Schönherr, R., Rehders, D., Lahey-Rudolph, J. M., Aquila, A., Barty, A., Basu, S., Doak, R. B., Duden, R., Frank, M., Fromme, R., Kassemeyer, S., Katona, G., Kirian, R., Liu, H., Majoul, I., Martin-Garcia, J. M., Messerschmidt, M., Shoeman, R. L., Weierstall, U., Westenhoff, S., White, T. A., Williams, G. J., Yoon, C. H., Zatsepin, N., Fromme, P., Duszenko, M., Chapman, H. N. & Betzel, C. (2020). *Nat. Commun.* **11**, 620.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.

Oberthuer, D., Knoška, J., Wiedorn, M. O., Beyerlein, K. R., Bushnell, D. A., Kovaleva, E. G., Heymann, M., Gumprecht, L., Kirian, R. A., Barty, A., Mariani, V., Tolstikova, A., Adriano, L., Awel, S., Barthelmess, M., Dörner, K., Xavier, P. L., Yefanov, O., James, D. R., Nelson, G., Wang, D., Calvey, G., Chen, Y., Schmidt, A., Szczepek, M., Frielingsdorf, S., Lenz, O., Snell, E., Robinson, P. J., Šarler, B., Belšak, G., Maček, M., Wilde, F., Aquila, A., Boutet, S., Liang, M., Hunter, M. S., Scheerer, P., Lipscomb, J. D., Weierstall, U., Kornberg, R. D., Spence, J. C. H., Pollack, L., Chapman, H. N. & Bajt, S. (2017). *Sci. Rep.* **7**, 44628.

Orville, A. M. (2020). *Curr. Opin. Struct. Biol.* **65**, 193–208.

Owen, R. L., Axford, D., Sherrell, D. A., Kuo, A., Ernst, O. P., Schulz, E. C., Miller, R. J. D. & Mueller-Werkmeister, H. M. (2017). *Acta Cryst.* D**73**, 373–378.

Park, J., Park, S., Kim, J., Park, G., Cho, Y. & Nam, K. H. (2019). *Sci. Rep.* **9**, 2525.

Pearson, A. R. & Mehrabi, P. (2020). *Curr. Opin. Struct. Biol.* **65**, 168–174.

Redecke, L., Nass, K., DePonte, D. P., White, T. A., Rehders, D., Barty, A., Stellato, F., Liang, M., Barends, T. R. M., Boutet, S., Williams, G. J., Messerschmidt, M., Seibert, M. M., Aquila, A., Arnlund, D., Bajt, S., Barth, T., Bogan, M. J., Caleman, C., Chao, T.-C., Doak, R. B., Fleckenstein, H., Frank, M., Fromme, R., Galli, L., Grotjohann, I., Hunter, M. S., Johansson, L. C., Kassemeyer, S., Katona, G., Kirian, R. A., Koopmann, R., Kupitz, C., Lomb, L., Martin, A. V., Mogk, S., Neutze, R., Shoeman, R. L., Steinbrener, J.,

Timneanu, N., Wang, D., Weierstall, U., Zatsepin, N. A., Spence, J. C. H., Fromme, P., Schlichting, I., Duszenko, M., Betzel, C. & Chapman, H. N. (2013). *Science*, **339**, 227–230.

Reed, L. J. & Muench, H. (1938). *Am. J. Epidemiol.* **27**, 493–497.

Roedig, P., Duman, R., Sanchez-Weatherby, J., Vartiainen, I., Burkhardt, A., Warmer, M., David, C., Wagner, A. & Meents, A. (2016). *J. Appl. Cryst.* **49**, 968–975.

Roedig, P., Ginn, H. M., Pakendorf, T., Sutton, G., Harlos, K., Walter, T. S., Meyer, J., Fischer, P., Duman, R., Vartiainen, I., Reime, B., Warmer, M., Brewster, A. S., Young, I. D., Michels-Clark, T., Sauter, N. K., Kotecha, A., Kelly, J., Rowlands, D. J., Sikorsky, M., Nelson, S., Damiani, D. S., Alonso-Mori, R., Ren, J., Fry, E. E., David, C., Stuart, D. I., Wagner, A. & Meents, A. (2017). *Nat. Methods*, **14**, 805–810.

Roedig, P., Vartiainen, I., Duman, R., Panneerselvam, S., Stübe, N., Lorbeer, O., Warmer, M., Sutton, G., Stuart, D. I., Weckert, E., David, C., Wagner, A. & Meents, A. (2015). *Sci. Rep.* **5**, 10451.

Roessler, C. G., Kuczewski, A., Stearns, R., Ellson, R., Olechno, J., Orville, A. M., Allaire, M., Soares, A. S. & Héroux, A. (2013). *J. Synchrotron Rad.* **20**, 805–808.

Sawaya, M. R., Cascio, D., Gingery, M., Rodriguez, J., Goldschmidt, L., Colletier, J.-P., Messerschmidt, M. M., Boutet, S., Koglin, J. E., Williams, G. J., Brewster, A. S., Nass, K., Hattne, J., Botha, S., Doak, R. B., Shoeman, R. L., DePonte, D. P., Park, H.-W., Federici, B. A., Sauter, N. K., Schlichting, I. & Eisenberg, D. S. (2014). *Proc. Natl Acad. Sci.* **111**, 12769–12774.

Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.

Schönherr, R., Klinge, M., Rudolph, J. M., Fita, K., Rehders, D., Lübber, F., Schneegans, S., Majoul, I. V., Duszenko, M., Betzel, C., Brandariz-Nuñez, A., Martinez-Costas, J., Duden, R. & Redecke, L. (2015). *Struct. Dyn.* **2**, 041712.

Schönherr, R., Rudolph, J. M. & Redecke, L. (2018). *Biol. Chem.* **399**, 751–772.

Schulz, E. C., Mehrabi, P., Müller-Werkmeister, H. M., Tellkamp, F., Jha, A., Stuart, W., Persch, E., De Gasparo, R., Diederich, F., Pai, E. F. & Miller, R. J. D. (2018). *Nat. Methods*, **15**, 901–904.

Shelby, M. L., Gilbile, D., Grant, T. D., Seuring, C., Segelke, B. W., He, W., Evans, A. C., Pakendorf, T., Fischer, P., Hunter, M. S., Batyuk, A., Barthelmess, M., Meents, A., Coleman, M. A., Kuhl, T. L. & Frank, M. (2020). *IUCrJ*, **7**, 30–41.

Shimada, A., Kubo, M., Baba, S., Yamashita, K., Hirata, K., Ueno, G., Nomura, T., Kimura, T., Shinzawa-Itoh, K., Baba, J., Hatano, K., Eto, Y., Miyamoto, A., Murakami, H., Kumasaka, T., Owada, S., Tono, K., Yabashi, M., Yamaguchi, Y., Yanagisawa, S., Sakaguchi, M., Ogura, T., Komiya, R., Yan, J., Yamashita, E., Yamamoto, M., Ago, H., Yoshikawa, S. & Tsukihara, T. (2017). *Sci. Adv.* **3**, e1603042.

Sierra, R. G., Batyuk, A., Sun, Z., Aquila, A., Hunter, M. S., Lane, T. J., Liang, M., Yoon, C. H., Alonso-Mori, R., Armenta, R., Castagna, J.-C., Hollenbeck, M., Osier, T. O., Hayes, M., Aldrich, J., Curtis, R., Koglin, J. E., Rendahl, T., Rodriguez, E., Carbajo, S., Guillet, S., Paul, R., Hart, P., Nakahara, K., Carini, G., DeMirci, H., Dao, E. H., Hayes, B. M., Rao, Y. P., Chollet, M., Feng, Y., Fuller, F. D., Kupitz, C., Sato, T., Seaberg, M. H., Song, S., van Driel, T. B., Yavas, H., Zhu, D., Cohen, A. E., Wakatsuki, S. & Boutet, S. (2019). *J. Synchrotron Rad.* **26**, 346–357.

Soares, A. S., Mullen, J. D., Parekh, R. M., McCarthy, G. S., Roessler, C. G., Jackimowicz, R., Skinner, J. M., Orville, A. M., Allaire, M. & Sweet, R. M. (2014). *J. Synchrotron Rad.* **21**, 1231–1239.

Spence, J. C. H. (2020). *Crystals*, **10**, 135.

Standfuss, J. & Spence, J. (2017). *IUCrJ*, **4**, 100–101.

Stellato, F., Oberthür, D., Liang, M., Bean, R., Gati, C., Yefanov, O., Barty, A., Burkhardt, A., Fischer, P., Galli, L., Kirian, R. A., Meyer, J., Panneerselvam, S., Yoon, C. H., Chervinskii, F., Speller, E., White, T. A., Betzel, C., Meents, A. & Chapman, H. N. (2014). *IUCrJ*, **1**, 204–212.

Sugahara, M., Mizohata, E., Nango, E., Suzuki, M., Tanaka, T., Masuda, T., Tanaka, R., Shimamura, T., Tanaka, Y., Suno, C., Ihara,

K., Pan, D., Kakinouchi, K., Sugiyama, S., Murata, M., Inoue, T., Tono, K., Song, C., Park, J., Kameshima, T., Hatsui, T., Joti, Y., Yabashi, M. & Iwata, S. (2015). *Nat. Methods*, **12**, 61–63.

Tenboer, J., Basu, S., Zatsepin, N., Pande, K., Milathianaki, D., Frank, M., Hunter, M., Boutet, S., Williams, G. J., Koglin, J. E., Oberthuer, D., Heymann, M., Kupitz, C., Conrad, C., Coe, J., Roy-Chowdhury, S., Weierstall, U., James, D., Wang, D., Grant, T., Barty, A., Yefanov, O., Scales, J., Gati, C., Seuring, C., Srajer, V., Henning, R., Schwander, P., Fromme, R., Ourmazd, A., Moffat, K., Van Thor, J. J., Spence, J. C. H., Fromme, P., Chapman, H. N. & Schmidt, M. (2014). *Science*, **346**, 1242–1246.

Tenney, K., Hunt, I., Sweigard, J., Pounder, J. I., McClain, C., Bowman, E. J. & Bowman, B. J. (2000). *Fungal Genet. Biol.* **31**, 205–217.

Tolstikova, A., Levantino, M., Yefanov, O., Hennicke, V., Fischer, P., Meyer, J., Mozzanica, A., Redford, S., Crosas, E., Opara, N. L., Barthelmess, M., Lieske, J., Oberthuer, D., Wator, E., Mohacsi, I., Wulff, M., Schmitt, B., Chapman, H. N. & Meents, A. (2019). *IUCrJ*, **6**, 927–937.

Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. (2010). *J. Struct. Biol.* **172**, 45–54.

Tsutsui, H., Jinno, Y., Shoda, K., Tomita, A., Matsuda, M., Yamashita, E., Katayama, H., Nakagawa, A. & Miyawaki, A. (2015). *Mol. Cell*, **58**, 186–193.

Urzhumtseva, L., Afonine, P. V., Adams, P. D. & Urzhumtsev, A. (2009). *Acta Cryst.* D**65**, 297–300.

Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., Spence, J. C. H., Doak, R. B., Nelson, G., Fromme, P., Fromme, R., Grotjohann, I., Kupitz, C., Zatsepin, N. A., Liu, H., Basu, S., Wacker, D., Han, G. W., Katritch, V., Boutet, S., Messerschmidt, M., Williams, G. J., Koglin, J. E., Seibert, M. M., Klinker, M., Gati, C., Shoeman, R. L., Barty, A., Chapman, H. N., Kirian, R. A., Beyerlein, K. R., Stevens, R. C., Li, D., Shah, S. T. A., Howe, N., Caffrey, M. & Cherezov, V. (2014). *Nat. Commun.* **5**, 3309.

Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D., Gashi, D., Vera, L., Marsh, M., Jaeger, K., Dworkowski, F., Panepucci, E., Basu, S., Skopintsev, P., Doré, A. S., Geng, T., Cooke, R. M., Liang, M., Prota, A. E., Panneels, V., Nogly, P., Ermler, U., Schertler, G., Hennig, M., Steinmetz, M. O., Wang, M. & Standfuss, J. (2017). *Nat. Commun.* **8**, 542.

White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.

White, T. A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K. R., Chervinskii, F., Galli, L., Gati, C., Nakane, T., Tolstikova, A., Yamashita, K., Yoon, C. H., Diederichs, K. & Chapman, H. N. (2016). *J. Appl. Cryst.* **49**, 680–689.

Wiedorn, M. O., Awel, S., Morgan, A. J., Ayyer, K., Gevorkov, Y., Fleckenstein, H., Roth, N., Adriano, L., Bean, R., Beyerlein, K. R.,

Chen, J., Coe, J., Cruz-Mazo, F., Ekeberg, T., Graceffa, R., Heymann, M., Horke, D. A., Knoška, J., Mariani, V., Nazari, R., Oberthür, D., Samanta, A. K., Sierra, R. G., Stan, C. A., Yefanov, O., Rompotis, D., Correa, J., Erk, B., Treusch, R., Schulz, J., Hogue, B. G., Gañán-Calvo, A. M., Fromme, P., Küpper, J., Rode, A. V., Bajt, S., Kirian, R. A. & Chapman, H. N. (2018). *IUCrJ*, **5**, 574–584.

Wiedorn, M. O., Oberthür, D., Bean, R., Schubert, R., Werner, N., Abbey, B., Aepfelbacher, M., Adriano, L., Allahgholi, A., Al-Qudami, N., Andreasson, J., Aplin, S., Awel, S., Ayyer, K., Bajt, S., Barák, I., Bari, S., Bielecki, J., Botha, S., Boukhelef, D., Brehm, W., Brockhauser, S., Cheviakov, I., Coleman, M. A., Cruz-Mazo, F., Danilevski, C., Darmanin, C., Doak, R. B., Domaracky, M., Dörner, K., Du, Y., Fangohr, H., Fleckenstein, H., Frank, M., Fromme, P., Gañán-Calvo, A. M., Gevorkov, Y., Giewekemeyer, K., Ginn, H. M., Graafsma, H., Graceffa, R., Greiffenberg, D., Gumprecht, L., Göttlicher, P., Hajdu, J., Hauf, S., Heymann, M., Holmes, S., Horke, D. A., Hunter, M. S., Imlau, S., Kaukher, A., Kim, Y., Klyuev, A., Knoška, J., Kobe, B., Kuhn, M., Kupitz, C., Küpper, J., Lahey-Rudolph, J. M., Laurus, T., Le Cong, K., Letrun, R., Xavier, P. L., Maia, L., Maia, F. R. N. C., Mariani, V., Messerschmidt, M., Metz, M., Mezza, D., Michelat, T., Mills, G., Monteiro, D. C. F., Morgan, A., Mühlig, K., Munke, A., Münnich, A., Nette, J., Nugent, K. A., Nuguid, T., Orville, A. M., Pandey, S., Pena, G., Villanueva-Perez, P., Poehlsen, J., Previtali, G., Redecke, L., Riekehr, W. M., Rohde, H., Round, A., Safenreiter, T., Sarrou, I., Sato, T., Schmidt, M., Schmitt, B., Schönherr, R., Schulz, J., Sellberg, J. A., Seibert, M. M., Seuring, C., Shelby, M. L., Shoeman, R. L., Sikorski, M., Silenzi, A., Stan, C. A., Shi, X., Stern, S., Sztuk-Dambietz, J., Szuba, J., Tolstikova, A., Trebbin, M., Trunk, U., Vagovic, P., Ve, T., Weinhausen, B., White, T. A., Wrona, K., Xu, C., Yefanov, O., Zatsepin, N., Zhang, J., Perbandt, M., Mancuso, A. P., Betzel, C., Chapman, H. & Barty, A. (2018). *Nat. Commun.* **9**, 4025.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

Yabashi, M. & Tanaka, H. (2017). *Nat. Photon.* **11**, 12–14.

Yamamoto, M., Hirata, K., Yamashita, K., Hasegawa, K., Ueno, G., Ago, H. & Kumasaka, T. (2017). *IUCrJ*, **4**, 529–539.

Yefanov, O., Mariani, V., Gati, C., White, T. A., Chapman, H. N. & Barty, A. (2015). *Opt. Express*, **23**, 28459.

Yuan, P., Jedd, G., Kumaran, D., Swaminathan, S., Shio, H., Hewitt, D., Chua, N.-H. & Swaminathan, K. (2003). *Nat. Struct. Mol. Biol.* **10**, 264–270.

Zarrine-Afsar, A., Barends, T. R. M., Müller, C., Fuchs, M. R., Lomb, L., Schlichting, I. & Miller, R. J. D. (2012). *Acta Cryst.* D**68**, 321–323.

3.3.3 Synopsis

The availability of beamtime is better at synchrotron sources than at XFELs, where access is limited to a small user group, and routine use thus prevented.[256] However, even for static structures, outrunning radiation damage while benefiting from the more intense and ultra-short X-ray pulses makes SFX an attractive method. At a synchrotron source the radiation dose must not exceed approximately 30 MGy at cryogenic temperatures, and the radiation limit is even lower at RT.[257] Otherwise, the elucidated structure from the recorded diffraction data can show artefacts of radiation damage, rather than the native fold.[258] Thus, using synchrotron radiation there is a minimal size for small crystals where the diffraction volume limits detectable resolution. The exact minimal volume depends on many factors, most importantly crystal quality and beamline characteristics (**Table 2**); This situation is different if XFEL radiation is used[55,259]:

Due to its extreme brightness, the energy of a focused XFEL pulse destroys any interacting material in tens to hundreds of femtoseconds. However, prior to the destruction of the sample, its almost damage-free diffraction can be captured on the detector.[55,260] *In cellulo* crystals are often in the µm-range or smaller[71], which limits the exploitation of synchrotron radiation, as previously discussed. In 2012, Lars Redecke demonstrated with his working group that CatB crystals grown in rBV infected Sf9 insect cells can be isolated and used for structure elucidation using XFEL radiation.[41,148] The method was optimized in this work towards structure elucidation of protein crystals grown in insect cells using an XFEL, but with crystal-delivery still inside the cells.

Efficacy comparison of FT and GDVN liquid jet sample delivery for SFX

GDVN-jetting SFX tests of PFA-fixated HEX-1 cyto crystal containing-High Five cells demonstrated that diffraction can be recorded this way. Crosslinking introduces covalent bonds at random positions that are physically near[261]; This explains the moderate loss of diffraction power (apparent 2.2 Å) compared to the FT-SFX experiment (structure solved at 1.7 Å, PDB 7ASI) and even compared to the later solved HEX-1 cyto structure with *in cellulo* SSX data (PDB 8CD4, solved at 1.83 Å). My results show that liquid jet injection is not well suited to deliver densely packed circa 30 µm-sized particles like infected insect cells that are prone to co-adhesion. A general drawback of liquid jets is the large sample consumption[244], since the jet is continuous while the pulse is not. Therefore, an immense quantity of crystals are being jetted but not irradiated between subsequent X-ray pulses - in particular between seperate X-ray pulse trains at the EuXFEL.[244] Consequently, upscaling of cell culture and *in cellulo* crystallization may be necessary, and was established for measurements at the EuXFEL. While upscaling rBV infection of insect cells[262] and upscaling *in cellulo* crystallization[249] is certainly possible, the process is less efficient if isolated crystals are required. A sample flow-back has not been established so far.

This might be worth to consider if crystallization in quantity is a particular challenge, however, reusing once-jettet sample would also introduce the risk to gather data of damaged crystals and increase the chance of nozzle-clogging.

Fixed-target SFX, on the other hand, yields high hit-rates with low sample consumption. This is consistent across multiple fixed-target setups, like silicone nitride chips, microfluidic devices within a graphene sandwich, or polymer-based supports.[183,263–267] A relevant factor for effective FT-SFX experiments are fast and precise automated positioning goniometers that allow alignment of the crystals fixed to the support medium. For scarce protein samples, it is possible, but time-consuming in the off-beam time, to locate and optimally position each crystal before the X-ray exposure. Using only one micropatterned silicon chip attached to the high-speed Roadrunner II goniometer[242,268], on which rBV-infeced insect cells from three 6-wells were combined (3×10^6 High Five cells), it was possible to collect a complete dataset within only 12 minutes. Subsequently the X-ray structure of HEX-1 cyto could be solved to 1.75 Å resolution.[183]

A complete dataset with the hitrate achieved at the second EuXFEL beamtime (max. 1 % estimation) using the GDNV liquid jet sample application, to a comparable multiplicity with ~ 12 000 indexed lattices, would require at least 255×10^6 High Five cells. In other words, the fixed-target approach requires 85 times less cells, not counting cell loss due to filtering in preparation for jetting.

For other sample types the liquid jet approach has resulted in many excellent results, in particular with adaptations like the lipidic cubic phase (LCP) jet that result in a slower flow-rate.[54,234,251,269] For SFX crystal delivery *in cellulo* a fixed-target approach appears best suited.[183]

3.4 Alternative phasing strategies: staining with heavy atoms *in cellulo*

The InCellCryst pipeline[181], regardless of the used data collection method, is limited to structures that can be solved by MR. In effect, it depends on the availability of a structure homologue from which the phase can be derived. Towards *ab initio* phasing of intracellular protein crystals by SAD, two approaches were investigated in this thesis.

3.4.1 First heavy atom staining approach: Fusion of a lanthanide binding tag to the target protein for soaking

The first concept idea is staining with terbium ions, a rare-earth lanthanide that binds to the protein of interest via a fused LBT (methods, 5.2.11). To achieve metal ion binding, crystals can be directly soaked with terbium ions after cell lysis. Alternatively, a cell penetrating peptide (CPP) fused to a terbium-saturated LBT could be used to shuttle the lanthanides across the plasma membrane and into the living cell. Prerequisite is a lanthanide binding affinity smaller than that of the target-protein-fused LBT. Once inside the cell, terbium ions should diffuse to and bind the high-affinity LBT attached to the protein molecules, both soluble and packed in the crystal.

pFB1 vectors with N- and C-terminal lanthanide binding tag (LBT-N and LBT-C, respectively) targeting the cytoplasm or the endoplasmatic reticulum (ER), with the retention signal 'KDEL' (pFB1 ER-KDEL LBT-N/C) or without it (pFB1 ER LBT-N/C) were generated in a co-effort with Robert Schönherr. The attached LBT sequence 'DTNNDGWIEGDEL' that binds terbium ions with a K_D in the nanomolar range was determined by the research group of Christian Löw (EMBL Hamburg)[270].

Dense elongated structures observed in rBV IMPDH N-LBT infected High Five cells.

High Five cells were observed daily by light microscopy after infection with P3 rBV IMPDH N-LBT and P3 rBV IMPDH C-LBT for eight days. First needle-shaped dense structures were detected four days *p.i.* with rBV IMPDH N-LBT. In **Figure 20**, High Five cells five to eight days after infection are shown. Some of the elongated structures occuring in 17 % of cells remained intact even when free-floating in ESF921 medium with the supplemented antibiotics penicillin and streptomycin (denoted +P/S, 100 U/mL P, 100 µg/mL S) after cell lysis (**Figure 20 C+ D**). High Five cells expressing IMPDH C-LBT showed first indications for potentially crystalline structures only at day six after rBV infecton, with a crystallization efficiency below 0.01 %.[271] Due to the rare occurrence of ordered structures in rBV IMPDH C-LBT infected High Five cells, further experiments were restricted to cells producing IMPDH N-LBT.
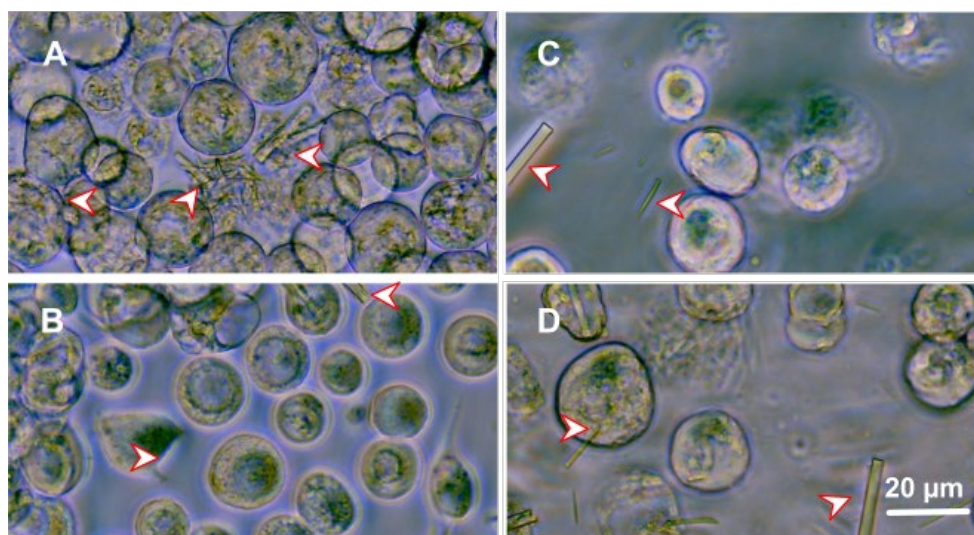
Figure 20 Morphology of thick needle-shaped IMPDH N-LBT *in cellulo* structures.
High Five cells were imaged for eight subsequent days after infection with 20 μL P3 rBV, using a DM ID LED with 20 x objective and an Axio Cam MRm camera in 6-well plates with $0.8 \cdot 10^6$ cells per well. **A** Five days *p.i.* with P3 rBV IMPDH N-LBT **B** Six days *p.i.* **C** Seven days *p.i.* **D** Eight days *p.i.*.The arrows mark potentially crystalline structures both *in cellulo* and free-floating after release from lysed cells.

A *Tb*IMPDH construct with N-terminal His-tag, followed by a TEV-cleavage site and the native C-terminal peroxisomal target sequence 'SKL' (denoted IMPDH N-His) was mentioned already in this work for *in cellulo* crystallization and subsequent structure solution applying the SSX approach (Nass et al (2020), PDB *6rfu*). IMPDH N-His forms needle-shaped *in cellulo* crystals with a maximal length of 70 μm,[134] an average length of 13 μm and a thickness of usually 2 to 4 μm. The N-terminal exchange of the His-tag and the TEV-cleavage site against the LBT does not significantly affect the morphology: IMPDH N-LBT forms comparable elongated structures, which are, however, on average 3.5 μm shorter and thicker. Since the plasma membrane surrounding intracellular crystals could hamper the direct accessibility for soaking experiments with terbium ions, isolation of potential crystals may be necessary for heavy atom staining. Therefore, the stability of isolated IMPDH-structures in different buffers was tested. The impact of the N-terminal LBT on the stability of the regular structures (marked in **Figure 20**) was investigated in direct comparison to likewise-treated IMPDH N-His crystals.

LBT reduces stability of dense IMPDH N-His structures.

High Five cells rBV infected with IMPDH N-His were subjected to different environmental conditions, including detergent, hypotonic buffer, denaturing agent, and a pH ranging from 3.5 to 9, all supplemented by proteinase inhibitor cocktail (PIC), and monitored by light microscopy over 24 h at 26 °C. IMPDH N-His crystals are remarkably robust over the entire time period at most conditions tested, including 2 M urea and RIPA buffer (Methods, **Table 11**). However, after one-hour incubation in 25 mM ammonium bicarbonate (pH 9), only a few and damaged crystals remained. Signs of deformation have also been observed in hypotonic buffer with 50 % glycerol and in an EDTA-containing buffer (100 mM Na₃PO₄,

1.5 M NaCl, 10 mM EDTA, pH 8). Under denaturing conditions with 4 to 6 M urea in hypotonic buffer, all crystals dissolved in less than five minutes (**Figure 21**).
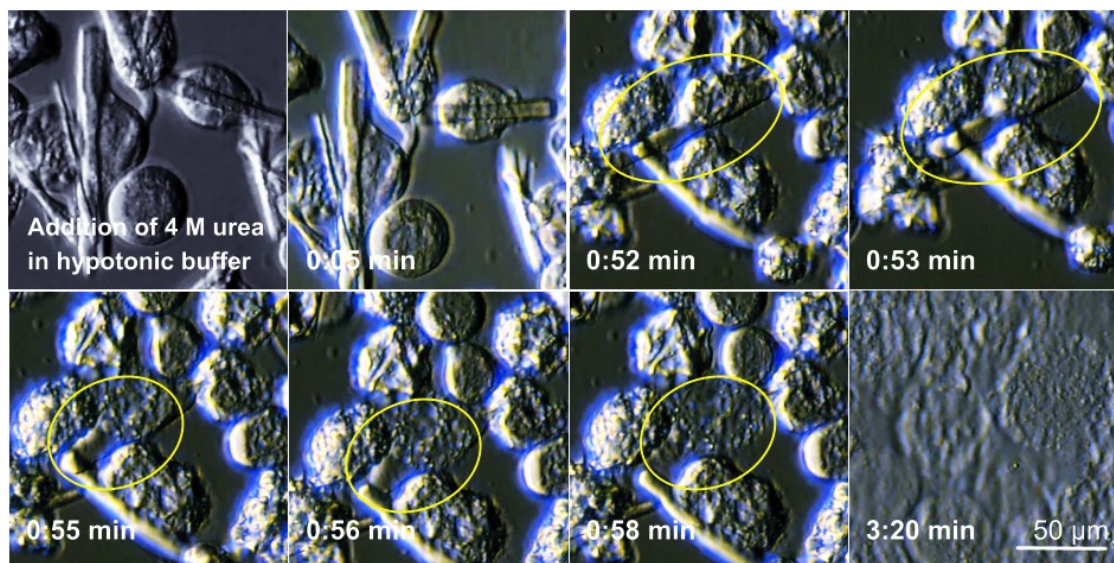
Figure 21 Live-cell imaging of the osmotic shock and subsequent resolubilisation of IMPDH N-His crystals 5 days *p.i.* in High Five cells after addition of hypotonic buffer containing 4 M urea.
Cells shrank by dehydration within 5 sec, crystal morphology was significantly affected within 1 min. During resolubilization of IMPDH N-His, needle-shaped crystals contracted, broadened, to the center of the cell, as exemplarily highlighted by yellow circles. The crystal completely dissolved within approximately 3 minutes. Cells were imaged at a Leica DM IL LED microscope.

The stability of the crystal-like structures detected in rBV IMPDH-LBT infected High Five cells was likewise tested. At five days *p.i.*, medium of rBV IMPDH N-LBT infected High Five cells (MOI 1) was removed and substituted with destilled water, hypotonic buffer, PBS/ TBS supplemented with 0.5 % TritonX100, RIPA buffer, or ESF921 medium as a control. All solutions were supplemented with PIC to inactivate cellular proteases freed by cell lysis. Light microscopy monitoring over 24 hours revealed that the quantity of potentially crystalline structures was reduced at hypotonic conditions and in the presence of detergent, likely due to dissolving of crystals after cell lysis. Lysis of cells containing *in cellulo* IMPDH N-LBT crystals with RIPA buffer for more than 2 h resulted in morphological changes of the structures and complete dissolution of cells, whereas the macroscopic structure of IMPDH N-LBT assemblies remained intact if the cells were incubated in water supplemented with 10 mM magnesium chloride, or in ESF921 medium for at least one day.

Concluding from these experiments, isolation of the IMPDH N-LBT structures from the living cells should be avoided if possible and kept in low-salt water if not. Long-term storage of isolates is not recommended for diffraction data collection.

Observed intracellular dense IMPDH N-LBT structures diffract X-rays.

At day four *p.i.*, $1 \cdot 10^6$ High Five cells were prepared for SSX as described in **Figure 12**. Serial data collection with helical line scans was performed at cryogenic conditions at a wavelength

of 0.97 Å (12.7 keV, not the Tb-edge) in mesh grids. Diffraction tests of IMPDH N-LBT structures within the intact cells at EMBL beamline P14 resulted in detector images that contained detectable Bragg diffraction. The majority of crystals diffracted up to approximately 3.5 Å resolution, while individual detector frames revealed diffraction up to a maximum resolution of 3.0 Å resolution (**Figure 22**). In comparison, the structures of IMPDH ori (pdb 8C53) and IMPDH cyto (pdb 8C51) could be solved by *in cellulo* SSX at a resolution of 2.3 Å and 2.4 Å, respectively.
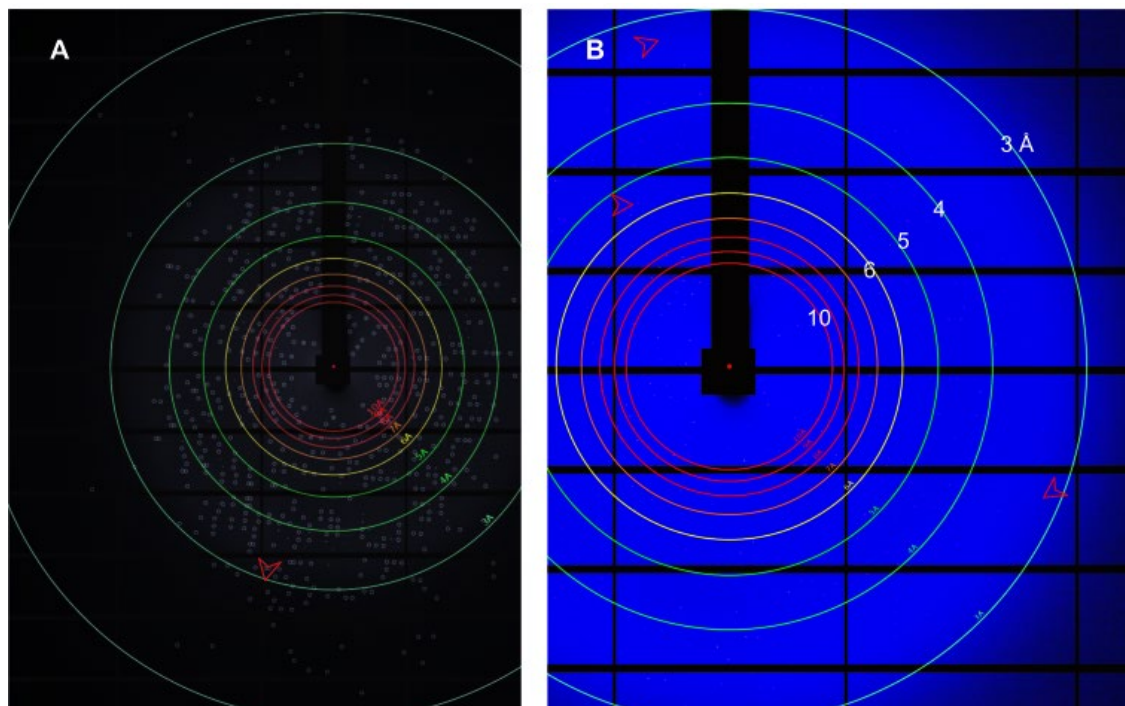
Figure 22 Diffraction image section of irradiated High Five cells 4 days *p.i.* with rBV IMPDH-LBT-N, collected at 100 K in Mitegen MicroMeshes.
Diffraction parameters were a beam focus of 10 x 5 μm², an exposure time of 150 ms, an EIGER16(4M) detector distance of 615 mm and an oscillation range of 0.5° with no overlap. The beamstop and panels were masked in the geometry file used for indexamajig of the CrystFEL suite. **A** Bragg peaks were successfully identified (check_near_Bragg) up to 3 Å and are encircled. The detector frame is depicted in inverted greyscale with resolution rings in colour. **B** Color representation diffraction picture. More than 100 peaks in light blue are present in the picture, but difficult to see. More intense peaks in the low-resolution are shown in pink. Exemplary Bragg spots are marked by red arrows. Resolution rings from 10 to 3 Å are shown.

After this confirmation of the crystalline character of the IMPDH N-LBT assemblies and identifying reasonable short-term storage conditions for free-floating crystals in 10 mM magnesium chloride water or ESF921 medium, intracellular crystals were isolated for soaking with Tb^{3+} at the day of the diffraction experiment. Since the saturation concentration of terbium chloride in ESF921 medium is below 15 μM, medium was exchanged to TBS buffer prior to soaking experiments with 50 mM terbium chloride in TBS buffer. Diffraction data were collected at PETRA III, EMBL beamline P14 as described before applying a helical line scan approach with a 5 x 10 μm focus beam and using a photon flux of 10^{13} ph/s, but at an energy of 7.514 keV (1.64 Å), corresponding to the L_{III} X-ray absorption edge of terbium.

Isolated, but unsoaked IMPDH N-LBT crystals diffracted up to 3.4 Å visible resolution. Bragg spots identified on the recorded diffraction patterns in *adxv* after the soaking experiment reached Bragg-spacings up to 6 Å.

The presence of terbium ions in the soaked, twice washed sample was confirmed by an X-ray fluorescence spectrum of isolated crystals in TBS with 1 s exposure at 7.5598 keV. The spectrum (**Figure 23 A**) exhibits a peak at approximately 7.5 keV, characteristic for the terbium Lγ emission line (7.5140 keV). The results of an energy scan for terbium with an inflection energy at 7.5199 keV and a peak energy at 7.5395 keV is shown in **Figure 23 B**.
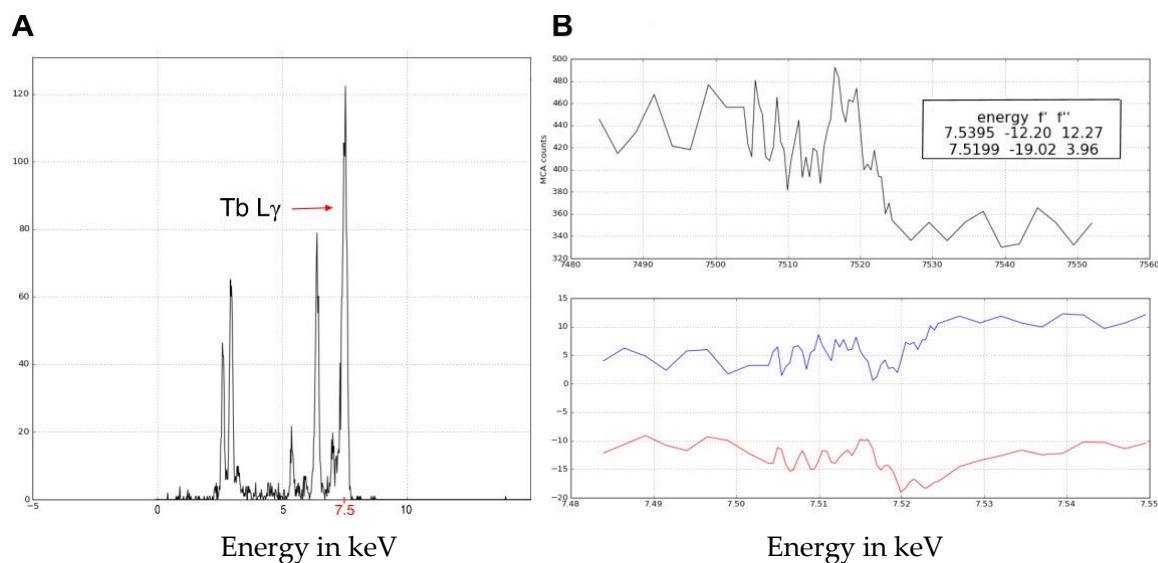
Figure 23: **Terbium can be detected in isolated and soaked crystals by X-ray fluorescence.**
Isolated IMPDH N-LBT crystals were soaked one hour in 50 mM terbium chloride dissolved in TBS buffer. **A** Shown is the X-ray fluorescence spectrum excited at 7.5598 keV. The arrow marks an emission line matching the terbium Lγ at approximately 7.5 keV; **B** Result of the energy scan for terbium.

Dozor found few hits, all appearing at low resolution, only 12,233 images were collected at the terbium edge and after the soaking with 50 mM terbium chloride, resolution was diminishing during data collection. This (incomplete) dataset was processed with *CrystFEL*. *Indexamajig* settings were *peakfinder8* with an integration radius of 3.8, 5.6 and 6.7, a tetragonal lattice type with the unique axis c, a minimal SNR of 5, a threshold at 100 and the *–highres* option at 2, including multi-hits in the 'delete-found-pattern, retry' algorithm (*--multi*). 775 crystals were successfully indexed with only a single frame containing two lattices. *Cell_explorer* found the unit cell axis a = b = 209.5±2.77 Å, with the unique axis c = 93.81±0.93 Å, and all angles at 90 ±0.5°, which was extracted from 2.024.106 measurements in total with 152.535 total reflections.

Indexed patterns were merged at the lower symmetry point group (-y 422 instead of 4/mmm) to account for unequal Friedel pairs, and scaled without modelling partialities (*unity* model) in two *partialator* iterations. Figures of merit of this small merged dataset were: 100 % completeness, CC* 8136 (0.1039), CC 0.4948 (0.0054), R_{split} 158.35 %, SNR 0.93 (0.13). Figures of merit for the highest resolution shell at 4.39 Å are given in brackets.

As to be expected for such a small number of indexed patterns at medium to low resolution, these statistics, including the overall $CC_{ano}= 0.0775$ (0.0037) was insufficient for SAD and otherwise below the self-set quality parameters (listed at page 52) suitable for structure solution. With the apparent damage to IMPDH LBT-N crystals resulting in low resolution data after terbium treatment and an overall low hitrate, the approach to soak isolated IMPDH N-LBT crystals was not continued. Instead, the possibility to transfer Tb^{3+} into the cytosol of the viable cell by diffusion and direkt binding of the LBT of the target protein was evaluated.

Lanthanide transfer into living insect cells

Förster or fluorescence resonance energy transfer (FRET) of a tyrosine or tryptophane residue in the binding tag to Tb^{3+} can be triggered by excitation at 280 nm.[270] This indirect excitation of Tb^{3+} via FRET was used to test whether Tb^{3+} binds to the LBT[272].

Tb^{3+} exhibits three emission maxima at 490 nm, 545 nm and 590 nm, respectively. To measure Tb^{3+} binding, fluorescence emission was detected at 544 nm using a *Jasco® FP-6200 spectrofluorometer*, as described in the methods section. Tb^{3+} is coordinated by six oxygen atoms when bound to the LBT.[273] Soaking experiments in TBS with 5 mM $TbCl_3$ were performed with High Five cells four days *p.i* with rBV IMPDH N-LBT and rBV IMPDH with an N-terminal hemagglutinin tag (IMPDH HA-N), as well as with uninfected High Five cells and isolated IMPDH HA-N crystals. The objective, implemented as a bachelor thesis in the working group, was to study possible diffusion of terbium ions through the cellular membranes and binding to the LBT.

Isolated IMPDH HA-N crystals showed the Tb^{3+} specific fluorescence emission at 544 nm. Cells in the absence of terbium ions did not produce a fluorescence signal upon excitation[272]. However, High Five cells expressing IMPDH N-LBT and also IMPDH HA-N (no LBT tag) both showed the characteristical fluorescence at 544 nm when excited at 280 nm, indicating additional unspecific binding of Tb^{3+}, possibly to outer membrane proteins[272]. Time monitoring revealed that the FRET signal is highest directly after terbium admission and diminished in a non-linear fashion over time, therefore subsequent measurements were performed just after mixing.

To verify free diffusion into the cells, Tb^{3+} binding to intracellular DNA was examined.[274,275] For this purpose, the uninfected cells were resuspended in TBS with 5 mM $TbCl_3$ and excited at 260 nm, thus exciting intracellular DNA, prompting FRET for bound Tb^{3+}. A signal was detectable at 544 nm, confirming binding of Tb^{3+} to DNA.[272] In addition, because nucleic acids are only present intracellularly, this suggests free diffusion of the lanthanide ions across the membrane into the cells.

To sum up results of the first approach of heavy atom staining: pFB1 vectors with N- and C-terminal LBT were created and the Tb*IMPDH* sequence extracted from a cytoplasmic

IMPDH pFB1 vector was succesfully cloned into these vectors. Infection with harvested P3 rBV IMPDH N-LBT resulted in growth of thick needle-like crystals, that showed less stability than IMPDH N-His crystals at the tested conditions, e.g. in RIPA buffer. Bragg diffraction confirmed the crystalline identity of these structures, but low and diminishing diffraction quality prevented the collection of a complete dataset. Unfortunately, SAD phasing of diffraction data obtained from Tb^{3+}-loaded crystals in the intact cell was not possible, due to unspecific binding, and both because of too few indexable patterns and too little anomalous signal found in the diffraction data.

3.4.2 Second heavy atom staining approach: Selenium incorporation.

The second concept for staining with heavy atoms *in cellulo* towards *ab initio* phasing involves the substitution of sulfur in methionine (Met) with selenium. To incorporate selenium into the recombinant proteins in terms of selenomethionine (SeMet),[276] cells were grown in a Met deficient medium with SeMet supplementation for the specific sequence-dictated integration of selenium into the amino acid chain[277]. *In cellulo* crystals and the extent of SeMet incorporation were characterized for multiple protein targets (chapter 3.4.3 to 3.4.5). Subsequently, first anomalous diffraction data were collected of HEX-1-SeMet crystals (chapter 3.4.6 - 3.4.7).
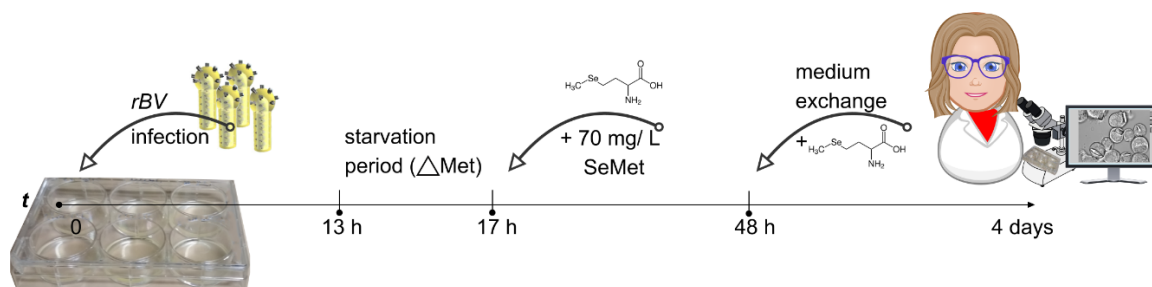
Figure 24 SeMet incorporation protocol with experiment timeline.
Adherent cell cultures (Sf9 or High Five) are infected with rBV (MOI 1) at 70 % confluency in ESF921 medium with antibiotics. At 13 hours *p.i.*, cells are washed once with ESF92-ΔMet medium following a four-hour starvation period with Met-deficient ESF921 medium as nutrition. The cell culture is supplemented with 70 mg/L SeMet in ESF921-ΔMet at 17 hours *p.i.*, and the ESF921-ΔMet medium exchanged at 48 hours *p.i.*. At the same time, SeMet supplementation is renewed. Crystal growth and cell viabilities are documented by LM four or five days *p.i.* As control, insect cells are mock-treated, leaving out any SeMet supplementation, thus only performing medium exchanges with washing.

The first step to incorporate selenium at known positions in the target protein that forms the intracellular crystals is the adaptation of the cell culture to a Met-deficient medium (ESF921ΔMet, Expression Systems). A starvation period was implemented into the SeMet incorporation protocol (**Figure 24**) for depletion of resident methionines in the cell culture, followed by supplementation of SeMet.

As mentioned already, polyhedrin was replaced by the gene of interest in the BEVS.[181] The baculovirus gene expression has a temporal regulation (see p.22); Expression of genes under control of the polyhedrin promotor is initiated very late in the infection cycle.[160,161]

However, expression of earlier viral genes like *gp64*, encoding for an envelope fusion protein, or *VP39* that is important for capsid formation, is necessary for successful viral infection and replication.[160] Thus, sufficient expression of early to late viral genes should be ensured with Met supply before the switch to ESF921ΔMet medium.

The timepoint of infection is crucial for maximization of the SeMet content of the target protein.[277] No robust data was found regarding the exact timing of the baculovirus life cycle in High Five cells.[160] SeMet addition is timed such that depletion of Met stocks in the cells is ensured just before the expression of late genes starts within the first 16 hours after virus infection; YFP fluorescence is detectable from 18 hours on after rBV infection. Ideally, SeMet supplementation should start before the insect cells switch to the alternate methyl cycle pathway and start synthesizing Met independently, possible from homocysteine metabolized from cysteine in the nutrition. Chosen timepoints of medium exchange and starvation period are depicted in **Figure 24**.

Toxicity of SeMet titration with subsequent cell vitality assay

Both High Five and Sf9 cells were tested for the maximum tolerable concentrations of SeMet in the cell culture. The cell viability was measured by PI staining (**Table 4**). PI is a red fluorescent agent that intercalates with DNA, and cannot cross the plasma membrane of living insect cells. The PI stain therefore is frequently used as a viability assay. Absolute numbers of cell vitality in the cell cultures differed from experiment to experiment. The live-cell ratio however consistently dropped for a SeMet dose exceeding ~ 90 mg/L SeMet and ~ 50 mg/L for High Five cells and Sf9 cells, respectively. These values represent the highest tolerable dose, assuming that rBV infection does not influence SeMet toxicity. If not stated otherwise, 70 mg/mL SeMet equivalent to 357 μmol/L SeMet supplementation was used for High Five cells.

Table 4 Cell vitality results of PI stain 5 days after addition of different SeMet concentrations to uninfected cell cultures.
Shown is the percentage of stained, respectively dead, High Five and Sf9 cells after manual counting from > 200 cells per condition. Zeiss Observer Z1 using the DsRed channel overlay with bright field LM.
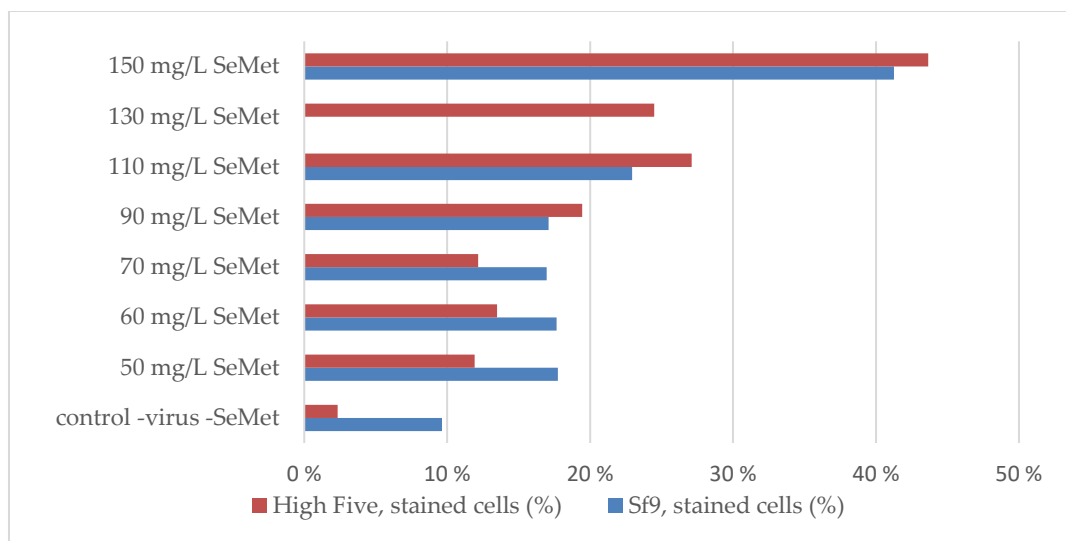
SeMet concentration [mg/L]	PI stained High Five cells	PI stained Sf9 cells
control (0)	4.0 %	5.3 %
50	5.0 %	6.8 %
70	6.8 %	97.0 %
90	19.3 %	97.3 %
110	68.0 %	100.0 %
130	84.2 %	100.0 %
140	98.0 %	100.0 %
150	100.0 %	100.0 %

Medium exchange to fresh ESF921ΔMet supplemented with virus stock at chosen MOI and fresh 70 mg/L SeMet supplementation at day 2 *p.i.* was found to improve cell viability.

In the next step, toxic effects of SeMet supplementation with respect to a possible influence of the virus infection were investigated. Cells were infected with P3 virus stock and allowed to express the early to late viral genes in ESF921 medium with antibiotics for up to 13 h, followed by a starvation period (4 h) to use up remaining Met before SeMet supplementation (**Figure 24**). At MOI 1, statistically 37 % of cells remain uninfected in the primary infection round, as described by the Poisson distribution:

$$P(0) = e^{-MOI} = e^{-1} = 0.37.$$

Results of a SeMet titration of Sf9 and High Five cells infected with a mock virus at MOI 1 are presented in **Figure 25**.

Figure 25 Impact of mock-virus infection using MOI 1 on the SeMet toxicity of high Five and Sf9 cells. Results of vitality assay of High Five cells (red bar) and Sf9 cells (blue bar) via PI stain 1:2000, 4 days *p.i.* with MOI of 1 mock virus. SeMet concentrations tested between 50 mg/L and 150 mg/L L-SeMet in Met-deficient ESF921 insect cell medium. No data exist for Sf9 cells at 130 mg/L. The proportion of stained cells was manually counted from overlays of DsRed fluorescence and transmitted light imaged using a Zeiss Observer Z1 microscope at 20 x obj, counting at least 300 cells per condition from three random areas in the infected 12-well culture plate.

In the absence of virus infection, more than 65 % of cells of both insect cell lines would be dead after four days at 110 mg/L SeMet supplement, as previously shown. Virus infection with MOI 1 reduced the cell mortality to less than half, indicating an unexpected increased resistance against toxic effects of SeMet supplementation at higher SeMet concentrations. Even though the absolute percentage of live cells at 70 mg/L is lower (now 30 %) at infected cells (MOI 1), this effect is partly due to a less healthy cell culture with 10 % dead cells in the control cells without SeMet supplementation.

My observation was recently supported by Wenzel *et al.*[278] The authors used a Mediator Head module of yeast as a single model protein complex composed of several subunits involved in transcription with RNA polymerase II. Wenzel *et al.* observed an 'estimated effectivity of infection' (eMOI) 24 h after infection, as a so-called titer estimation of quality control (TEQC) method, and observed that a high virus titer protected High Five cells from toxic effects of SeMet supplementation.[279]

Inspired by this result, intracellular protein crystallization with SeMet incorporation and a high-titer virus infection was tested. The TCID$_{50}$ titer estimation calculated with respect to MOI was used, as described in the InCellCryst publication, where MOI 1 has been evaluated best for maximisation of crystal growth.[181] The setup developed in this work compares a MOI 1 with that of a very high titer, MOI 10, which still shows a majority of live cells.[280] Two SeMet concentrations were investigated: 70 mg/ L SeMet, a tolerable dose for more than 90 % of uninfected Sf9 and High Five cells, and 150 mg/L SeMet supplementation.

High Five cells were infected with rBVs encoding HEX-1 cyto, IMPDH or CatB SKLG using a MOI of 1 and a MOI of 10, otherwise following the SeMet incorporation protocol described above. The CatB variant with C-terminal SKLG addition was chosen due to the increased crystal volumes compared to the native CatB *in cellulo* crystals (previous observations in the AG Redecke). As a control, rBV mockvirus-infected cells were treated likewise. A vitality assay was performed 4 days *p.i..* Results are shown in **Figure 26**. Panels **A**, **C, E** and **G** depict overlays of the bright field (BF) and DsRed fluorescence channels. At least 400 cells were counted per condition from four images at random positions of two independently infected wells to determine the total cell numbers at BF (20x obj) and the number of stained cells (10x obj., charts in **Figure 26** pannels **B**, **D**, **F** and **H**).

Independent of the MOI, crystal growth of HEX-1 cyto, IMPDH N-His and CatB SKLG was observed for all conditions 4 days *p.i.* with 70 mg/L SeMet supplementation, as marked by yellow arrows in **Figure 26** (**A**, **C, E**). Visual inspection indicates that an MOI of 10 increased the percentage of cells with intracellular target protein crystals both at 70 mg/L SeMet and 150 mg/L SeMet supplementation. Less cells infected with MOI 1 survived the high-dose SeMet supplementation. The tolerable SeMet dose that allows *in cellulo* crystal growth is also target dependant: For example, no IMPDH N-His crystals grew at MOI 1 in the presence of 150 mg/L SeMet. At MOI 10, IMPDH N-His crystals did grow, but the crystals grown with 150 mg/L SeMet supplementation appeared deformed.

Occasionally, at high SeMet dosage, several virus-infected cells build a syncytium that is covered by an enlarged outer membrane and permeable for PI (**Figure 26 C** and **E**). The syncytium formation does not prevent intracellular crystal growth in these uncharistically large cells, as observed in **Figure 26 C** (HEX-1 cyto, upper right panel) and **E** (*Tb*IMPDH, lower right panel).

There is a trend for an increased cell viability at high MOI compared to MOI 1 with one exception (again Tb*IMPDH, 70 mg/ L SeMet, MOI 1 vs MOI 10*). On average, 86 % of High Five cells survive the treatment (*MOI 1 + 70 mg/ L SeMet*), while more cells survive with the higher MOI (89 %, *MOI 10 + 70 mg/ L SeMet*). 82 % of High Five cells survive the high SeMet concentration supplementation (*MOI 1 + 150 mg/ L SeMet*) and even 89 % survive at *MOI 10 + 150 mg/ L SeMet*. Consequently, future experiments on SeMet incorporation into protein crystals in High Five cells should use a high MOI that enables a SeMet supplementation above 70 mg/L. Unfortunately, no further beamtime was available since obtaining this result.

Therefore, the characterization of SeMet incorporated crystals below focuses on results from MOI 1 infections of High Five cells.

CatB SKLG

HEX-1
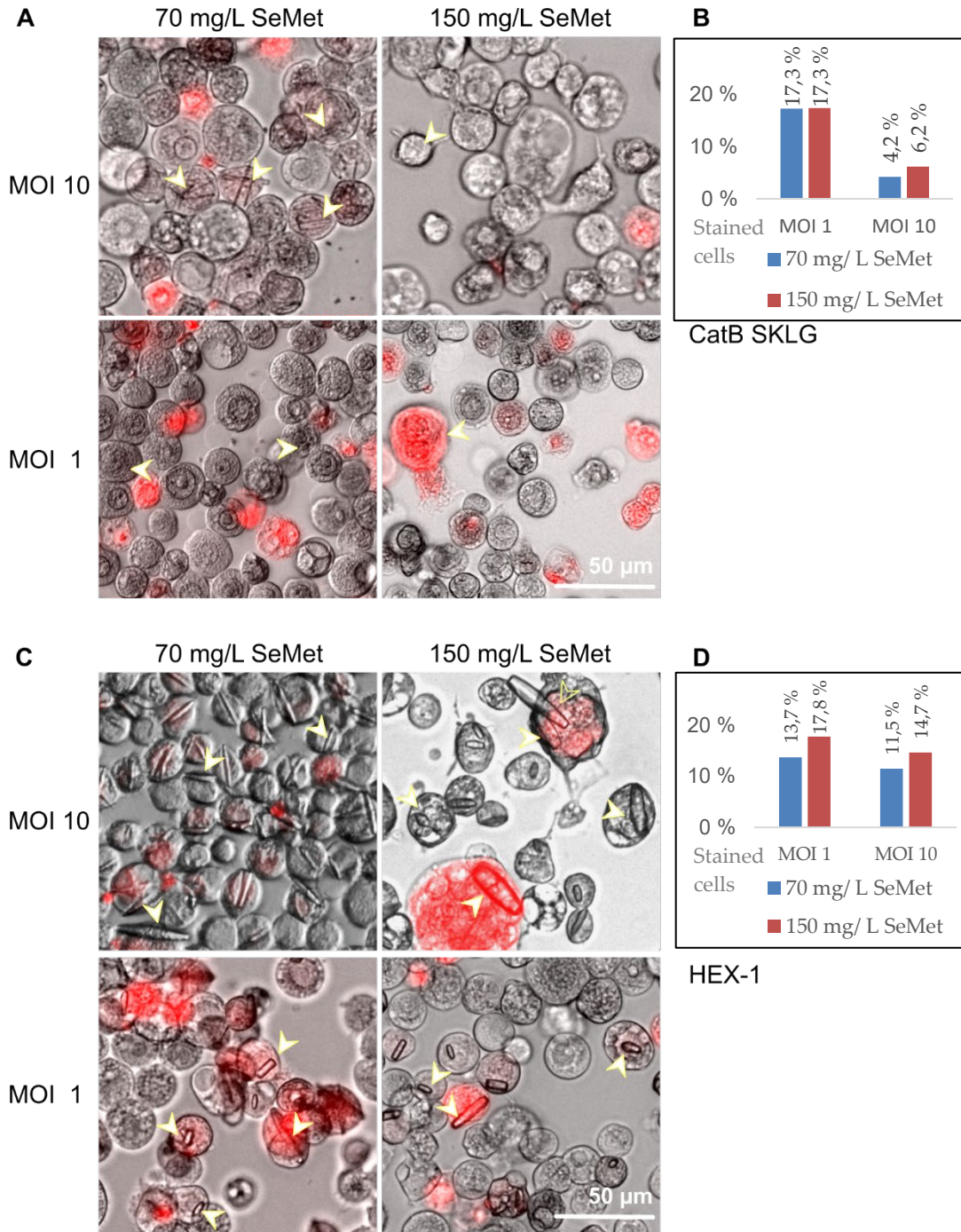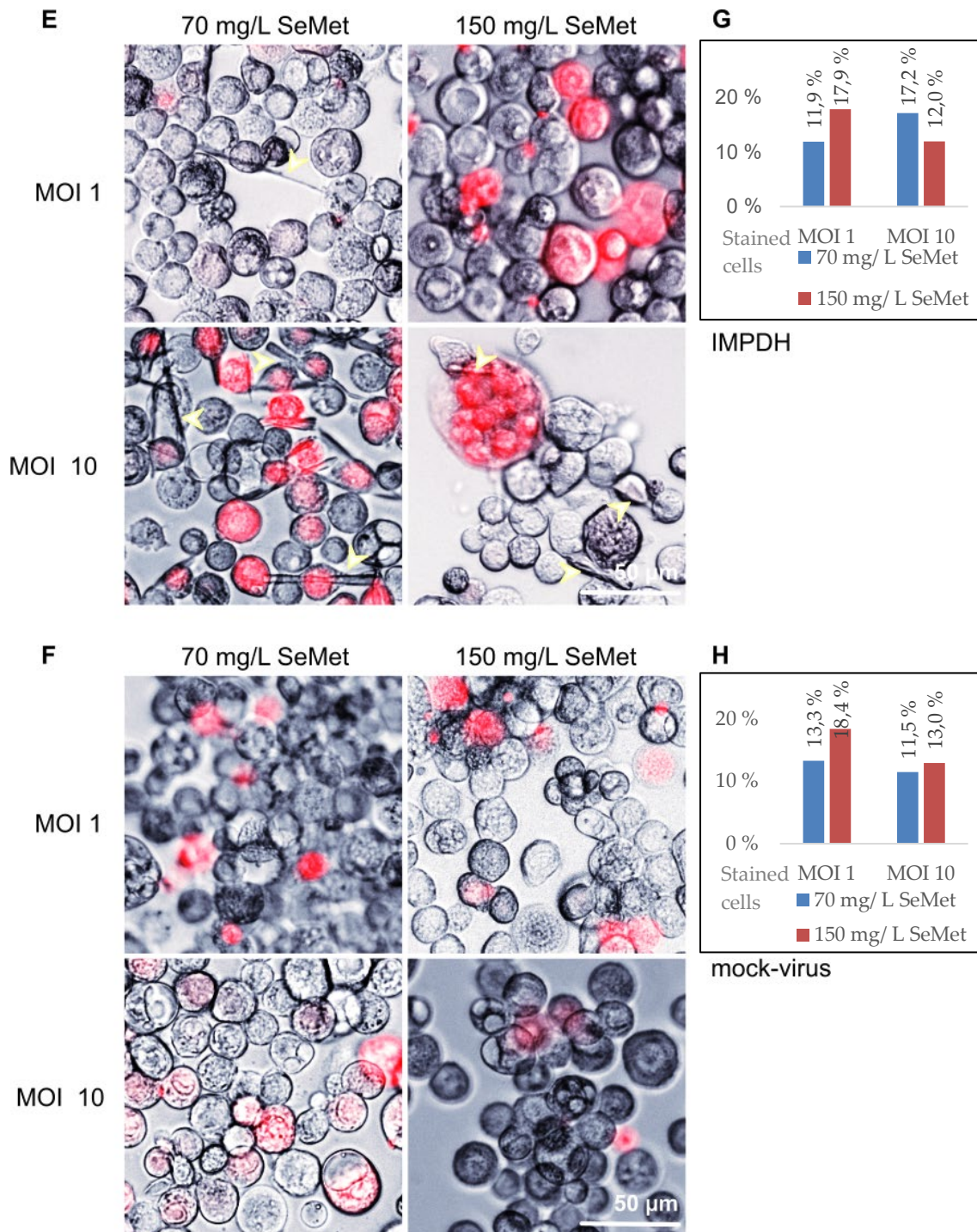
Figure 26 The influence of virus infection on SeMet cytotoxicity, using MOI 1 and MOI 10.
PI cell vitality assay. High Five cells were imaged 4 days *p.i.* using a Zeiss Observer Z1. Microscopy images shown are overlays of bright field - and DsRed channels. Fluorescence of PI diffused into damaged cells is depicted in red. Cell mortality in the bar charts is given in % for over 500 manually counted cells per condition. High Five cells infected with rBV at high MOI are presented in upper panels and cells infected with rBV at low MOI in lower panels. Cells supplementated with 70 mg/L SeMet are shown in left panels, those supplemented with 150 mg/L SeMet in right panels. **A** High Five cells infected with P3 rBV CatB. **B** Bar chart displaying the percentage of PI-positive cells infected with rBV CatB. **C+D** High Five rBV HEX–1 cyto infected, likewise treated. Circa 85 % of infected cells remain viable with a MOI 10 and 150 mg/L SeMet supplementation. **E+F** High Five cells infected with P3 rBV IMPDH likewise treated. **G+H** rBV mock–virus infected High Five cells including bar chart of cell mortality rates.

3.4.3 Characterization of SeMet incorporated crystals

For successful mapping of heavy atoms into an electron density map, not only the position and incorporation percentage within all molecules forming the crystal matters, but also the number of heavy atoms, i.e. SeMet positions that are available in the individual molecule. Next to the protein CatB SKLG (7 Met, 2 in propeptide within the target protein sequence), also HEX-1 cyto (6 Met) and IMPDH N-His (16 Met) crystallized with incorporated SeMet intended for SAD phasing.

High Five cells treated with the developed SeMet incorporation protocol (**Figure 24**) and infected with recombinant rBV CatB SKLG (**Figure 27**), HEX-1 cyto (**Figure 28**) or IMPDH N-His (**Figure 29**) were imaged using DIC. As before, the efficiency of crystal formation, representing the percentage of High Five cells that grow at least one crystal visible by LM was determined from a sample size of at least 400 cells per condition. Crystal sizes were measured using the software ImageJ, and crystal morphologies compared to the apparent crystal morphologies within cells grown and infected in Met-containing medium. The needle-like crystal morphology of CatB in Sf9 cells has been characterized elsewhere.[41,148]

CatB SKLG is a CatB mutant with the four additional residues serine, lysine, leucine and glycine (SKLG) at the C-terminus, plus an alanine remaining from the blunt-ending cloning strategy. This protein grows roubust needle-like structures of up to one µm in thickness and 40 µm in length, which are thus longer than the wild-type CatB crystals. In High Five cells, more than 50 % of cells exhibited crystal growth, often with five or more crystals randomly oriented in a single cell (**Figure 27 A**). The average rBV infected High Five cell displays a diameter of 30 µm, the diameter ranging between 16 to 45 µm, as opposed to an average diameter of ~20 µm of uninfected High Five cells. The average cell size does not change with SeMet supplementation.

Grown in Met-deficient medium with SeMet supplementation, rBV CatB SKLG infected High Five cells exhibit an increased formation of pseudopodiae. Characteristic evidence for cellular stress is an increase in vacuoles volume and quantity (**Figure 27 B**). Furthermore, with SeMet in ESF921∆Met, some cells fuse at the plasma membrane building syncytia. This phenomenon occurs even more frequently at SeMet concentrations above 100 mg/L (data not shown). However, the needle-like crystal morphology is not influenced by SeMet incorporation.
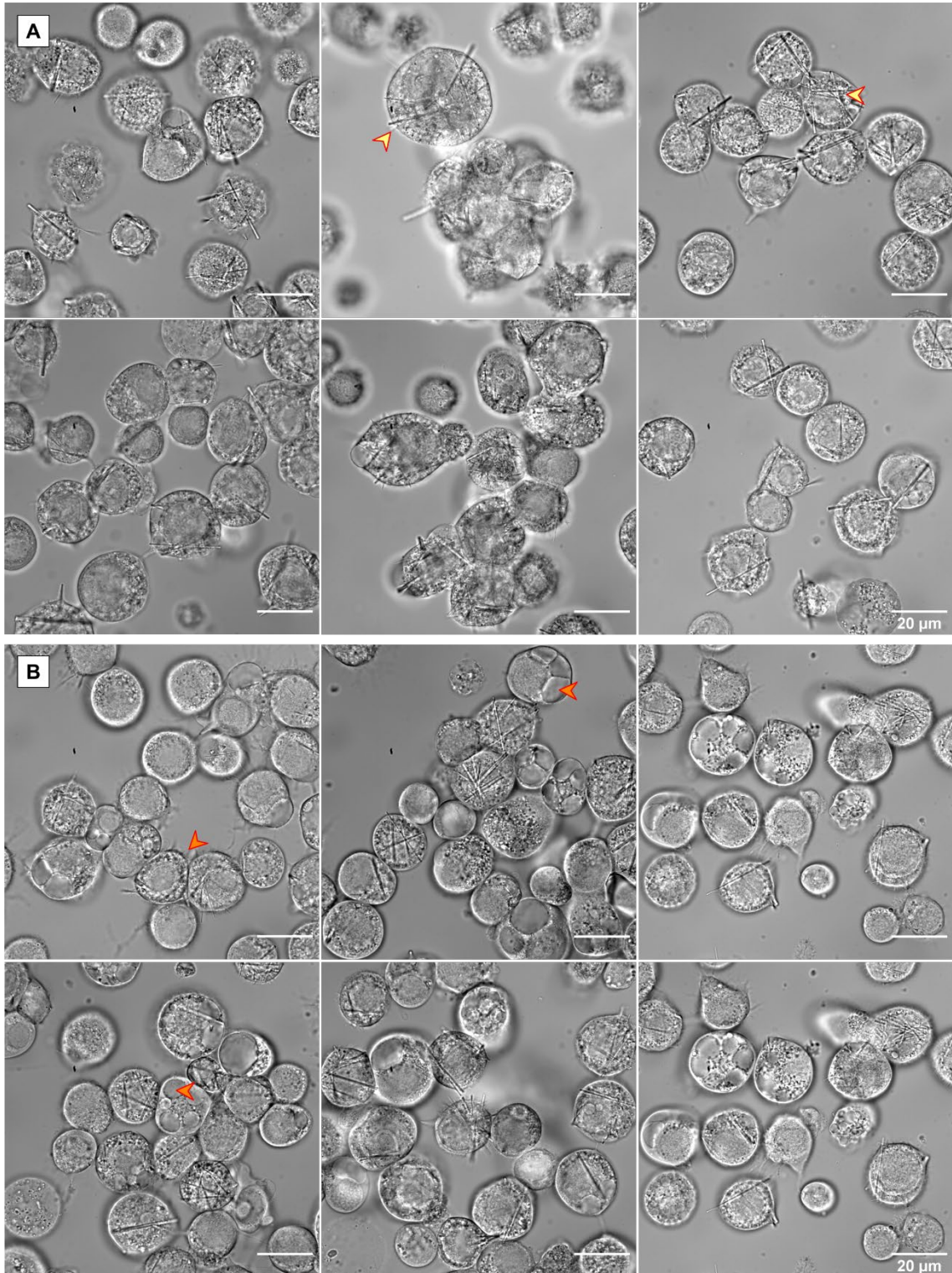
Figure 27 Microscopic comparison of CatB SKLG protein crystals. Grown either **A** in ESF921 medium with Met or **B** applying the developed SeMet incorporation protocol with 70 mg/L SeMet supplemented. Cells were grown in 6-well plates on a glass-slide. Images were taken on day 4 *p.i.* with MOI 1, applying DIC using a Nicon eclipse Ti2-E (100x objective). Increased cell diameter due to virus infection and multi needle-like crystal growth in single cells are highlighted with yellow arrows. Characteristics specific for SeMet addition are highlighted with orange-filled arrows, such as the formation of large vacuoles, filling up to ¾ of cell volume, and development of pseudopodiae. CatB SKLG SeMet crystals exhibit the same needle-like morphology as those grown with a Met source.
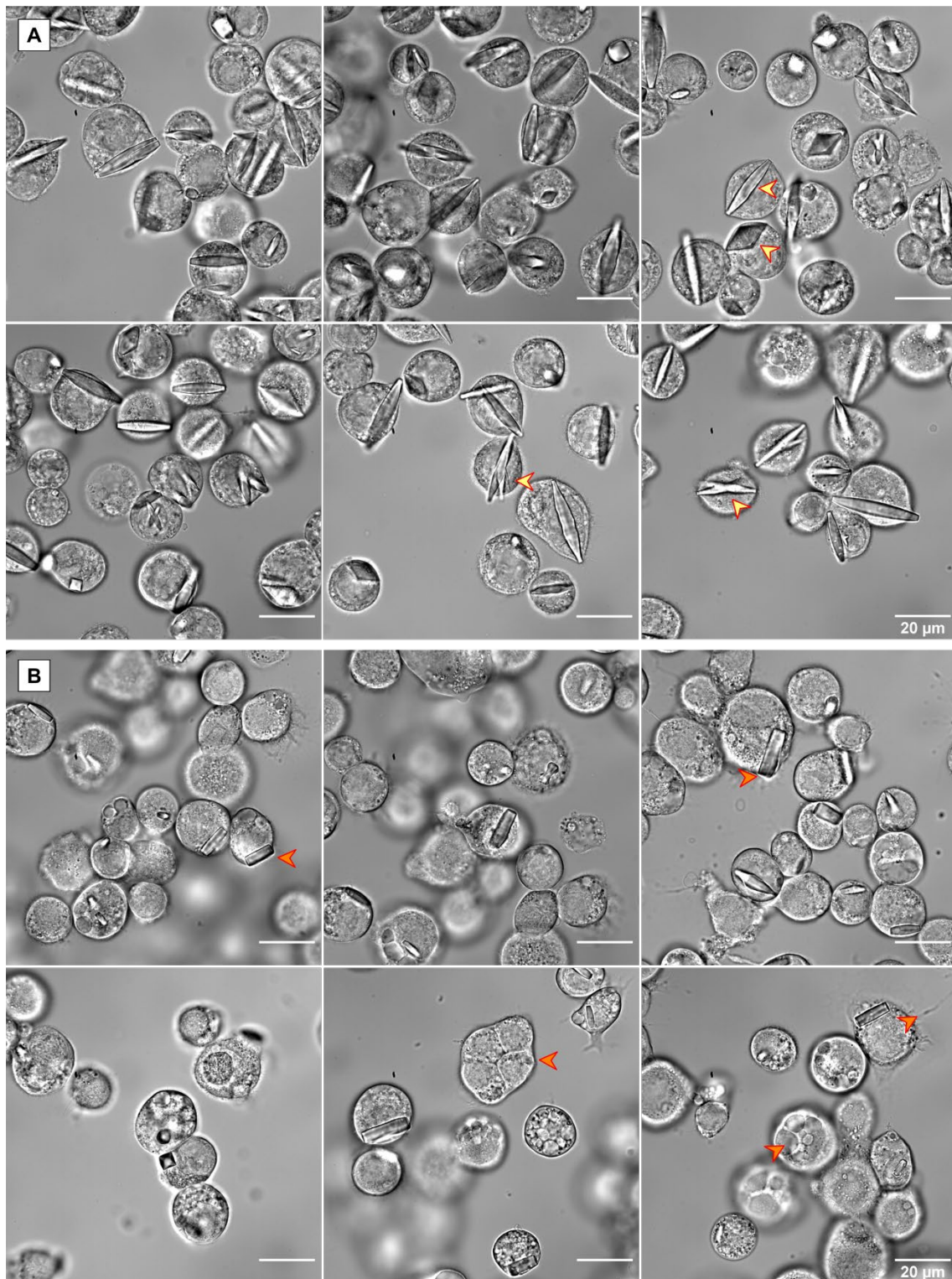
Figure 28 HEX-1 cyto crystals in rBV (MOI 1) infected High Five cells, imaged by DIC.
A At normal growth conditions with Met containing insect cell medium and **B** after growth in Met-deficient medium with 70 mg/L SeMet supplement (protocol scetched in **Figure 24**). Cells were grown in 6-well plates on a glass-slide. Cells were imaged with DIC using a Nicon eclipse Ti2-E with a 100x objective on day 4 *p.i.*. The proportion of crystal-cotaining cells in **B** is reduced, and growth of multiple crystals per cell was not observed. Cell membanes may fuse (lower middle panel).

Smooth, needle-like shapes are clearly visible that can exceed the diameter of the cell, and a maximum of one CatB SKLG SeMet crystal was observed per cell. However, crystal growth occurs 2.5 x more frequently under normal growth conditions using the same virus stock and MOI for infection. Using the SeMet incorporation protocol at MOI 1, only 20.2 % of cells contain crystals on day 4 *p.i.*.

High Five cells were rBV HEX-1 cyto infected and imaged at similar conditions to rBV CatB SKLG infected cells. In ESF921 medium, two basic crystal morphologies occur within one infected cell culture (**Figure 28 A**). Approximately twenty percent of crystals display a bipyramidal morphology, ranging between 8.5 x 8.5 x 18 μm^3 to 3 x 3 x 6 μm^3 in size. The majority of HEX-1 cyto crystals form an elongated needle-like or spindle-shaped morphology of about 3.8 x 3.8 x 24 μm^3, with some crystals growing up to 40 μm in length and 8 μm in thickness. These spindle-like crystals grow thinner towards the edges, resulting in pointy endings. Frequently, a crystal breaks in the middle, leading to the growth of more pointy endings, resulting in starfish-like structures, with all extensions fused in the centre. In cells cultured in Met-deficient ESF921 medium supplemented with 70 mg/L SeMet, 46 % of rBV infected cells grow a HEX-1 cyto crystal, seldom more than one *(***Figure 28 B***)*. The bipyramidal crystals exhibit a smaller characteristic size and volume. As observed for CatB SKLG crystals containing SeMet, on day 4 *p.i.*, stress-symptoms with pseudopodiae- and large vacuole formation are observed more frequent. Furthermore, the spindle-shaped morphology, which is more abundant under normal growth conditions and diffracts to high resolution, changes: Starfish or cross-like structures were not observed at 70 mg/L SeMet. Crystals grow shorter and more compact, because the edges do not in most cases thin down, leading to a more building block-shape appearance.

Under normal growth conditions, IMPDH N-His crystals exhibit a thick needle-like morphology with a maximum length of ~ 70 μm and a thickness of between 2 to 4 μm on average which sometimes extends up to 10 μm in width (**Figure 29 A**). IMPDH N-His crystals usually end in a smooth edge, which can break, then fanning out to smaller needles. The crystals grown with the SeMet incorporation protocol are less stable (dissolve in RIPA buffer) and less than 5 % of cells grow a crystal (**Figure 29 B**). These needle-shaped crystals are often smaller than their equivalents grown in the presence of Met, exhibit thinned-out endings, presumably as a sign of degradation. In addition to pseuropodiae and large cellular vacuoles, cases of lysis of the plasma membrane leading to cytosolic leakage occur.
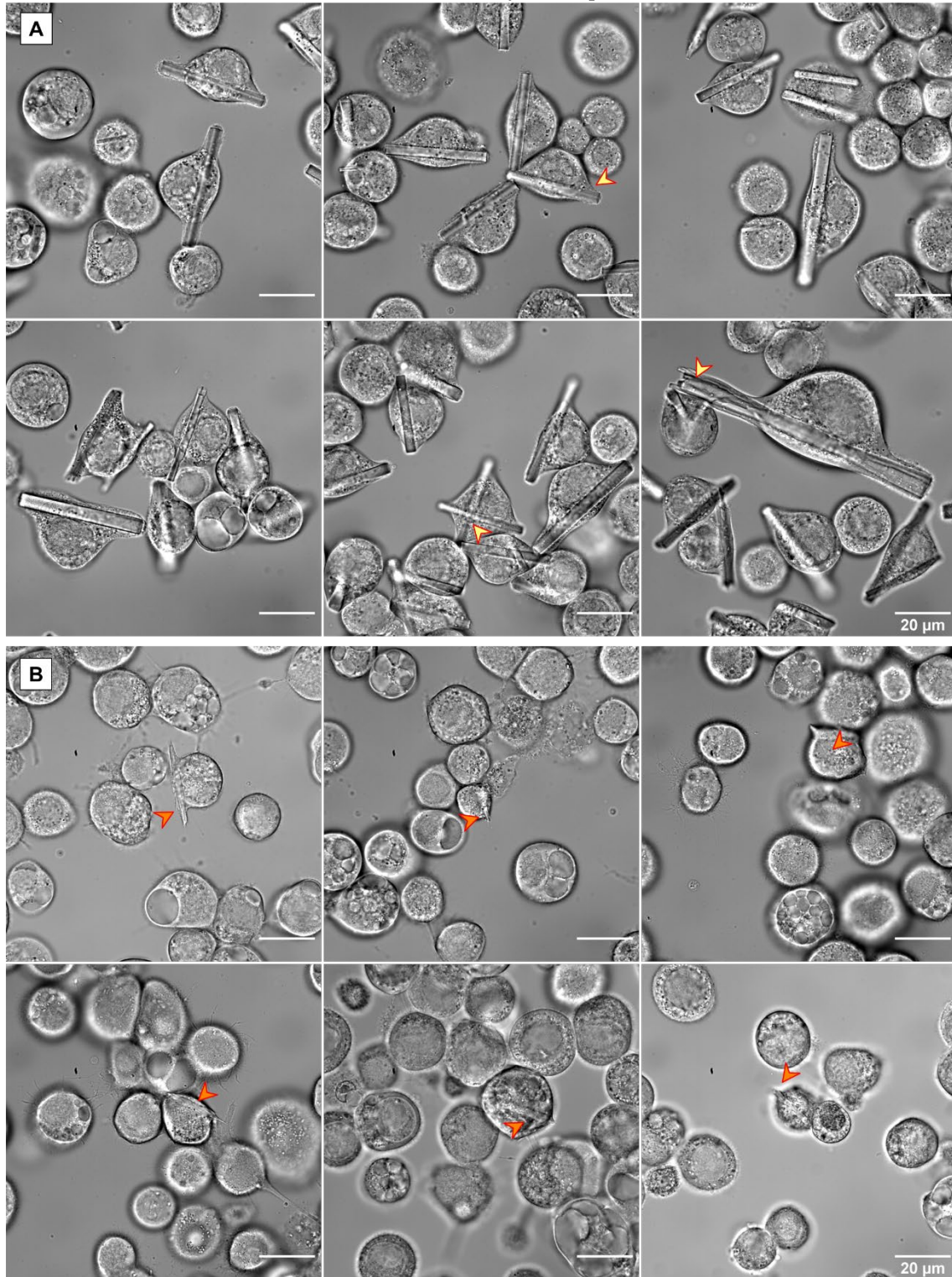
Figure 29 IMPDH N-His crystals can grow in High Five cells both with a Met source or SeMet-containing nutrition only, imaged by DIC.
A At growth conditions with reguar ESF921 medium and **B)** after growth in Met-deficient ESF921 medium with 70 mg/L SeMet supplement according to the SeMet incorporation protocol. Insect cells were infected with rBV IMPDH N-His (MOI 1) in 6-well plates on a glass-slide and imaged on day 5 *p.i.* using a Nicon eclipse Ti2-E with a 100x objective. **A** *Tb*IMPDH crystals grow in thick needle-shaped crystals that frequently exceed the outer diameters of the cell. The plasma membrane spans around the whole crystal(s) as highlighted by a yellow-filled arrow. In a single cell, zero up to five microscopically visible crystals grow. Approximately 60 % of cells show crystal growth.

Longer IMPDH N-His crystals appear silightly broader outside the main cell body and tend to break longitudinally. **B** Although some also exceed the outer diameter of a cell, crystals grown with 70 mg/L SeMet are generally both thinner – less than 2 µm thick - and shorter, approximately 20 µm in length.

To sum up, rBV infection of High Five cells with the three model rBVs encoding the model proteins CatB SKLG, cytoplasma-remaining HEX-1 cyto, and *Tb*IMPDH[134,148] leads to crystal growth with a high percentage of cells growing one or more crystals. The frequency of crystallization events decreases with the application of the developed SeMet incorporation protocol depicted in **Figure 24**. Furthermore, crystal morphologies of HEX-1 cyto spindle-like crystals and IMPDH N-His crystals (**Figure 28 B** and **Figure 29 B**) change. Stress-induced cellular changes due to the addition of SeMet have been observed. Given these changes, an incorporation of SeMet into the *in cellulo* grown crystals seems to be probable but needs experimental evidence.

3.4.4 Qualitative proof of SeMet incorporation to intracellular crystals

The hypothesis of SeMet incorporation into intracellular crystals was investigated by X-ray fluorescence spectroscopy. As a first step, fluorescence emission spectra of intact High Five cells containing *in cellulo* crystals of the model proteins described above, grown in the presence of SeMet, were recorded. These spectra showed the characteristic selenium Kα peak at 11.2 keV (data not shown). However, it is possible that soluble (target) protein that is still abundant in the cytoplasm around the crystals incorporated SeMet, whereas only protein *without* SeMet could be preferably selected for crystal growth.

To exclude the latter possibility, crystals of CatB SKLG crystals were isolated from their cellular environment as previously established for native CatB,[148] and thoroughly washed with PBS. Furthermore, *Nc*HEX-1 crystal isolation was established: Four days *p.i.*, rBV HEX-1 cyto infected High Five cells (MOI 1) were pelleted by centrifugation at 100 rcf for ten minutes at four °C. Excess medium was removed, and cells washed once in PBS. A cocktail containing protease inhibitors (PIC) was added to the cell pellet. To disrupt the membranes, cells were then incubated in RIPA buffer at RT for 10 min, while rolling on a gel shaker inside a falcon tube. The pellet was imaged by LM using a DM IL LED microscope. In cases where the majority of crystals was found free-floating in the buffer, RIPA buffer was exchanged to PBS and cell debris separated from free-floating HEX-1 cyto crystals by sequential centrifugation steps with LM inspection of pellet and supernatant after each step. Otherwise, incubation with RIPA buffer was prolonged for another 10 min shaking at 4 °C before sequential centrifugation steps were initiated. The solution was supplemented with PBS and large cell aggregates that were not chemically disrupted were filtered out using a 50 µm-pore plastic filter. **Figure 31** shows that this isolation procedure results in a dense crystal pellet of HEX-1 cyto crystals, containing very few macroscopically intact cells and cellular debris.

On the resulting crystal pellets, X-ray fluorescence emission spectra were recorded at the EMBL P14 beamline at DESY (**Figure 30**).
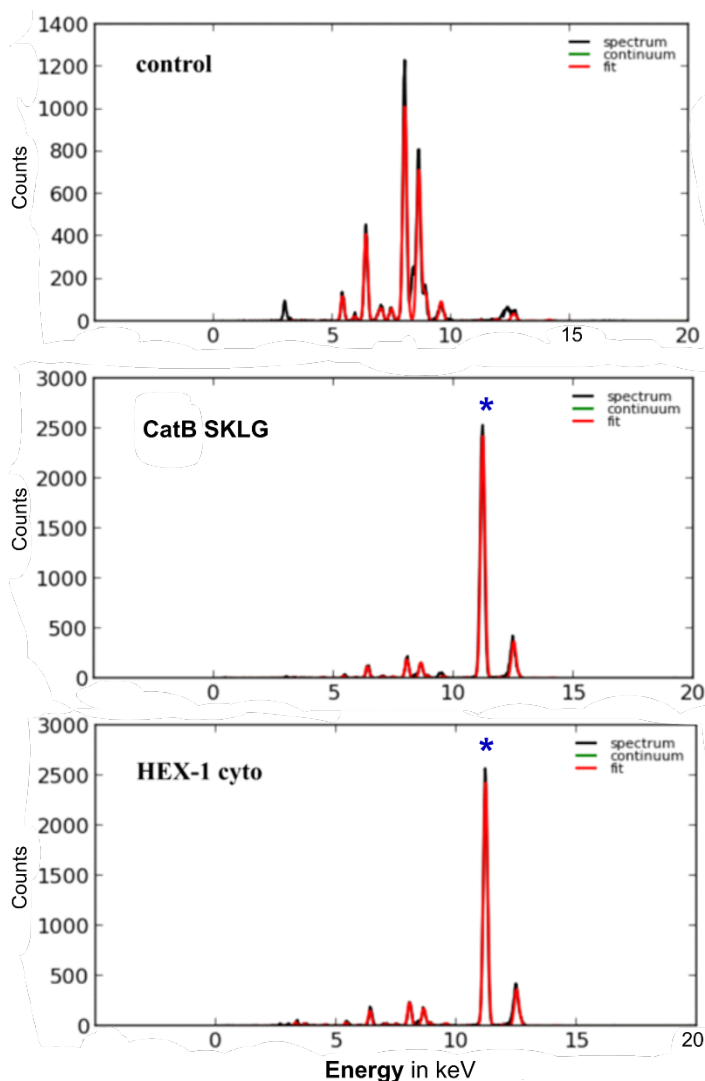

Figure 30 Fluorescence emission spectra of isolated and washed *in cellulo* crystals with a PBS buffer spectrum as control. At day 4 p.i., HEX-1 cyto crystals and CatB SKLG crystals were isolated from High Five cells grown in Methionine-deficient ESF921 medium, supplemented with 70 mg/L SeMet and transferred to 6 °C. X-ray absorption was measured at 12.7 keV in PBS at the EMBL beamline P14 at DESY. The most intense peak is the Se K-alpha at 11.2 keV. As expected, the control did not contain selenium.

Selenium incorporation was verified from isolated crystals via fluorescence emission spectrum for both CatB SKLG and HEX-1 cyto (**Figure 30**). IMPDH N-His crystals grown in the presence of selenium dissolved within the isolation step with RIPA buffer within minutes, preventing data collection of a fluorescence emission spectrum. In contrast, intracellular IMPDH N-His crystals grown in ESF921^(+P/S) including Met are stable within RIPA buffer for hours. The changed crystal properties point towards a successful incorporation of SeMet.

3.4.5 Mapping and quantification of selenium incorporation by mass spectrometry

Following the qualitative proof of selenium incorporation, the objective of the following experiment was to quantify SeMet incorporation into cytoplasmic *Nc*HEX-1 protein crystals by the mass difference between sulfur and selenium, together mapping SeMet incorporation positions. *Nc*HEX-1 was chosen as first target for proof-of-principle experiments. The precise determination of mass to charge ratio is possible by mass spectrometry (MS), a technique used for the analysis of the ionized forms of molecules in the gas phase. By Newton's third law, $F = m \cdot a$, with the known force F, the mass m and the acceleration a, a more massive ion will be accelerated to a lesser extend by the force. Thus, the mass is determined by measurement of the acceleration of an ion in an applied electric field of known intensity. *electrospray ionisation* (ESI) is a technique developed to generate a sufficiently high concentration of ionized protein molecules in the gas phase. A solution of a protein or peptide in a salt-free solvent is sprayed through a thin metallic tip with an applied electric potential, leading to the generation of fine droplets. The solvent evaporates, so that, only protein like-charged molecules remain within the droplet. Due to repulsion, single molecular ions, usually multiply charged, fly in the applied electric field towards the detector to determine the mass, a process called *time of flight* (TOF) analysis.[2]

As a first step, HEX-1 cyto crystal isolation as described above was was performed on High Five cells treated with the SeMet incorporation protocol. For upscaling, liquid cell culture in upright 75 cm² flasks on an orbital shaker, 100 rcf, was used severalfold. Crystal isolates with incorporated SeMet were stored in 0.1 M sodium acetate trihydrate (pH 7.0) and 2 M sodium formate with 0.5 mM DTT instead of PBS, as suggested by Yuan et al;[84] The crystal growth efficiency was found to be higher by visual inspection if cells were infected in an adherent cell culture. However, this finding needs to be validated with suffucient statistics in the future.
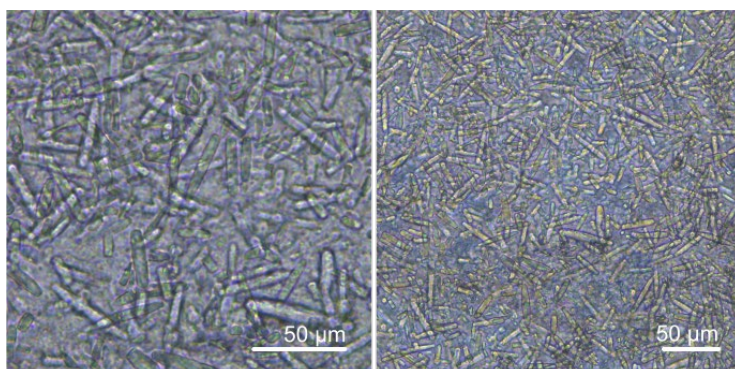

Figure 31 Rod-like HEX-1 cyto SeMet crystals, isolated from High Five cells.
Crystals were isolated from rBV infected High Five cells (MOI 1) 4 days *p.i.* treated with the SeMet incorporation protocol, after cell lysis with RIPA buffer and subsequential centrifugation steps. Inspected with a Leica DM IL LED LM in 0.1 M sodium acetate trihydrate at pH 7.0, and 2 M sodium formate with PIC. Crystals were filtered through a 50-µm filter to remove cell aggregates and imaged with a 40x objective (left panel), and a 20x objective (right panel), using a Leica MC170 HD microscope camera.

Since soluble protein is needed for MS-ESI, the stability of HEX-1 cyto crystals with- and without SeMet incorporation was tested by incubation with the same set of conditions used to examine IMPDH N-His crystal stability (**Table 11**), including detergent, hypotonic buffer, denaturing agent, and a pH range from 3.5 to 9, supplemented with PIC. Monitoring by light microscopy over 24 h at 26 °C revealed that HEX-1 cyto$^{\pm SeMet}$ crystals remained stable in all conditions, except for denaturing conditions using 2, 4 and 6 M urea. The highest urea concentration was found to resolve the crystals best as judged by coomassie stained SDS polyacrylamide gelelectrophoresis (not shown). Since the influence of denaturing agent in MS experiments is unknown, any residual denaturing agent should be removed during gel extraction.

Sample preparation for MS: dissolving *in cellulo* crystals and SDS-PAGE

In preparation for SDS-PAGE, the dense crystal pellet fraction (**Figure 31**) was transferred into a reaction vessel and allowed to settle. The pellet (14 µL yield) appeared in a white-translucent colour. After removal of the supernatant, crystals were mixed with urea in PBS and diluted with HEX-1 sample buffer^{+PIC}, resulting in a concentration of 6 M urea, followed by incubation for 5 min at RT. After thermal denaturation, SDS-PAGE was performed with 3.5 µL per lane. In order to avoid protein cross-contaminants, only one sample type was loaded per gel in addition to the marker. The Coomassie stained 15 % SDS gel from resolubilized HEX-1 cyto^{+SeMet} crystals is shown in the following figure, before (**Figure 32 A**) and after (**Figure 32 B**) gel band extraction.
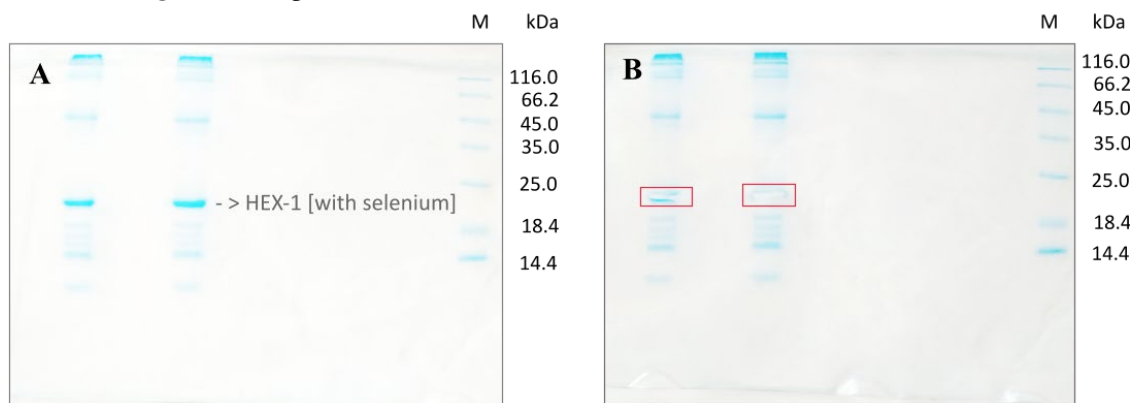
Figure 32 Coomassie stain of a 15 % SDS-gel from resolubilized intracellular protein crystals isolated from rBV HEX-1 infected High Five cells (MOI 1) in Met-deficient ESF921 medium supplemented with 70 mg/L SeMet. The mixture was heated to 90°C after addition of sample buffer while shaking for 10 min. 7 µL of the final sample were loaded for each SDS-gel lane. **A** shows the gel before, **B** after cutting out of the relevant band (marked with red box) supposedly containing the target protein with incorporated selenium. *Neurospora crassa* HEX-1 is a 176 aa protein with a calculated molecular weight of 19.13 kDa. Two bands of isolated *in cellulo* HEX-1 protein crystals, resolubilized in 6 M urea separated by SDS-PAGE, were combined for MS experiments. Remaining cellular debris or membrane proteins resolubilized from whole cells or cell parts within the crystal pellet explain the remaining smear and unknown bands. M: 5 µL Thermo Fischer unstained protein MW marker.

The gel-piece containing the protein HEX-1 cyto and, for evaluation of successful protein gel extraction, the Pierce™ unstained Protein MW marker band were cut out, stored at -20 °C and sent to the group of Dominik Schwudke at the Leibniz research center Borstel. InGel trypsin digestion with reduction and alkylation with iodoacetamide, following the MS-MS analysis with a Quadropol-Time of flight (Q-ToF) and Q Exactive were performed there. An error tolerant MASCOT database search confirmed the identity of HEX-1 cyto (**Figure 33**).
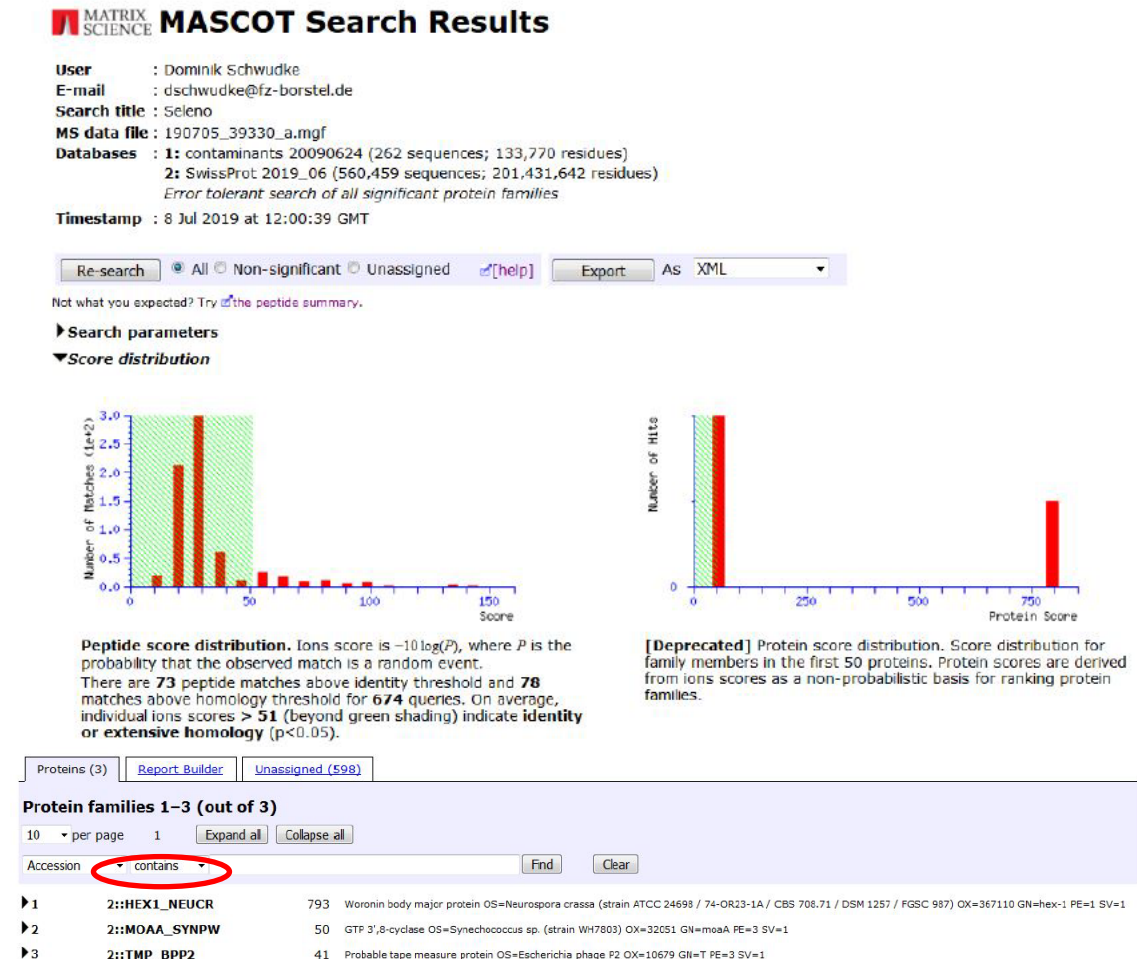
Figure 33 MASCOT database error-tolerant search results confirms protein identity of HEX-1 cyto. The applied significance threshold *p* was chosen to be < 0.05. The MASCOT search takes MS1 spectra of peptides after trypsin digestion as input and uses a probability-based score for the identification of proteins.

This result was confirmed by a BLASTP search using the fasta sequence format of *Nc*HEX-1 (data not shown). The MS^1 spectrum enabled a search for SeMet-containing peptides (**Figure 34**).
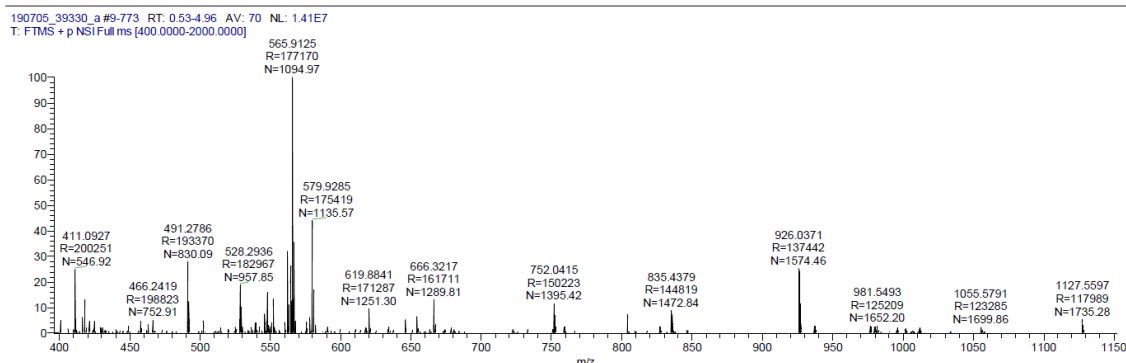
Figure 34 MS¹ spectrum of resolubilized HEX-1 crystals isolated from rBV infected High Five cells.
Infected with MOI 1, following in-Gel-digestion with reduction and alkylation using iodoacetamide.

Unfortunately, the modification S -> Se is not included in standard databases, therefore the analysis of Met / SeMet containing peptides could not be automated.

Selenium has a very characteristic distribution that can be detected with a sensible mass spectrometer. The predicted peptide signals were searched for in the MS¹ spectra by Dominik Schwudke employing *Q Exactive* and aligned using the *LipidXplorer* software.

An example spectrum analysis of the peptide -VLD**M**QDGSIVA**M**TETGDVK- (aa 108-126 of HEX-1 cyto) is shown in **Figure 35**.

Mass/ charge
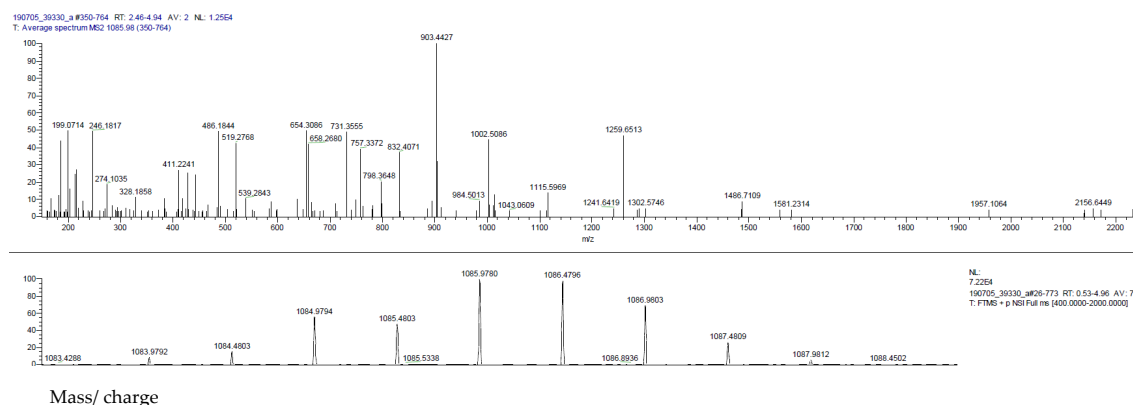
Figure 35 MS¹ spectrum of the peptide from aas 108 to 126: VLDMQDGSIVAMTETGDVK, + 1 SeMet and + 2 carbamidomethyles. The upper panel shows the experimentally determined normalized m/z values of the peptide, the lower panel gives the calculated theoretical spectrum of m/z values.

The isotopic contribution for each signal in the Met containing peptide spectra was determined with *Xcalibur* and employed to normalize mass spectrometric responses for every peptide. This allowed a determination of the percentage of SeMet incorporation for every quantifiable peptide. The intensity based and normalized incorporation rate of selenium was estimated for all three peptides in the sequence of HEX-1 cyto (**Figure 36**) that are containing Met or SeMet, at each MS measurement condition.

```
        10         20         30         40         50         60         70         80         90        100
MGYYDDDAHG HVEADAAPRA TTGTGTGSAS QTVTIPCHHI RLGDILILQG RPCQVIRIST SAATGQHRYL GVDLFTKQLH EESSFVSNPA PSVVVQTMLG
                                                                                         0.38
       110        120        130        140        150        160        170        177
PVFKQYRVLD MQDGSIVAMT ETGDVKQNLP VIDQSSLWNR LQKAFESGRG SVRVLVVSDH GREMAVDMKV VHGSRLA
     0.58                                                              0.42
```

Figure 36 Peptide sequence of *Nc*HEX-1. There are three peptides resulting from trypsin digestion that contain one or multiple methionines, marked in bold and underlined. The peptide containing the start Met was not found in sufficient quantity. The relative incorporation rate of SeMet after SeMet incorporation protocol (MOI 1) as determined from the normalized isotopic contribution of selenium containing signal is given in percent and marked in red.

The relative incorporation rate in the peptide **EMAVDMK** is estimated to be 0.245 (1 SeMet) and 0.177 (two SeMets in the peptide), added up to an incorporation probability of 0.42 (**Figure 36**). Which of the two possible positions in the peptide harboring the SeMet cannot be distinguished with this method.

As previously mentioned, the trend was observed, that rBV infection with a high MOI could maintain cell viability by evading the SeMet toxicity (chapter 3.4.2). Although this was not tested in detail, this should allow the supplementation of the cell culture with higher SeMet concentrations which should result in increased selenium incorporation.

Consequently, the protocol for SeMet incorporation was adapted to a high-titer experiment with a heightened SeMet dose to 90 mg/L. As each virus stock addition introduces a new source of Met (with a high titer in high quantity) to the cells, which in turn reduces incorporation rates, virus was only added once for the primary infection of insect cells. Moreover, 0.5 mM DTT was added starting from the cell lysis step as reducing agent in order to minimize oxidation.

These protocol optimizations have almost doubled SeMet incorporation into HEX-1 cyto (**Figure 37**) to 73 %, respective 92 %.

```
                                                                                           0.73
        10         20         30         40         50         60         70         80         90        100
MGYYDDDAHG HVEADAAPRA TTGTGTGSAS QTVTIPCHHI RLGDILILQG RPCQVIRIST SAATGQHRYL GVDLFTKQLH EESSFVSNPA PSVVVQTMLG
       110        120        130        140        150        160        170        177
PVFKQYRVLD MQDGSIVAMT ETGDVKQNLP VIDQSSLWNR LQKAFESGRG SVRVLVVSDH GREMAVDMKV VHGSRLA
     0.92                                                              0.92
```

Figure 37 Peptide sequence of *Nc*HEX-1, highlighting M-containing peptides. The three peptides resulting from trypsin digestion that contain one or multiple methionines are marked in bold and underlined. The peptide containing the start Met was not found in sufficient quantity. The relative SeMet incorporation rate after SeMet incorporation protocol (MOI 5) as determined from the normalized isotopic contribution of selenium containing signal is given in percent and marked in red.

Since the application of the SeMet incorporation protocol results in significant incorporation rates, as proven by the MS experiments, diffraction data collection for subsequent phase determination using SAD was tested.

3.4.6 *In cellulo* single-wavelength anomalous dispersion tests

Serial *in cellulo* Se-SAD data collection was tested with all three model proteins CatB SKLG, HEX-1 cyto and IMPDH N-His using rBV infections at MOI of 1 and 70 mg/L SeMet medium supplementation. Cell batches were identical to the ones used to isolate crystals for the X-ray fluorescence emission tests described above. Unfortunately, by the time high-titer SeMet incorporation had been established, no further beamtime was available, therefore no diffraction tests with high incorporation rates could be performed. Crystal-containing cells were loaded on mesh-grids (MiTeGen) with 37 % PEG400 as cryo protectant (**Figure 12**). Diffraction data was recorded at the EMBL P14 beamline at DESY, at an X-ray photon energy of 12.660 keV in helical line scans at 100 K, using a beam focus of 5 x 10 μm².

By visual inspection, IMPDH N-His_SeMet exhibited weak diffraction up to approximately 6 Å Bragg-spacing, as opposed to 3 Å resolution detected for intracellular IMPDH N-His crystals without SeMet incorporation. CatB SKLG^{+SeMet} diffracted up to approximately 3 Å resolution, but only 30 of the 1,100 recorded frames showed diffraction at 500 ms exposure, which is comparable to the 3.6 Å resolution detected for CatB SKLG crystals without SeMet incorporation.

3.4.7   SeMet incorporation influences crystal unit cell parameters

Using cells with HEX-1 cyto^{+SeMet} crystals, a total of 25 000 recorded frames were collected at the selenium edge. Bragg spots are detectable up to 2 Å resolution, which is about the same maximum visible resolution as observed for HEX-1 cyto crystals without incorporated SeMet and therefore a very promising result. *CrystFEL indexamajig* processing (indexing with *mosflm*, *DirAx*, *asdf*, *XDS*, without prior lattice or unit cell information) identified 10,480 frames containing diffraction from 3,289 indexable crystals in total, representing an indexing rate of 31.4 %.

Indexing with CrystFEL *indexamajig* without prior information of the symmetry or the unit cell dimensions revealed a wide spread of unit cell parameters, mostly with primitive (black P) unit cells, as shown in the following histogram of unit cell parameters in
Figure 38.

58 crystal patterns were indexed using a triclinic cell symmetry (0.7 % indexing rate).
Using the most prominent unit cell of HEX-1 cyto at cryogenic temperatures (hexagonal, a=b=57.57 Å, c=189.41 Å; α=90° β=90° γ=120°), only six patterns were successfully indexed; The indexing results revealed a very heterogenic composition of the sample consisting of different crystal types; The drastic influence of the SeMet incorporation on the unit cell parameters (
Figure 38) suggests an impact on protein folding, resulting in a highly diverse mixture of different crystals in the cells.
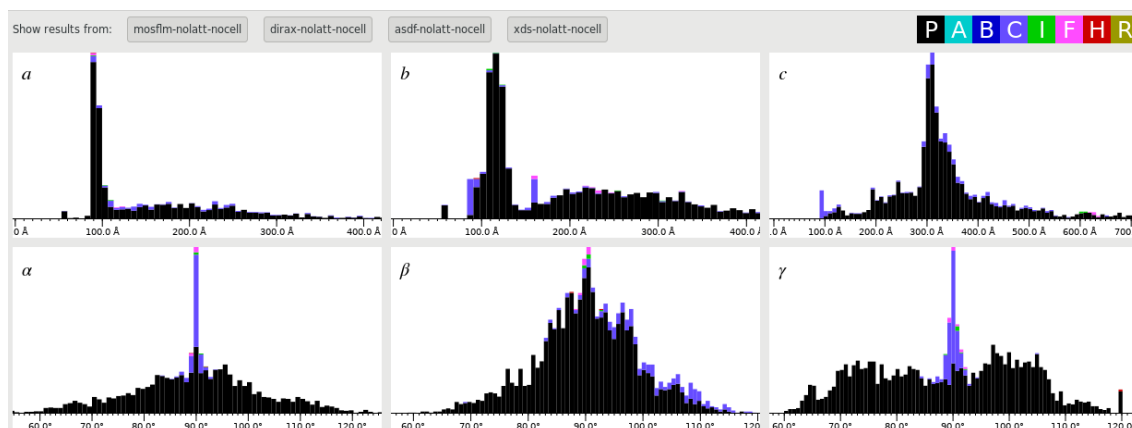
Figure 38 Histogram of unit cell parameters of HEX-1 cyto with incorporated SeMet shows diversification compared to crystals without selenium. Indexing result (mosflm, DirAx, asdf, XDS) without prior information on lattice or cell parameters plotted with the CrystFEL script *cell_parameter*.

Therefore, HEX-1 was investigated with the PDBePISA interface databank (input PDB *8CD4* determined from HEX-1 cyto crystals, data collected at 100 K *in cellulo*) to check if any Met residues were involved in crystal contacts, which could explain changes of the crystal latice. However, all five methionines (Met98, Met111, Met119, Met164 and Met168) that are resolved in the structure are solvent-accessible in the crystal lattice. None forms hydrogen bonds or salt bridges with symmetry-related molecules. Thus, SeMet incorporation is unlikely to influence the crystal lattice, at least without also changing protein structure.

To extract phases of *in cellulo* crystallized HEX-1 by SAD in the future, it is required to significantly reduce the number of unit cell combinations and thus to increase the homogeneity of the sample in terms of crystal types.

3.4.8 Discussion

For truly new and unbiased structures to be solved using the InCellCryst pipeline, methods are needed to specifically place heavy atoms into intracellular protein crystals, towards alternative phasing. In conventional "*in situ*" grown crystals, soaking experiments are a common and well-established method to enable direct phasing; only a skin of mother liquor "protects" the crystal from environmental influences, thus soaking bears the risk that a protein crystal does not survive the soaking experiment. So far, conventional soaking experiments with heavy metals were unsuccessful for intracellular protein crystals in viable insect cells. The *in cellulo* crystals are surrounded by at least one membrane, the plasma membrane, and depending on the crystallization compartment by additional membranes. This effectively shields the crystals from any hydrophilic extracellular molecules that are neither actively transported across nor can diffuse through lipids, leading to a challenge of accessibility in staining *in cellulo* crystals. Heavy atom staining is possible on isolated crystals with soaking experiments, or by permeabilization of the plasma membrane of crystal-harboring cells before soaking, as reported by Boudes *et al.*.[111] However, isolation and

purification of *in cellulo* crystals is experimentally elaborate and needs to be optimized for each crystallized protein target. Also, membrane permeabilization may result in a reduced order of the crystalline lattice and thus in a reduced diffraction power.[71,101,111]

In the scope of this thesis, two approaches towards heavy atom staining were investigated for *in cellulo* crystals, paving the way towards *ab initio* phasing of intracellular protein crystals using SAD. As a first approach, marking with terbium ions via a high-affinity binding tag covalently attached to a crystallized protein was established using *Tb*IMPDH as a model protein. The lanthanide terbium was shosen because it has several properties that make it a good choice for SAD: Terbium has a a large atomic number (Z=65) and a high anomalous scattering power. Tb^{3+} has a relatively small ionic radius (0.923 Å, depending on the coordination)[281], which should allow it to bind to specific sites on the macromolecule without significantly perturbing the protein structure. Also, terbium is commercially available in high purity. However, the relatively low solubility in water compared to other lanthanides with lower atomic number can be a limiting factor.

It could be shown that *Tb*IMPDH with an N-terminal LBT crystallizes in insect cells. It shows a thick needle-like morphology similar to that observed in rBV IMPDH N-His infected insect cells, but of shorter length and a diameter of up to 10 μm. The similarity is not unexpected, because the artificial N-terminus in the SFX IMPDH N-His crystal structure (PDB *6rfu*) is highly flexible[134]. Both the N-terminal LBT and His-tag rest flexibly in water channels of the crystal and do not influence crystal contacts and thus morphology, which is further confirmed by an identical scattering observed by SAXS/XRPD[181]. Since changes in folding have been shown to impact the unit cell of a crystal, this is also deemed unlikely. Of course, the addition of an artificial tag influences both folding and protein structure in other cases.[282,283]

Anomalous differences between members of Bijvoet pair reflections in a diffraction dataset would be the ultimative proof for ordered lanthanide binding to the IMPDH N-LBT crystals. The diffraction tests at the EMBL beamline P14 at DESY revealed a maximum resolution of 3-4 Å for IMPDH N-LBT *in cellulo* without prior soaking at 12.7 keV, but, at a wavelength above the terbium absorption edge, isolated IMPDH N-LBT crystals soaked with Tb^{3+} only showed visible resolution dropped down to 6 Å, further faiding during data collection with less than 1000 indexable collected patterns. In comparison, SFX experiments with isolated IMPDH N-His crystals had allowed structure solution at 2.8 Å (PDB *6rfu*, naturally with crystals irradiated at 12.7 keV).[134] The reduced crystalline order of IMPDH N-LBT is attributed to heavy atom binding. Analysed patterns did not contain detectable anomalous signal – perhaps Tb^{3+} does bind, but unspecifically. Although spectoscropic analysis of High Five cells with IMPDH N-LBT crystals indirectly revealed Tb^{3+} binding to the LBT via FRET[272], there may be competitive lanthanide binding sites.[284,285] FRET is distance dependant

and occurs when the donor fluorophore (here tryptophane) and acceptor (here Tb^{3+}) are in close proximity within 10 nm. Also nucleic acids can act as FRET donors for Tb^{3+} ions,[284,285] and the FRET signal detected at 544 nm from mock-infected insect cells without LBT excited at 260 nm could result from Tb^{3+} affinity to nucleic acids that are abundant in living cells. This finding shows two things: Firstly, Tb^{3+} freely diffuses into the insect cell. Consequently, no lanthanide shuttle system, e.g. a cell penetrating peptide, is needed to deliver Tb^{3+} to the intracellular crystals. Secondly, a FRET signal from terbium ions *in cellulo* is unsufficient evidence for binding to the LBT attached to crystallized model protein. Competitive binding to nucleic acids in the living cell reduces the Tb^{3+} concentration in the cell, and thus the chance of a high occupancy of the target protein-attached LBT with the heavy atoms, which would be necessary for SAD. Increasing the terbium concentration would be necessary to saturate binding, but this directly correlates with cytotoxcicity effects[111,286]. Thus, intracellular Tb^{3+} binding is not a suitable method to allow experimental phasing in the context of *in cellulo* crystallography. Applicable heavy metal ion candidates should interact with the target protein with high specificity and exhibit a low unspecific binding[272], while cytotoxicity should not prevent *in cellulo* crystal growth at required metal ion concentrations. In future studies one could test ions that are already abundant in the cell such as zinc, and clone the corresponding natural binding motive, such as a zinc-finger, to the protein of interest.

Competing binding sites at nucleotides in the nucleus may not be a problem for phasing *Tb*IMPDH after all, because terbium ions could also attach to the IMPDH-bound adenosyl or guanosyl residues in an ordered manner. Therefore, for future studies, a complete dataset of isolated IMPDH N-LBT crystals, soaked at increased lanthanide concentration, remains interesting.

In a second approach to heavy atom staining *in cellulo*, selenium was incorporated during gene expression as the Met derivative SeMet. The advantage of SeMet-SAD is that the native target protein sequence can be used as no tag for heavy atom binding needs to be accommodated. Instead, selenium is incorporated into the protein during translation as SeMet. This has been successfully applied in insect cells before for labeling proteins that were then conventionally crystallized.[277]

It was estimated that one SeMet would be required every 75 to 100 aa for a suffessful SAD/MAD experiment. This corresponds to approximately 80 % of all proteins that all have a Met content of at least 1 %.[287,288] However, this calculation assumes incorporation rates close to 100 % and ordered SeMet residues within the protein crystal. Whether a residue is largely ordered depends on its position within the target sequence and cannot be influenced, reasoning that efforts need to concentrate on improving SeMet incorporation efficiency.[289]

Toxicity of SeMet to insect cells is well recognized in the community.[278,290] However, no concluding studies were published at the time of the experiments reported herein regarding which quantities of SeMet induce toxicity as measured by means of cell viability. Studies on the mechanisms of SeMet toxicity in eucaryotes have mostly been performed with the yeast *Saccharomyces cerevisiae*[290–292]. Two mechanisms may play a role in SeMet cytotoxicity: Firstly, induction of reactive oxygen species generation due to SeMet[290], effectively leading to cell growth inhibition. Secondly, random incorporation of SeMet to different sites in proteins may promote protein aggregation and additionaly lead to protein fragmentation that cause proteotoxic stress in the cell.[293]

Experiments confirmed toxicity of SeMet to insect cells by an increased cell mortality, depending on the SeMet concentration and morphological changes in the cell culture, such as the formation of syncytiae and an increase in vacuole formation. Furthermore, the influence of the virus titer on SeMet toxicity was investigated and revealed a trend that higher virus titer reduces the cell mortality during SeMet supplementation, even though absolute cell vitalities vary between experiments and cell cultures. This is in agreement with Wenzel and coworkers[278].

Why should a high virus titer reduce cell toxicity of SeMet? One possibility is that multiple infections of the insect cell and the associated boost in protein production forces all SeMet to be incorporated into proteins, leaving less to cause damaging reactive oxygen species. Another is related to the cell cycle arrest in the G_2/M phase that is initiated by the rBV infection.[294] It culminated in a consistent inability of cellular DNA replication, thus any DNA damage induced by SeMet could lose vital importance.[278]

Does incorporation of SeMet influence the intracellular protein crystals? A qualitative proof that selenium was built into the crystals was obtained by X-ray fluorescence emission from isolated crystals. The diffraction tests performed *in cellulo* for rBV MOI 1 infected High Five cells carrying IMPDH N-His, *Nc*HEX-1 and CatB-SKL crystals, all treated with the SeMet incorporation protocol, highlighted the importance of the target sequence. This makes sense because both Met content and incorporation positions differ for the three model systems. Some diffraction images were obtained for all three samples. IMPDH N-His SeMet crystals showed reduced order due to SeMet incorporation, diffracting only up to 6 Å. Unsufficient data could be collected from CatB-SKL[+SeMet] crystals due to reduced crystallization efficiency. HEX-1 cyto[+SeMet] crystals showed the same order and crystal formation efficiency as HEX-1 cyto[-SeMet]. However, subsequent processing with *CrystFEL* revealed a highly inhomogenous spread of unit cell parameters of the indexed lattices, with no apparent dominating subgroup of which a dataset could be formed for phasing and structure solution.

This is an unexpected finding: Why does selenium incorporation into intracellular HEX-1 protein crystals influence the unit cell parameters? SeMet incorporation is widely used with

in vitro crystallized proteins, because in most cases, the exchange of sulfur with selenium does *not* influence the protein fold.[295] In fact, the *aim* of direct phasing is to preserve isomorphism.[14,296] Sulfur and selenium are both elements in the sixth main group of the periodic table; the 2.32 Å atomic diameter of covalently bound selenium is 0.28 Å larger than the diameter of covalently bound sulfur. In nature, this can facilitate the formation of selenenyl-sulfide bridges in enzymes such as mammalian thioredoxin reductases. These enzymes have a naturally occurring selenocysteine in the active center, resulting in increased reactivity.[297] Therefore, the starting hypothesis was that the unit cells are still the same, but diffraction processing needs optimization. However, even after through optimization of the collected data (double-checked dark maps, iteratively optimized peak detection, optimized detector geometry), still no consistent unit cell with more than 50 indexed patterns was found (**Figure 38**). In addition, none of the Met in the structure by PDBePISA[298] analysis were proposed to participate in crystal contacts (Protein Interfaces Surfaces and Assemblies, http://www.ebi.ac.uk/pdbe/prot_int/pistart.html).

For *Nc*HEX-1, there are 64 different incorporation distribution possibilities for the six methionines in the sequence: Two exponentiated by the number of methionines in the structure. It is possible but unusual that proteins with SeMet incorporations at different positions crystallize together at a fixed ratio, which would result in a crystal with statistically determined occupancies at each position in each crystal. Crystallization is known to purify isoforms[42], even though there are exceptions like the heterogenous glycosylation of the native *in cellulo* crystals from the viviparous cockroach *Diploptera punctata*.[73] It is furthermore possible that fixed combinations (e.g. Met98, SeMet112, SeMet120, Met165, and SeMet168) would result in crystallization.

The latter hypothesis was challenged by MS experiments on isolated and dissolved HEX-1 cyto[+SeMet] crystals, to determine the amount and the position of SeMet incorporation in the protein. The analysis of MS data relies on the frequency of isotopes in the sample.

It includes the comparison of signal intensity between the theoretical signal strength expected if this isotope were 100 % of the sample, to the measured intensity. This results in the approximated percentage of the isotope for the specific peptide, for example EMAVDMKV with one SeMet, one Met and two carbamidomethyl derivatives. As expected, quantitative analysis by MS revealed an inhomogenous incorporation of different peptides: The initial experiments with rBV HEX-1 cyto infected High Five cells using a MOI 1 resulted in an estimated SeMet incorporation of 38 to 52 %. Several oxidation states occured at the Met/ SeMet, which may explain the heterogeneity of crystal unit cell dimensions, provided that these oxidation states do not change during crystal isolation. The possibility to incorporate SeMet at all positions increases with higher incorporation rates of course, which minimizes the number of combinations and thus could result in a more homogenous distribution of unit cell parameters of the crystal batch. The observed trend of increased cell vitality, when infected with a high virus titer before SeMet treatment, allowed useage of

higher SeMet concentrations during gene expression. Indeed MS analysis revealed that SeMet incorporation of HEX-1 cyto increased to 73 % – 92 % using the optimized procedure at MOI 5; this is higher than the 70 – 75 % SeMet incorporation reported by Cronin *et al.*,[277] and in the upper range of other reports with SeMet incorporation intracellularly produced protein from 40 % to 100 %.[277,299–301]

Directly measuring the amount of selenium ejected in the gas phase would allow scaling of these SeMet quantifications from the sulfur-selenium stoichiometry of ESI-MS results gained in this work. This could be achieved by inductively coupled plasma-MS[302,303], but only after first resolving the challenge to get the resolvated *in cellulo* crystals into the plasma phase.

Whether the increased incorporation rate of SeMet into HEX-1 cyto indeed reduces the unit cell heterogeneity of the crystal lattices and whether this applies to other proteins remains to be verified in future experiments. However, in the light of the recent protein folding prediction revolution with AlphaFold 2[171], further efforts into this direction have become less pressing.

4 Conclusion

A phenomenom occurring throughout all domains of life,[71] it is still not clear whether intracellular crystallization of proteins after recombinant gene expression is a rare phenomenon, occurring only in a small number of proteins prone to self-assemble in an ordered way, or rarely observed. Intracellular protein crystals are often hard to detect[182] and most working groups that perform protein production and purification from insect cells do not seriously look for signs of crystallization.

To exploit the immense potential of *in cellulo* protein crystallization, it is necessary to understand the cellular processes involved in crystal formation in more depth. Identification of biological parameters that are suitable for screening approaches like the developed SAXS/XRPD method could further widen the applications of *in cellulo* crystallization.[182] First parameters that influence and control the size and shape of crystals as well as their crystallization efficiency have been determined:

Compartment screening has been identified as a screening parameter. The InCellCryst pipeline optimized in the scope of this thesis allows rapid screening of target proteins into various insect cell organelles. Each cellular compartment has a unique chemical environment and provides a confined space for expressed genes to crowd. This, combined with an intrinsic crystallization tendency of the target protein, apparently can trigger *in cellulo* crystallization, maybe in in actively promoted manner by the living cell.[71,101,130,148] It could be demonstrated that while some proteins have the propensity to crystallize in several compartments, and *Nc*HEX-1 even in all compartments (intrinsic crystallization propensity), most crystallize only with or without posttranslational modifications. Consequently, proteins crystallize only in the ER or only outside, for example in the cytoplasm or the nucleus.[181] Proteins may crystallize in a compartment different to the targeted one, because transport pathways may be saturated[304] due to protein production based on the strong polyhedrin promotor. Saturation would cause further protein accumulation. Provided that an ordered aggregation was energetically more likely or somehow actively promoted by the cell – transport pathway saturation could also cause nucleation of proteins[121] still within the secretory pathway. For example, *Tb*CatB is equipped with a native signal sequence and crystallizes in the ER, not further along the secretory pathway.[41,101] In contrast, live cell imaging with co-infected cells of *Tb*IMPDH and fluorescent membrane compartment tags show that *Tb*IMPDH crystals grow within the cytosol[134], despite of the native C-terminal 'SKL' translocation signal targeting peroxisomes.

Also, it could be shown that by optimizing the virus titer/ MOI, the crystallization efficiency can be increased, but only up to a target-dependant limit.[181] Intrestingly, high virus titers can reduce cellular stress, as induced by Met starvation and SeMet supplementation.

Therefore, I recommend systematically checking the influence of virus titers, away from the optimal MOI 1 determined for under standard conditions, when testing the influence of

supplemented chemicals during gene expression on intracellular protein crystallization in the future.

Three examples (CatB SKL, IMPDH N-LBT and Xpa[305]) have shown that sequence modifications may alter either maximal crystal sizes[306] and/ or crystal stability. Therefore, it appears reasonable to assume that exploiting knowledge for example of crystal contact positions might allow growing a targeted *in cellulo* crystallite size-range in the future. Amino acids involved in crystal contacts could be mutated to increase the binding energy by designing salt bridges and facilitating hydrogen bond formation, which should stabilize the crystal and possibly increase crystal order. On the downside, without additional biochemical assays on the mutated enzyme, as performed by Buey and coworkers for AgIMPDH[211,307], assumed enzyme activity of protein constructs with any artificial amino acids remains hypothetical.

Furthermore, it could be shown that the choice of the insect cell line influences intracellular crystal size. Initially this was demonstrated by Fan *et al* in 1996[132]. This finding was confirmed for variants of *Tb*IMPDH, firefly luciferase, *Tb*CatB, HEX-1 cyto and EGFP-μNS, published in the InCellCryst publication. In addition to crystal size, also the number of crystallites per cell is influenced by the insect cell line.[181]

What is the reason for choosing insect cells as containers for crystal growth, as opposed to bacterial, yeast or mammalian cells? The most straight-forward answer would be: Because it works. More recombinant proteins have been successfully crystallized in insect cells than in cells of any other domain of life[71], and the outcome of a costs/merits/handling-time evaluation as opposed to crystallization in mammalian cells supports this choice: Being eucaryotes, insect cells, as opposed to bacteria or archae, are organized in membrane-devided cellular compartments, which opens up a small degree of control on crystallization conditions *in cellulo* by enabling the aforementioned compartment screening. Yeast cells can grow only crystallite sizes limited to the confinements of their rigid cell wall[121,308]. Insect cells provide the opportunity for posttranslational modification similar if not identical to that of mammalian cells, including *N*-linked and in some cases *O*-linked glycosylation patterns. Humanized *N*-glycosylation baculovirus-insect cell systems have been developed[203] and could be used for crystallization tests of glycoproteins with a more authentic posttranslational modification.

Finding parameters that limit rather than enhance protein crystallization *in cellulo* is easier: Cellular stress resulting from shear forces in rBV infected suspension culture (rather than grown in adhesion) declines the crystallization efficiency, compared to rBV infection in adhesion culture. Limiting nutrition by using a minimal medium as for SeMet labeling experiments also has a dampening effect on crystallization efficiency.

What about a temperature influence during crystal nucleation and growth? Sommer asserts in his PhD thesis that reduced temperature during infection resulted in reduced
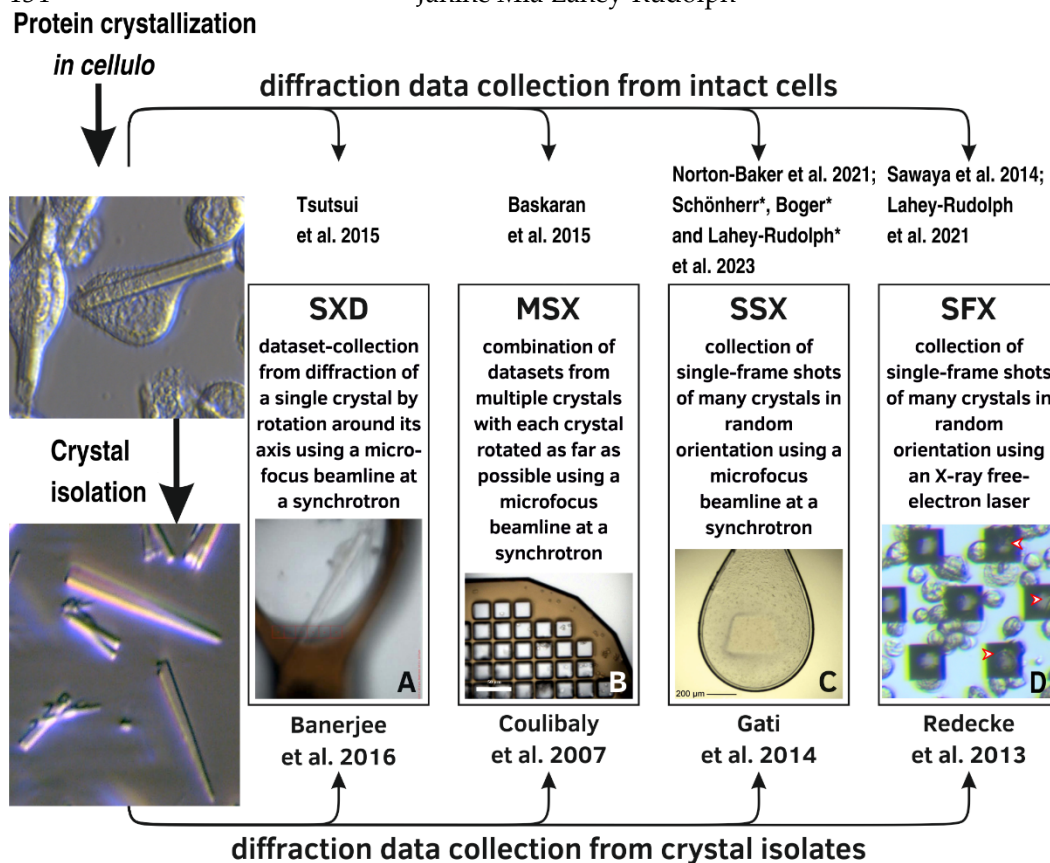
crystallization efficiency.[309] Maybe lower temperature during late gene expression and thus in the crystal growth phase can lead to more ordered crystals that diffract better due to reduced mosaicity.

Additional parameters need to be identified in the future as tools to influence the phase diagram of recombinant proteins *in cellulo,* towards controlled synthesis of intracellular protein crystals with regard to crystal quantity, size, order and morphology.

Once intracellular crystallization is achieved *in cellulo* crystals can be exploited for structure determination.

Method development of data collection *in cellulo* started in 2014 with SFX from toxin crystal-containing bacteria. Here, data collection of crystals from recombinant proteins inside viable insect cells was established using both synchrotron radiation[181,233] and XFEL sources[183], continuing and contributing to method development in the field of SX using *in cellulo* crystals (**Figure 39**). No apparent drop in resolution due to background scattering by the cellular environment was observed, consistent with results by other groups who compared polyhedra diffraction approaches on CPV polyhedral crystals[111,232]. Diffraction in intact insect cells is strongly recommended; An exception are methods where dense crystal pellets are beneficial, as the surrounding cells naturally limits a crystal pellett concentration. Also, whenever pure protein is needed, such as for MS with resolubilized proteins from *in cellulo* crystals, or if technical reasons prohibit data collection in intact cells, isolation and purification of intracellular protein crystals remains necessary. Crystal isolation and purification protocols for HEX-1 cyto (± SeMet), IMPDH N-His (± SeMet) and CatB SKLG (± SeMet) have been successfully implemented in this work.

Regarding data collection strategy, to exhaust the maximum dose per interaction point of the crystal, in theory a 2 μm step-size in between each oscillation would be best to circumvent secondary radiation damage due to inelastic stattering by photo electron escape[33] (personal communication with Elspeth Garman, August 2019). However, in my experience, at least three, but better five, subsequent diffraction images of the same rotated crystal are needed for a successful identification of a rotation series with *dozor*. Therefore, irradiating with no additional translation other than the 5 μm-horizontal step-size is a compromise to allow successful XDS data processing. This compromise is unnecessary for targets with larger diffraction volumes, e.g. in the size-range of *Photinus pyralis* luciferase *in cellulo* needle-shaped crystals. Luciferase crystals could be positioned in a favourable orientation on the sample holder.

Protein crystallization

in cellulo

diffraction data collection from intact cells

Tsutsui et al. 2015

Baskaran et al. 2015

Norton-Baker et al. 2021; Schönherr*, Boger* and Lahey-Rudolph* et al. 2023

Sawaya et al. 2014; Lahey-Rudolph et al. 2021

Crystal isolation

SXD
dataset-collection from diffraction of a single crystal by rotation around its axis using a micro-focus beamline at a synchrotron

MSX
combination of datasets from multiple crystals with each crystal rotated as far as possible using a microfocus beamline at a synchrotron

SSX
collection of single-frame shots of many crystals in random orientation using a microfocus beamline at a synchrotron

SFX
collection of single-frame shots of many crystals in random orientation using an X-ray free-electron laser

Banerjee et al. 2016

Coulibaly et al. 2007

Gati et al. 2014

Redecke et al. 2013

diffraction data collection from crystal isolates

Figure 39: **Methods of X-ray diffraction data collection that have been applied successfully on intracellular protein crystals after crystal isolation (lower part) or on crystals within the intact cells (upper part), and selected publications. A** A single crystal within a cell mounted in a 200 μm cryogenic loop[126]. **B** Isolated crystals loaded onto a 25 μm MicroMesh[109]. **C** Suspension of isolated crystals from an Sf9 cell culture scooped up with a 700 μm cryogenic nylon loop[31]. **D** High Five cells harboring *Nc*HEX-1 crystals on a micro-patterned silicon chip. SXD, single-crystal X-ray diffraction using synchrotron radiation; MSX, multiple-crystal synchrotron crystallography. Image modified from Schönherr, Rudolph and Redecke (2018).

Pre-location of detected crystals off-beam with marked positions on the sample holder would increase hitrates and be applicable in cases of high-profile targets with low crystallization rates in cells. This approach has been established at the PSI beamline at the SwissMX for SFX with protein crystals of soluble targets and membrane proteins in LCP. Prelocating crystals minimizes the time needed to collect a complete dataset at the FEL, where beamtime is very limited, for the prize of longer so far mostly manual preparation time at a synchrotron that has to be established with the necessary soft- and hardware[244].

The choice between cryogenic or ambient temperatures during data collection is a weighting of radiation damage and observable dynamics. Data collection at cryogenic temperatures can be very useful to achieve high-resolution diffraction, because of the attenuated radiation damage at low temperatures.[18,19] However, room-temperature data collection has become an attractive alternative with the recent advances made in synchrotron sources, beamline optics and detector technique. Advantages of RT data collection are the possibility to see alternative conformations in important structural regions more easily observed at ambient temperature

due to increased flexibility[310,311]. Furthermore, no cryo-protectant is needed, and in principle time-resolved studies are possible[310,312–314], provided that the conformational changes do not destroy crystal contacts to a degree that the crystal collapses.

In this work, *in situ in cellulo* data collection at RT in CrystalDirect™ plates has proven worthwhile; Since noteworthy alternate confirmations in *Nc*HEX-1 structure (Val157 only) were not observed, not so much due to the ambient temperature, but rather as a quick screening opportunity that is also suitable for fragile crystals that would be damaged during transfer or lose diffraction power due to cryoprotectant treatment.

Outlook

Low crystal efficiency is remaining a serious limitation towards a broader usage of crystallography with *in cellulo* crystals. In addition to faster data collection, complete datasets could be obtained in reasonable time if it was possible to sort cells containing crystals from those that do not. Inspired by Coulibaly 2017[112] who achieved an enrichment with polyhedral crystals, we tested FACS on infected insect cells[181,306]. An indirect enrichment of crystal-carrying cells seems to be possible[111,232], but the challenge of diminished cell viability after shear forces still has to be solved.

Data collection should use robot-assited mounting to minimize an impact of the experimentator experience. The blotting process needs be optimized to prevent salt precipitation on MicroMesh grids.

Should method development towards *de novo* phasing of *in cellulo* crystals be continued, despite recent accomplishments and growing accuracy of in silico models predicted by AlphaFold2[171,172]? In favour of investments to advancing *de novo* phasing approaches, AlphaFold2 is dependant on accuracy and reliability of prediction models. Depending on sequence coverage it then returns a confidence (*pIDDT*) for each amino acid; an incertainty which can be avoided by determining the phase experimentally.

Folding predictions of multimeric proteins can still be difficult, especially long-range interactions, and model predictions must be validated by experimental data.[315] Even though the two approaches towards *de novo* phasing diffraction of intracellular protein crystals clearly succeeded in incorporating heavy atoms to viable insect cells, both would need considerable adaption to be used in phasing. The lanthanide terbium is not a suitable heavy-atom label for SAD due to competetive binding to nucleotides[275]. SeMet incorporation could enable direct phasing, provided that high virus titers during the starvation protocol are applied and the resulting optimized selenium incorporation does indeed homogenize the crystal batch properties. This needs to be validated, also on targets other than *Nc*HEX-1.

An aim of this PhD project as part of the InCellCryst project was to inspire structural biologist groups to use *in cellulo* crystallization as an alternative approach to conventional crystallization methods. Slowly, awareness and usage of the methodology grows.[316] The EuXFEL for instance enables *in cellulo* crystallization[317] in the XBI laboratory. Also, the synchrotron SOLEIL in France started building a platform for intracellular crystallization using mammalian HEK293 cells.[185] They plan to adopt the compartment screening presented in the InCellCryst[181] pipeline.[185]

The results of this work pave the way to SX *in cellulo* as a serious method tool-box addition to conventional *in vitro* crystallization approaches, with a more efficient detection and usage of crystal-containing cells as suitable targets for SX at synchrotron sources and XFELs.

5 Materials and Methods

5.1 Materials

5.1.1 Devices

Table 5 Used devices.

Device type	Name (manufacturer)
Agarose gel assembly	Easy Phor mini, 7x 7 cm and 7x 10 cm (Biozym Scientific) And: Easy Phor Midi, 10 x 7 cm and 10x 10 cm (Biozym Scientific)
Balance	ABS-N/ABJ-NM (KERN & SOHN GmbH) Kern PCB & Kern 572 (KERN & SOHN GmbH)
Cell incubator	Cooled Incubator MIR-154 (Panasonic Healthcare Co.) B 5050 E (Heraeus Holding GmbH) Kühlinkubator MIR-553 (Sanyo)
Centrifuges	Sigma 112 Mini Centrifuge (Sigma Laborzentrifugen GmbH) Centrifuge 5424 and 5804 R (Eppendorf)
Gel documentation	UV-transilluminator (UVP) UV-transilluminator (Bachhofer Laboratoriumsgeräte)
Incubation shaker	Mixing Block MB-102 (Biozym Scientific) Orbital Shaker PSU-10i (bioSan Laboratories)
Microscope	DM IL LED with MC170 HD microscope camera (Leica Camera AG) Nicon Ti Eclipse (Nicon Instruments Inc) Axiovert 40 CFL (Zeiss)
Micro-meshes	MicroMesh 25 μm apertures/ 700 μm diameter (Mitegen)
PCR cycler	Mastercycler personal, gradient (Eppendorf) T gradient (Biometra GmbH)
pH meter	FiveEasy (Mettler Torledo)
Spectrophotometer	Nanodrop 2000c (Thermo Scientific Peqlab) Jasco® FP-6200 spectrofluorometer (JASCO, Japan)
Sonificator	Branson Sonifier 250/450 (Emerson Electric Co, USA)
Thermo cycler	Biometra® TGradient (Biometra GmbH)
Voltage supply source	Power Pac 3000 (Bio-Rad Laboratories)
X-ray hybrid photon counting detectors	PILATUS 6M (Dectris) EIGER 4M (Dectris) EIGER 16M (Dectris)

5.1.2 Buffers and media

Table 6 List of used buffers and media compositions.

Name	Concentration
Coomassie stain	40 mL methanol
	150 mL ddH$_2$O
	16 g ammonium sulfate $(NH_4)_2SO_4$
	2.78 mL 85 % phosphoric acid
	160 mg Coomassie G625
	[mixing for 15 minutes]
Insect cell medium	ESF921 (Expression Systems)
50x TAE	2 M Tris
	50 mM EDTA
	0.5 M acetic acid
	pH 8.3
PBS	137 mM NaCl
	2.7 mM KCl
	12 mM Na$_2$HPO$_4$ pH 7.2
TBS	20 mM Tris
	137 mM NaCl
	2.7 mM KCl
	pH 7.0
TBS with TbCl$_3$	20 mM Tris
	137 mM NaCl
	2.7 mM KCl
	5 / 10/ 25/ 50/ 100 mM TbCl$_3$
	pH 7.0
LB medium per 1 L	10 g tryptone
	10 g NaCl
	5 g yeast extract
6 x Loading dye	0.35 M Tris-Cl, pH 6.8
	0.35 M SDS
	30 % (v/v) glycerol
	0.6 M DTT
	0.175 mM bromphenol blue
HEX-1 reservoir buffer	50 mM Tris, pH 7.5
	200 mM NaCl
	30 mM CaCl$_2$
HEX-1 SeMet reservoir buffer	50 mM Tris, pH 7.5
	200 mM NaCl
	30 mM CaCl$_2$
	0.1 M sodium acetate trihydrate
	2 M sodium formate, pH 7.0
	+ fresh reducing agent TCEP & PIC

IMPDH LBT-N storage buffer	10 mM magnesiumchloride-hexahydrate in dH$_2$O
Micro-mesh (Mitegen) cleaning solution	2 % SDS in dH$_2$O
SeMet supplemented insect cell medium	SeMet (Acros Organics) in ESF921 Met-deficient medium (Expression Systems)

5.2 Methods

5.2.1 Agarose gel electrophoresis

After boiling 0.8-2 % agarose in TAE buffer, the solution was poured into a chamber (Biozym) with 6 µL Midori Green advance (Nippon Genetics) and allowed to cool down with the inserted chosen comb. The gel was covered with TAE buffer. Samples were loaded with loading buffer. A Power Pac 300 power supply (Biorad) was connected to the chamber and typically adjusted to 95 V, approximately 70 mA, for nucleic acid separation, which takes approximately 30 minutes. Subsequent detection of DNA under UV light follows.

5.2.2 DNA restriction

To generate blunt-ends (5'-GGC|GGC-3') for blunt-end cloning, vectors were linearised with restriction enzyme *Ehe*I (Thermo Scientific), an analogon of *Sfo*I. In a total volume of 20 µL, 2 µg of vector DNA were mixed with 0.1 µL of the restriction enzyme and 2 µL of the FastDigest Buffer (10 x, Thermo Scientific). The solution is incubated for 30 minutes at 37 °C.

5.2.3 DNA ligation

Ligation of the PCR product –the insert – and a linearized vector was performed in a molar vector-insert ratio of 1:3 or 1:5 according to the following protocol. *Ehe*I was added in order to minimise empty-vector-religation.

Table 7 Ligation protocol optimised for blunt-end cloning.

Component	Concentration or amount
vector	50 ng
purified PCR product	3 to 5 x molar excess
T4 DNA ligase	1 U
T4 DNA ligase buffer	1x
*Ehe*I	0.1 µL
dH$_2$O	ad 20 µL

5.2.4 Generation of chemically competent *E. coli* with CaCl₂ method

On day 1, a preculture with 5 mL LB medium with antibiotics and previously prepared chemically competent *E. coli* was incubated overnight at 37 °C, 150 rpm.

On day 2, in a cold room or on ice, 100 mL LB medium was inoculated with 2 mL of the preculture. At 37 °C, the bacteria were grown up to an OD600 of 0.6 – 0.8, then cooled down on ice for 10 minutes in two 50 mL falcon tubes. Cells were pelleted at 5000 rcf for 10 minutes, and the supernatant carefully discarded. The cells were resuspended in 10 mL cold 0.1 M CaCl₂/10 % glycerin solution and cooled on ice for 30 minutes. Then, cells were pelleted for 10 minutes at 5000 rcf. Supernatant was discarded and each pellet carefully resuspended in 1 mL cold 0.1 M CaCl₂/10 % glycerine. 50 μL aliquots were shock-frozen in liquid nitrogen and stored at -80 °C.

5.2.5 Transformation of chemically competent *E. coli*

Per ligation reaction, one 20 μL aliquot of competent DH5α *E. coli* cells (Stratagene) were thawed on ice. 2 μL of the ligated vector or (100 to 300) ng DNA were added and incubated for 20 to 30 min on ice. After 45 seconds of heatshock at 42 °C, cells were cooled down on ice for 2 minutes. 200 μL of plain LB medium without antibiotics was added and the reaction vessel incubated at 37 °C for 1 hour. The cell suspension was then plated to agar plates with antibiotics for overnight incubation at 37 °C.

5.2.6 Colony PCR

A PCR reaction master mix was prepared as follows in **Table 8** and aliquoted to reaction tubes.

Table 8 Reaction master mix for colony PCR using a HighQu AllIn HiFi DNA polymerase.

Component	Concentration or amount
Water	14 μL
5x buffer	4 μL
ALLin HiFi DNA polymerase	0.05 μL
fwd insert primer	1 μL of 10 μM
rev sequencing primer vector	1 μL of 10 μM
total	20 μL

A colony was picked and submerged in the PCR reaction mix for 2 seconds. 5 mL of LB medium with appropriate antibiotics were inoculated with the same tip and incubated over night at 37 °C and 200 rpm.

5.2.7 Plasmid and bacmid isolation

For the isolation of plasmids and bacmids, well-isolated colonies were picked from a LB agar plate and inoculated with 5 mL LB medium over night, at 37 °C and 200 rpm. Plasmids and

bacmids were isolated according to manufacturer's instructions using a GeneJet Miniprep Kit (Thermo Scientific). DNA concentration was estimated by the 260 nm adsorbtion with a Nanodrop ND-2000 (Thermo Scientific), under the assumption that an adsorbtion of one equals a DNA concentration of 50 ng/ μL.

A PCR was run for 30 cycles with the following standard parameters for a colony PCR, with an elongation time of 30 s/kb of expected PCR product and an annealing temperature chosen with this calculator by BioLabs [*https://tmcalculator.neb.com/#!/main*].

Table 9 Standard colony PCR program with a HighQu AllIn HiFi DNA polymerase.

step	temperature	time
initial denaturation	95 °C	2 min
denaturation	95 °C	20 s
annealing	calculated	20 s
elongation	68 °C	30 s/kb
final elongation	68 °C	2.5 x elongation time

Successful ligation of the insert in the correct direction was evaluated by a 1 % agarose gel electrophoresis of the PCR product.

5.2.8 Sequencing of plasmids

Plasmid sequencing was outsourced to the LGC Genomics *Ready2Run* service. In a total volume of 10 μL, 1 μg of the plasmid were mixed with 4 μL of the 5 μM respective sequencing oligonucleotide.

pFB1 fwd Seq 5'- GTTGGCTACGTATACTCCGGA - 3'

pFB1 rev Seq 5' - CACCTCCCCCTGAACCTGAA - 3'

5.2.9 Bacmid generation

DH10Bac or DH10EmBacY *E. coli* were thawed on ice. 20 μL bacteria were mixed with 40 ng plasmid DNA and incubated on ice for 10-30 minutes. Heatshock was performed for 45 seconds at 42 °C followed by 2 minutes incubation on ice. 380 μL LB medium was added and the solution incubated for 4 hours, gently shaking (< 500 rpm) at 37 °C. Depending on the competency of the particular bacterial stock, 5 – 50 μL of the cell suspension were plated on agar plates containing gentamycin (7 μg/mL), kanamycin (50 μg/mL), tetracycline (10 μg/mL), X-Gal (100 μg/mL) and IPTG (40 μg/mL).

A blue-white screening was done during 2 days incubation at 37 °C. Two white colonies were picked, and re-plated with a 2 days incubation time to prevent false positives. Those were used for inoculation over night in LB medium with gentamycin (7 μg/mL), kanamycin (50 μg/mL) and tetracycline (10 μg/mL). Bacmids were isolated and purified (see Methods 5.2.7).

5.2.10 Test PCR

To verify integration of the insert sequence into the bacmid, a test PCR was performed and the result inspected via agarose gel electrophoresis. The PCR was basically run as a colony PCR (Methods 5.2.6), except with different primers:

pUC/M13 fwd 5'-CCCAGTCACGACGTTGTAAAACG-3' and
pUC/M13 rev 5'-AGCGGATAACAATTTCACACAGG-3',

with 1 μL of the isolated bacmid as target.
A positive sample has a size of 2300 bp in addition to the insert size, the negative band is 300 bp long.

5.2.11 Cloning to virus production of *Tb*IMPDH with terminal LBTs

Tb*IMPDH* (GenBank accession number M97794) with an N-terminal His-tag was amplified from an existing cytoplasmic pFB1 *Tb*IMPDH vector via PCR. The PCR product was purified with a GeneJET gel extraction kit, and ligated into the generated and *Ehe*I linearized pFB1 LBT-N and pFB LBT-C vectors. *E.coli* DH5α cells were transformed with the ligated plasmid, following incubation of positive colonies. Bacmids were prepared, Sf9 cells transfected and virus stock was generated over three cell passages. The resulting third passage rBV IMPDH N-LBT stocks and rBV IMPDH C-LBT stocks were used for infection of High Five insect cells at MOI 1, with preceding virus stock titration[318] via $TCID_{50}$ determination.

5.2.12 Biochemical characterization

To test the successful expression of the target genes within the insect cells, SDS polyacrylamide gel electrophoresis (SDS-PAGE) was applied. Further applications were analysis of the overall expression level of the gene of interest, and to detect the protein of interest either in the soluble or unsoluble fraction. The SDS-gels were imaged either after staining with Coomassie or, if an antibody for the specific protein or else an existing HA-tag was available, by Western blotting.

SDS polyacrylamide gel electrophoresis

Identification of the produced protein and evaluation of purity was evaluated by SDS PAGE. The following **Table 10** gives the components for two 6 % stacking gels and two 15 % separation SDS-gels. After assembly of the separation gel, an isopropanol cover prevents irregular drying of the gel during polymerisation.

Polymerized SDS-gels were either used directly for SDS-PAGE or else stored in wetted dust-free paper and aluminium foil at 6 °C for a maximum of 3 days prior to use. Samples were usually denatured by addition of 6x loading dye and heating to 90 C for 10 min.

SDS-PAGE were run at constant current and 25 mA for approximately 1 h 15 min. The running buffer was prepared according to the Hoefer manual for SE250 gel electrophoresis units.

Table 10 Gel casting instruction for two 15 % separating- and two 6 % stacking SDS-gels for SDS-PAGE

	Volumes for *two* 15 % separating gels	Volumes for *two* 6 % stacking gels
dH$_2$O	900 µL	1.35 mL
1.5 M Tris	(pH 8.8) 1.25 mL	(pH 6.8) 625 µL
10 % SDS	50 µL	25 µL
40 % glycerol	225 µL	-
30 % acrylamide/bis	2.5 mL	495 µL
10 % APS	50 µL	25 µL
TEMED	5 µL	2.5 µL
1 % BPB	-	2.5 µL

Terbium-binding monitoring via FRET

Verification of Tb^{3+} binding to isolated IMPDH N-LBT crystals, uninfected High Five cells, High Five cells carrying IMPDH N-LBT and IMPDH N-His crystals was performed using a FP-6200 spectrofluorometer *(Jasco International Company)*. 200 µL of isolated crystals/ cell pelett were mixed with 200 µL of a 5 mM terbium chloride solution in TBS (20 mM Tris, 137 mM NaCl, 2.7 mM KCl, 5 mM TbCl$_3$) and transferred to a quartz cuvette.

In a control experiment, isolated crystals ($78 \cdot 10^4$ crystals per mL) were mixed 1:1 with TBS (20 mM Tris, 137 mM NaCl, 2.7 mM KCl).

Incubation time varied from 10 s to 24 h. Incubation times up to 30 minutes were performed in the quartz cuvette with resuspension before each measurement. Samples were irradiated with a wavelength of 295 nm in the *spectrum measurement* mode. Emission spectra from 530-560 nm were recorded. Additional parameters were a 5 nm excitation bandwith, a 10 nm emission bandwith, a scanning speed of 125 nm/min, low to medium sensitivity.

Determination of protein concentration

Soluble protein concentrations were determined by measuring the absorption in solution using a NanoDrop spectrometer at 280 nm (A280). This method is not very sensitive and relies on the light absorbing features of aromatic amino acids (phenyl alanine, tryptophane and tyrosine) that exhibit an absorption maximum at approximately 280 nm. One µL was applied for each measurement.

The protein concentration was determined with the molar extinction coefficient of the relevant protein, calculated with the amino acid sequence using the program ProtParam of the *ExPASy* Proteomics Server, and the following formula:

$c = \frac{A_{280} MW}{\varepsilon d}$, with *c*, concentration of the protein solution, $A_{280,}$ measured adsorption at 280 nm, *MW*, molecular weight of the protein, ε, molar extinction coeffizient of the protein, *d*, layer thickness of the cuvette.

5.2.13 Insect cell culture

Sf9 and High Five insect cells were held in suspension culture in serum-free ESF921 insect cell culture medium (Expression Systems) at 27 °C on an orbital shaker at 100 rpm. Suspension culture cells were usually seeded at $0.5 – 1 \times 10^6$ cells/mL, in a total volume of 25 mL in an upright standing 75 cm² disposable T-flask except where stated otherwise. Cell density was counted daily in a Neubauer counting chamber and cultures were split when the density reached 4×10^6 cells/mL for High Five or 6×10^6 cells/mL for Sf9 cells, respectively.[181] All steps were performed under sterile conditions under a clean bench.

For upscaling of the insect cell culture, High Five cells and Sf9 cells were adopted to suspension culture in Erlenmeyer flasks without baffles (VWR). These were filled to a maximum of one third with cell culture to ensure sufficient oxygen supply, and were kept rocking at 80 rpm at 27 °C. A supplementation of 7.5 % fetal calve serum (FCS) was applied to increase cell viability.

Preparation of insect cell stocks

Cells of a low passage were grown to 5×10^6 cells/ mL. A reezing chamber (Mr. Freeze, Nalgene) was cooled on ice. Counted cells (Neubauer chamber) were collected in a 50 mL falcon tube and centrifuged for minutes at 700 rcf, 4 °C. All the following steps were done at 4 °C. The supernatant was discarded, and cells resuspended in freezing medium (sterile filtered EFS921 w/o antibiotics, 10 % DMSO and 10 g/L trehalose) to a final cell density of 5×10^7 cells/ mL. Cells were aliquoted at 1 mL to cryo tubes and stored for at least 4 hours in the freezing chamber at -80 °C before transfer to the storage box.

Thawing of insect cells

Insect cells are stored in 1.5 mL cryo reaction tubes in 10 % DMSO at -80 °C.

50 mL ESF921$^{+ \text{P/S}}$ were pre-warmed to 26 °C and distributed evenly to two 75 cm² cell culture flasks. Cells are rapidly thawed in a water bath, then 500 µL cells are immediately given to 25 mL medium. Either cells were directly given to suspension culture and allowed to grow up to 4×10^6 cells/mL at 100 rpm, then split to 1×10^6, before continuing with conventional cell culture. Alternatively, cells were allowed to adhere for 45 minutes, then medium was changed once to remove DMSO. Cells were then grown in adherent cell culture to confluence, with a medium change after two days, then carefully scraped with a cell scraper, then continued in suspension culture.

Recombinant virus production

Bacmid-DNA was used for lipofection. Sf9 insect cells grown in ESF921 serum free insect cell medium at 27 °C were transfected using Escort™ IV transfection reagent (Sigma-Aldrich). For transfection, Sf9 cells were plated out to 70 % confluency and allowed to adhere.

The DNA/liposome complex was prepared: 1 µL of bacmid DNA was diluted in 50 µL antibiotics-free ESF921 medium and 3 µL of Escort™ IV transfection reagent in a final

volume of 50 µL ESF921 medium. Both diluted solutions were gently mixed by resuspension. The DNA/liposome complex was allowed to form for 45 minutes at RT. Adherent cells were washed once with antibiotics-free medium. To a 12-well, 0.4 mL antibiotics-free medium were used to cover the cells.

The DNA/liposome complex was added to the well in a dropwise manner to cover as many cells as possible. Cells were incubated for 15-16 hours. Then, the medium was removed and the cells covered with 1 mL ESF921 + P/S to be incubated for 4 days under standard conditions. The supernatant is the P1 viral stock and was harvested by transfer to an aerosol-tight reaction vessel, 30 centrifugation at 14.000 rcf to remove cell debris and storage in a fresh reaction vessel at 4 °C. For high-titer stock production, 0.9×10^6 Sf9 cells/well in a 6-well plate were infected with 100 µL of P1 viral stock or 20 µL of P2 viral stock and incubated for 4 days and harvested as described above.

Infection of insect cells

In adherent insect cell culture, for crystal generation usually High Five cells were used, cells were plated out to 80 % confluency in ESF921 insect cell medium$^{+P/S}$ and allowed to adhere. Virus stock was usually added to a MOI of 1, if not stated otherwise.

Insect cells in CrystalDirect™ plates were for RT diffraction data collection were infected as follows: 20 min UV sterilization of the CrystalDirect plate, following coating with D-lys, 0.2 mg/ mL. The rBV, cells and ESF921 insect cell medium with antibiotics were pipetted into a reaction vessel and snip-mixed. Subsequently, 125 µL of the liquid were pipetted per 96-well, containing approximately $5.1 \cdot 10^4$ High Five or Sf9 cells.

Suspension cultures in cell flasks were infected in 15 mL ESF921 with P/S with 1×10^6 cells/mL by directly mixing virus stock to the desired MOI while continuosly shaking at 27 °C on an orbital shaker for 3 to 5 days. Suspension cultures in Erlenmeyer flasks without baffles were infected with rBV at an increased cell density of 2×10^6 cells/mL.

All rBV infected cells were routinely inspected for crystal growth under a Leica DM IL LED microscope with 20x and 40x objectives and documented with a Leica MC170 HD microscope camera.

IMPDH *in cellulo* crystal stability tests

IMPDH N-His crystals grown in High Five cells in 500 µL volume were subjected to a range of buffers listed in the following table 4 days *p. i..* Both High Five cells with intracellular IMPDH N-His crystals and free-floating IMPDH N-His crystals in ESF921 insect cell medium were observed by light microscopy using a Leica DM IL LED microscope at 20 x and 30 x obj., during 5 min after addition, and again 1 h and 24 h after buffer addition to the cells.

Table 11 Intracellular and free-floating IMPDH N-His crystals were subjected to buffers and solutions, whose components in water and their pH are listed.

Name of solution	components	pH
NaAc pH 3.5	50 mM sodium acetate	3.5
pH 10 buffer	10 mM KCl 5 mM $MgCl_2$ 100 mM CAPS	10.0
ddH_2O	$VE\text{-}H_2O$	7.5
hypo buffer	10 mM HEPES 10 mM KCl 5 mM $MgCl_2$	7.5
PEG400	hypotonic buffer +30 % (v/v) PEG400	7.5
2- 6 M urea	hypotonic buffer + 2/4/6 M urea	7.5
glycerol	hypotonic buffer + 50 % (v/v) glycerol	7.5
EDTA	100 mM Na phosphate 1.5 M NaCl 10 mM EDTA	8.0
PBS	1 x PBS	7.4
Ammonium bicarbonate	25 mM $(NH_4)HCO_3$	9.0

Fluorescence spectroscopy using High Five and Sf9 insect cells

$1 \cdot 10^6$ cells (either High Five or Sf9) were infected using rBV MOI 2 and left to incubate for four days at 26 °C. After, medium was removed and cells washed with 1 mL TBS. Cells were flushed from the well, either with TBS, or with TBS supplemented with 5 mM $TbCl_3$, then transferred to a 1.5 mL reaction vessel.

For fluorescence measurements performed to detect terbium ion binding to the cells, 100 µL of cells were diluted 1:3 with TBS buffer with or without 5 mM $TbCl_3$. Fluorescence emission was measured at a *Jasco® FP-6200 spectrofluorometer* in a 400 µL quartz cuvette, using parameters as described in the method 'Terbium-binding monitoring via FRET'. Cells were excited at 295 nm. Fluorescence emission was monitored at 530- 560 nm.

CatB SKLG crystal isolation

High Five cells were rBV CatB SKLG infected in adhesion culture as described above and pelleted on day 3 p.i. by centrifugation at 100 rcf for 5 minutes. Cells were resuspended in hypotonic buffer (10 mM HEPES pH 7.5, 10 mM KCl, 5 mM $MgCl_2$) and gently agitated at 37 °C for 1 h, following stepwise centrifugation to separate cell remnants from crystals. In a final step to improve purity, crystals were incubated in RIPA buffer for 10 min, and stored in PBS.

Cell culture upscaling and Nct ECD SKL structure isolation

Insect cells were adapted to suspension culture in Erlenmeyer flasks filled to maximal one third and incubated at continuos shaking (80 rpm) at 27 °C, using a supplementation of 7.5 % fetal calve serum (FCS).[249] Sf9 cells were used for virus stock production.[319]

Since by eye and experience, the crystallization efficiency is higher in suspension culture, High Five cells were infected with rBV nicastrin ECD SKL using a MOI 1 at a concentration of 2×10^6 cells per mL in cell culture disc stacks (Bio One cell discs).

Cells were harvested four days *p.i.* to isolate crystals in an optimized protocol based on cell lysis using detergent (RIPA buffer) and subsequent stepwise centrifugation steps to enrich a crystal pellet. This approach is complicated by a very similar density of *in cellulo* crystals and other cell content.[248] After lysis and buffer change to PBS, residual genomic DNA was removed by DNase I digestion according to manufacturers instructions. Cells were centrifuged in multiple steps at low rcf to harvest crystals in the supernatant, following concentration of the crystal pellet at 10,000 rcf.

Adherent cell culture infection in cell discs and PFA fixation

Medium, virus stock and cells were mixed prior to pouring into the cell disc.

Infected High Five cells were peletted 4 days *p.i.* for 10 min at 300 rcf in falcon tubes, and the supernatant was discarded. Cells were once washed in PBS (pH 6.5), gently spinned down, and the supernatant was again discarded. The pellet was slowly resuspended in a 10 x volume of fixation buffer (0.1 % paraformaldehyde (PFA) in PBS, pH 6.5).

4×10^7 High Five cells resulted in about 1 mL of pellet for which 10 mL fixation buffer were used. Cells were incubated in the fixation buffer at RT for 1.5 hours gently agitated by a gel-shaker. After 10 min centrifugation (300 rcf), the supernatant was discarded, and the pellet was washed by resuspension in the 10x volume of PBS. Fixated cells were pelleted for 10 min at 300 rcf, discarding the supernatant.

IMPDH N-LBT crystal isolation

High Five cell pellets were incubated for 30 min in RIPA buffer, allowing a separation of cellular debris and larger crystals by centrifugation at 100 rcf. The isolated IMPDH N-LBT crystals were washed twice with and then kept in water supplemented with 10 mM $MgCl_2$, followed by 20 minutes incubation in the presence of up to 20 mM Tb^{3+} for LBT loading where they stayed morphologically intact for a few hours.

Propidium iodide staining

High Five, TriEX or Sf9 cells were plated in 96-well plates with 200 μL, 1.2×10^4 cells per well, or in 24-well plates with 500 μL, 5×10^5 insect cells per well. PI (Invitrogen) was added in a 1: 2000 dilution with a final concentration of 500 ng/mL and left to incubate for 5 minutes.

After staining with PI, cells were imaged with fluorescence microscopy using an *Axio Observer Z1*, DsRed channel overlay with bright-field (BF), illuminations ~ 15 ms at BF and 350 ns at DsRed channel with 14 µm offset. Images were inspected and analysed using *AxioVision Rel 4.8.2 SP3* (2013), the overlay was created with *ImageJ*.

For SeMet cytotoxicity tests, 20 minutes after cell adhesion, ESF921 medium was exchanged to ESF921ΔMet medium. Final SeMet concentrations ranging from 50 to 150 mg/L were tested using the same timeline as in **Figure 24**. Insect cells grown without additives in ESF921 undeficient medium served as control. Five days after medium exchange to ESF921ΔMet, a PI stain as described above was performed.

Preparation of SeMet stock

The necessary amount of L-SeMet (Acros Organics) powder was weighted to a 1.5 mL reaction vessel. One drop of 1 M hydrochloric acid was added and filled with dH$_2$O to a 50 mg/ mL SeMet concentration solution. This was further diluted in the cell culture to desired concentrations. In pure water at RT, SeMet is only soluble up to approximately 1×10^{-6} mg/L.

5.2.14 Programs and scripts

Alignment of sequences

- o **BLAST**

  Sequence alignment was performed using the web server of the NCBI (https://blast.ncbi.nlm.nih.gov/Blast).

- o **ClustalW**

  To align more than three DNA or protein sequences, ClustalW[320] was used.

- o **Prot Param**

  Prot Param was useful to display numbering of the sequence.

- o **ESPript 3.0**

  Alignment results of IMPDHs from different organisms were visualized using *ESPript 3.0*. To keep in mind: The file output of ClustalW default is *.CLUSTAL. EXPript 3.0* only accepts files ending with *.ALN*.

CCP4 software suite

- o **Superpose**

  Superposition of X-ray model structures to determine differences

- o **pdbset**

  Renumbering of residues in one job: When using 'increment by', one needs to increment also former deletions/ missing aa, because renumbering will not alter the original file. Configurations in *pdbset* for comparison of IMP deydrogenases related to Tb are shown in **Figure 40**.

Figure 40 Configurations of pdbset, ccp4 suite. An example on how a common nomenclature can be implemented for related sequences to superpose structures and associated aa sequences. Shown is the adaption of aa numbering of the IMPDH N-His sequence from an overlay of the human IMPDH (*PDB 1JCN*) sequence.

Dosage calculation

RADDOSE 3D was used for dose estimation in diffraction experiments.[24,27,33]

Assumed was a Gaussean beam profile with 5 x 10 μm FWHM for diffraction data collected at EMBL beamline P14 at DESY. The chosen computational resolution was 10 pixels per μm.

CrystFEL

The use of tools and scripts belonging to the *CrystFEL* software suite with versions 0.6.4 to 0.9.1 and 0.10.1 were used for this work. The following geometry file (EMBL beamline P14 with the detector EIGER 16M in 4M mode) may be adjusted to any particular experiment.

eiger16m_4m.geom

clen = 0.1857
photon_energy = 12700
adu_per_photon = 1
res = 13333.3 ; 75 micron pixel size

Janine Mia Lahey-Rudolph

; Mask of beamstop_manually edited by using hdfsee *on a virtual powder pattern*

bad_v0/min_fs = 1004

bad_v0/min_ss = 1050

bad_v0/max_fs = 1045

bad_v0/max_ss = 2165

; Masked out strips between panels

;number 1

bad_v0/min_fs = 0

bad_v0/min_ss = 516

bad_v0/max_fs = 2069

bad_v0/max_ss = 547

;number 2

bad_h0/min_fs = 0

bad_h0/min_ss = 1069

bad_h0/max_fs = 2069

bad_h0/max_ss = 1100

;number 3

bad_h2/min_fs = 0

bad_h2/min_ss = 1622

bad_h2/max_fs = 2069

bad_h2/max_ss = 1653

; Bad pixels masked out

mask_file = eiger-badmap.h5

mask = /data/data

mask_good = 0x0

mask_bad = 0x1

panel0/min_fs = 0

panel0/min_ss = 0

panel0/max_fs = 2069

panel0/max_ss = 2166

panel0/corner_x = -1026

panel0/corner_y = -1078.54

panel0/fs = +1.000000x +0.000000y

panel0/ss = +0.000000x +1.000000y

6 Literature

1. Allen, F. H. *et al.* Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds. *J. Chem. Soc. Perkin Trans. 2* S1–S19 (1987). doi:10.1039/P298700000S1

2. Berg, J. M., Tymoczko, J. L., Stryer, L. & Gatto, G. J. *Biochemistry*. (2012).

3. Gossert, A. D. & Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **97**, 82–125 (2016).

4. Luchinat, E. & Banci, L. In-cell NMR: A topical review. *IUCrJ* **4**, 108–118 (2017).

5. Marion, D. An Introduction to Biological NMR Spectroscopy. *Mol. Cell. Proteomics* **12**, 3006–3025 (2013).

6. Billeter, M., Wagner, G. & Wüthrich, K. Solution NMR structure determination of proteins revisited. *J. Biomol. NMR* **42**, 155–158 (2008).

7. Nitta, R., Imasaki, T. & Nitta, E. Recent progress in structural biology: Lessons from our research history. *Microscopy* **67**, 187–195 (2018).

8. Shen, P. S. The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. *Anal. Bioanal. Chem.* **410**, 2053–2057 (2018).

9. Vonck, J. & Mills, D. J. Advances in high-resolution cryo-EM of oligomeric enzymes. *Current Opinion in Structural Biology* (2017). doi:10.1016/j.sbi.2017.05.016

10. Berman, H. M. *et al.* The Protein Data Bank www.rcsb.org. *Nucleic Acids Res.* **28**, 235–242 (2000).

11. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666 (1958).

12. Friedrich, W., Knipping, P. & Laue, M. von. Interferenz-Erscheinungen bei Röntgenstrahlen. **14**, (1912).

13. Bragg, W. L. The diffraction of short electromagnetic waves by a crystal. in *Proceedings of the Cambridge Philosophical Society* **17**, 43–57 (1913).

14. Taylor, G. The phase problem. *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 1881–90 (2003).

15. Burmeister, W. P. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 328–341 (2000).

16. Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci.* **97**, 623–628 (2002).

17. Garman, E. F. & Schneider, T. R. Macromolecular Cryocrystallography. *J. Appl. Crystallogr.* **30**, 211–237 (1997).

18. 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).

19. Garman, E. F., Walsh, S. P. & Bury, C. S. Radiation damage in macromolecular crystallography: the current knowns and unknowns. *Acta Crystallogr. Sect. A Found. Adv.* **73**, C1023–C1023 (2017).

20. Hope, H. Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallogr. Sect. B* **44**, 22–26 (1988).

21. Cowan, J. A. & Nave, C. The optimum conditions to collect X-ray data from very small samples. *J. Synchrotron Radiat.* **15**, 458–462 (2008).

22. Paithankar, K. S., Owen, R. L. & Garman, E. F. Absorbed dose calculations for macromolecular crystals: Improvements to RADDOSE. *J. Synchrotron Radiat.* **16**, 152–162 (2009).

23. Storm, S. L. S. *et al.* Measuring energy-dependent photoelectron escape in microcrystals. *IUCrJ* **7**, 129–135 (2020).

24. Bury, C. S., Brooks-Bartlett, J. C., Walsh, S. P. & Garman, E. F. Estimate your dose: RADDOSE-3D. *Protein Sci.* (2018). doi:10.1002/pro.3302

25. Brooks-Bartlett, J. C. *et al.* Development of tools to automate quantitative analysis of radiation damage in SAXS experiments. *J. Synchrotron Radiat.* **24**, 63–72 (2017).

26. Christensen, J. *et al.* Radiation damage in small-molecule crystallography: Fact not fiction. *IUCrJ* (2019). doi:10.1107/S2052252519006948

27. Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M. & Garman, E. F. Predicting the X-ray lifetime of protein crystals. *Proc. Natl. Acad. Sci.* **110**, 20551–20556 (2013).

28. Zeldin, O. B., Gerstel, M. & Garman, E. F. RADDOSE-3D: Time- and space-resolved modelling of dose in macromolecular crystallography. *J. Appl. Crystallogr.* **46**, 1225–1230 (2013).

29. Arndt, U. W. & Wonacott, A. J. data collection from macromolecular crystals. (1977).

30. Holton, J. M. A beginner's guide to radiation damage. *J. Synchrotron Radiat.* **16**, 133–142 (2009).

31. Gati, C. *et al.* Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* **1**, 87–94 (2014).

32. Wierman, J. L. *et al.* Fixed-target serial oscillation crystallography at room temperature. *IUCrJ* **6**, 305–316 (2019).

33. Zeldin, O. B., Gerstel, M. & Garman, E. F. Optimizing the spatial distribution of dose in X-ray macromolecular crystallography. *J. Synchrotron Radiat.* **20**, 49–57 (2013).

34. Chapman, H. N., et al. Femtosecond {X}-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).

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

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

37. Mariani, V. *et al.* OnDA: Online data analysis and feedback for serial X-ray imaging. *J. Appl. Crystallogr.* **49**, 1073–1080 (2016).

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

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

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

41. Redecke, L. *et al.* Natively Inhibited Trypanosoma brucei Cathepsin B Structure Determined by Using an X-ray Laser. *Science (80-.).* **339**, 227–230 (2013).

42. Rupp, B. *Biomolecular crystallography: Principles, Practice, and Application to Structural Biology.* (Garland Science, 2010).

43. Burkhardt, A. *et al.* Status of the crystallography beamlines at PETRA III. *Eur. Phys. J. Plus* **131**, 0–8 (2016).

44. Zielinski, K. A. *et al.* Rapid and efficient room-temperature serial synchrotron crystallography using the CFEL TapeDrive. *IUCrJ* **9**, 778–791 (2022).

45. Blanchet, C. E. *et al.* Highly automated and time-resolved BioSAXS at the P12 beamline of EMBL Hamburg. **23**, 2015 (2015).

46. Cianci, M. *et al.* P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **24**, 323–332 (2017).

47. Pellegrini, C. X-ray free-electron lasers: From dreams to reality. *Phys. Scr.* **2016**, (2016).

48. Ribic, P. R. & Margaritondo, G. Status and prospects of x-ray free-electron lasers (X-FELs): A simple presentation. *J. Phys. D. Appl. Phys.* **45**, (2012).

49. Spence, J. C. H. XFELs for structure and dynamics in biology. *IUCrJ* **4**, 322–339 (2017).

50. Lee, J., Zatsepin, N. A. & Kim, K. H. Time-resolved serial femtosecond X-ray crystallography. *BioDesign* **6**, 15–22 (2018).

51. Colletier, J. P. *et al.* De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, (2016).

52. Gati, C. *et al.* Atomic structure of granulin determined from native nanocrystalline granulovirus using an X-ray free-electron laser. *Proc. Natl. Acad. Sci.* **114**, 2247–2252 (2017).

53. Liu, W. *et al.* Serial femtosecond crystallography of G protein-coupled receptors. *Science (80-.).* **342**, 1521–1524 (2013).

54. Kupitz, C. *et al.* Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 (2014).

55. Nass, K. Radiation damage in protein crystallography at X-ray free-electron lasers. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 1–8 (2019).

56. Chapman, H. N. ., Caleman, C. & Timneanu, N. Diffraction before destruction. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2011).

57. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).

58. Duyvesteyn, H. M. E. *et al.* Towards in cellulo virus crystallography. *Sci. Rep.* **8**, 3771 (2018).

59. Chapman, H. N. X-Ray Free-Electron Lasers for the Structure and Dynamics of Macromolecules. *Annu. Rev. Biochem.* **88**, annurev-biochem-013118-110744 (2019).

60. Govada, L. & Chayen, N. E. Choosing the Method of Crystallization to Obtain. *Crystals* **9**, 106 (2019).

61. McPherson, A. Protein crystallization. *Methods Mol. Biol.* **1607**, 17–50 (2018).

62. Asherie, N. Protein crystallization and phase diagrams. *Methods* **34**, 266–272 (2004).

63. Charcot, J. M. & Robin, C. Observation de leucocythemie. *CR Mem Soc Biol* **5**, 450–454 (1853).

64. Hartig, T. Über das Klebermehl. *Bot. Zeitung* **13**, 881–882 (1855).

65. Strunk, S. W. The formation of intracellular crystals in midgut glands of Limnoria lignorum. *J. Biophys. Biochem. Cytol.* **5**, 385–92 (1959).

66. Freddo, T. F., Townes-Anderson, E. & Raviola, G. Rod-shaped bodies and crystalloid inclusions in ocular vascular endothelia of adult and developing Macaca mulatta. *Anat. Embryol. (Berl).* **158**, 121–31 (1980).

67. Veenhuis, M., Harder, W., van Dijken, J. P. & Mayer, F. Substructure of crystalline peroxisomes in methanol-grown Hansenula polymorpha: evidence for an in vivo crystal of alcohol oxidase. *Mol. Cell. Biol.* **1**, 949–57 (1981).

68. Hawkes, F. Presence of a crystal in the cytoplasm of the male germ cells of the garden dormouse Eliomys quercinus L. *J. Submicrosc. Cytol. Pathol.* **25**, 407–15 (1993).

69. Vayssié, L., Skouri, F., Sperling, L. & Cohen, J. Molecular genetics of regulated secretion in paramecium. *Biochimie* **82**, 269–88 (2000).

70. Nürnberger, S. *et al.* Giant crystals inside mitochondria of equine chondrocytes. *Histochem. Cell Biol.* **147**, 635–649 (2017).

71. Schönherr, R., Rudolph, J. M. & Redecke, L. Protein crystallization in living cells. *Biol. Chem.* **399**, 1–22 (2018).

72. Ward, R. T. The origin of protein and fatty yolk in Rana pipiens. II. Electron microscopical and cytochemical observations of young and mature oocytes. *J. Cell Biol.* **14**, 309–41 (1962).

73. Banerjee, S. *et al.* Structure of a heterogeneous, glycosylated, lipid-bound, in vivo-grown protein crystal at atomic resolution from the viviparous cockroach Diploptera punctata. *IUCrJ* **3**, 282–293 (2016).

74. Giembycz, M. & Lindsay, M. Pharmacology of the eosinophil. *Pharmacol. Rev.* **51**, 213–340

       (1999).

75. Sawaya, M. R. *et al.* Protein crystal structure obtained at 2.9 A resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl. Acad. Sci.* **111**, (2014).

76. Turkewitz, A. P. Out with a bang! Tetrahymena as a model system to study secretory granule biogenesis. *Traffic* **5**, 63–8 (2004).

77. Arnott, H. J. & Smith, K. M. An ultrastructural study of the development of a granulosis virus in the cells of the moth Plodia interpunctella (Hbn.). *J. Ultrastruct. Res.* **21**, 251–68 (1967).

78. Coulibaly, F. *et al.* The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22205–10 (2009).

79. Veenhuis, M., van Dijken, J. P., Pilon, S. A. & Harder, W. Development of crystalline peroxisomes in methanol-grown cells of the yeast Hansenula polymorpha and its relation to environmental conditions. *Arch. Microbiol.* **117**, 153–63 (1978).

80. Vonck, J. & van Bruggen, E. F. Architecture of peroxisomal alcohol oxidase crystals from the methylotrophic yeast Hansenula polymorpha as deduced by electron microscopy. *J. Bacteriol.* **174**, 5391–9 (1992).

81. Sára, M. & Sleytr, U. B. S-Layer proteins. *J. Bacteriol.* **182**, 859–68 (2000).

82. Heinze, M., Reichelt, R., Kleff, S. & Eising, R. High Resolution Scanning Electron Microscopy of Protein Inclusions (Cores) purified from Peroxisomes of Sunflower (Helianthus annuus L.) Cotyledons. *Cryst. Res. Technol.* **35**, 877–886 (2000).

83. Hammersen, F., Gidlöf, A., Larsson, J. & Lewis, D. H. The occurrence of paracrystalline mitochondrial inclusions in normal human skeletal muscle. *Acta Neuropathol.* **49**, 35–41 (1980).

84. Yuan, P. *et al.* A HEX-1 crystal lattice required for Woronin body function in Neurospora crassa. *Nat. Struct. Biol.* **10**, 264–270 (2003).

85. Papassideri, I. S., Trougakos, I. P., Leonard, K. R. & Margaritis, L. H. Crystalline yolk spheroids in Drosophila melanogaster oocyte: freeze fracture and two-dimensional reconstruction analysis. *J. Insect Physiol.* **53**, 370–6 (2007).

86. Massover, W. H. Intramitochondrial yolk-crystals of frog oocytes. I. Formation of yolk-crystal inclusions by mitochondria during bullfrog oogenesis. *J. Cell Biol.* **48**, 266–79 (1971).

87. Lange, R. H., Grodziński, Z. & Kilarski, W. Yolk-platelet crystals in three ancient bony fishes: Polypterus bichir (Polypteri), Amia calva L., and Lepisosteus osseus (L.) (Holostei). *Cell Tissue Res.* **222**, 159–65 (1982).

88. Snigirevskaya, E. S., Hays, A. R. & Raikhel, A. S. Secretory and internalization pathways of mosquito yolk protein precursors. *Cell Tissue Res.* **290**, 129–42 (1997).

89. Ingram, M. J., Stay, B. & Cain, G. D. Composition of milk from the viviparous cockroach, Diploptera punctata. *Insect Biochem.* **7**, 257–267 (1977).

90. Williford, A., Stay, B. & Bhattacharya, D. Evolution of a novel function: nutritive milk in the viviparous cockroach, Diploptera punctata. *Evol. Dev.* **6**, 67–77 (2004).

91. Tetreau, G. *et al.* De novo determination of mosquitocidal Cry11Aa and Cry11Ba structures from naturally-occurring nanocrystals. *Nat. Commun.* **13**, (2022).

92. Perera, S., Li, Z., Pavlik, L. & Arif, B. *Insect Virology*. (Caister Academic, 2010).

93. Sanahuja, G., Banakar, R., Twyman, R. M., Capell, T. & Christou, P. Bacillus thuringiensis: a century of research, development and commercial applications. *Plant Biotechnol. J.* **9**, 283–300 (2011).

94. Chiu, E. *et al.* Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization. *Proc. Natl. Acad. Sci.* **112**, 3973–3978 (2015).

95. Schnepf, E. *et al.* Bacillus thuringiensis and its pesticidal crystal proteins. *Microbiol. Mol. Biol. Rev.* **62**, 775–806 (1998).

96. Adalat, R., Saleem, F., Crickmore, N., Naz, S. & Shakoori, A. R. In vivo crystallization of three-domain cry toxins. *Toxins (Basel).* **9**, 1–13 (2017).

97. Arvan, P. & Castle, D. Sorting and storage during secretory granule biogenesis: looking backward and looking forward. *Biochem. J.* **332 (Pt 3)**, 593–610 (1998).

98. Dodson, G. & Steiner, D. The role of assembly in insulin's biosynthesis. *Curr. Opin. Struct. Biol.* **8**, 189–94 (1998).

99. Frigas, E. & Gleich, G. J. The eosinophil and the pathophysiology of asthma. *J. Allergy Clin. Immunol.* **77**, 527–37 (1986).

100. Soragni, A. *et al.* Toxicity of eosinophil MBP is repressed by intracellular crystallization and promoted by extracellular aggregation. *Mol. Cell* **57**, 1011–21 (2015).

101. Schönherr, R. *et al.* Real-time investigation of dynamic protein crystallization in living cells. *Struct. Dyn.* **2**, 041712 (2015).

102. Ke, Z. *et al.* Promotion of virus assembly and organization by the measles virus matrix protein. *Nat. Commun.* **9**, 1736 (2018).

103. Smith, K. M. *Virus-Insect Relationships.* (Longman, 1976).

104. Coulibaly, F. *et al.* The molecular organization of cypovirus polyhedra. *Nature* **446**, 97–101 (2007).

105. Rohrmann, G. F. Polyhedrin structure. *J. Gen. Virol.* **67 (Pt 8)**, 1499–513 (1986).

106. Payne, C. C. & Mertens, P. P. C. *The cytoplasmic polyhedrosis viruses.* (Plenum Press, 1983).

107. Ji, X. *et al.* Polyhedra structures and the evolution of the insect viruses. *J. Struct. Biol.* **192**, 88–99 (2015).

108. Ji, X. *et al.* How baculovirus polyhedra fit square pegs into round holes to robustly package viruses. *EMBO J.* **29**, 505–514 (2010).

109. Axford, D., Ji, X., Stuart, D. I. & Sutton, G. In cellulo structure determination of a novel cypovirus polyhedrin. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1435–1441 (2014).

110. Ginn, H. M. *et al.* Structure of CPV17 polyhedrin determined by the improved analysis of serial femtosecond crystallographic data. *Nat. Commun.* **6**, 1–8 (2015).

111. Boudes, M., Garriga, D., Fryga, A., Caradoc-Davies, T. & Coulibaly, F. A pipeline for structure determination of in vivo-grown crystals using in cellulo diffraction. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 576–585 (2016).

112. Boudes, M., Garriga, D. & Coulibaly, F. Microcrystallography of Protein Crystals and In Cellulo Diffraction. *J. Vis. Exp.* 1–9 (2017). doi:10.3791/55793

113. Sleytr, U. B., Schuster, B., Egelseer, E.-M. & Pum, D. S-layers: principles and applications. *FEMS Microbiol. Rev.* **38**, 823–64 (2014).

114. Albers, S.-V. & Meyer, B. H. The archaeal cell envelope. *Nat. Rev. Microbiol.* **9**, 414–26 (2011).

115. Kerfeld, C. A. & Erbilgin, O. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol.* **23**, 22–34 (2015).

116. Plegaria, J. S. & Kerfeld, C. A. Engineering nanoreactors using bacterial microcompartment architectures. *Curr. Opin. Biotechnol.* **51**, 1–7 (2017).

117. Luijsterburg, M. S., Noom, M. C., Wuite, G. J. L. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.* **156**, 262–72 (2006).

118. Karas, V. O., Westerlaken, I. & Meyer, A. S. The DNA-Binding Protein from Starved Cells (Dps) Utilizes Dual Functions To Defend Cells against Multiple Stresses. *J. Bacteriol.* **197**, 3206–15 (2015).

119. Veenhuis, M., Kiel, J. A. K. W. & Van der Klei, I. J. Peroxisome assembly in yeast. *Microsc. Res. Tech.* **61**, 139–150 (2003).

120. Tsukada, H., Mochizuki, Y. & Fujiwara, S. The nucleoids of rat liver cell microbodies. Fine structure and enzymes. *J. Cell Biol.* **28**, 449–60 (1966).

121. Doye, J. P. K. & Poon, W. C. K. Protein crystallization in vivo. *Curr. Opin. Colloid Interface Sci.* **11**, 40–46 (2006).

122. Adalat, R., Saleem, F., Crickmore, N., Naz, S. & Shakoori, A. R. In Vivo Crystallization of Three-Domain Cry Toxins. *Toxins (Basel).* **9**, 80 (2017).

123. Würtz, C., Schliebs, W., Erdmann, R. & Rottensteiner, H. Dynamin-like protein-dependent formation of Woronin bodies in Saccharomyces cerevisiae upon heterologous expression of a single protein. *FEBS J.* **275**, 2932–2941 (2008).

124. Hasegawa, H. *et al.* In vivo crystallization of human IgG in the endoplasmic reticulum of engineered chinese hamster ovary (CHO) cells. *J. Biol. Chem.* **286**, 19917–19931 (2011).

125. Tsutsui, H. *et al.* A diffraction-quality protein crystal processed as an autophagic cargo. *Mol. Cell* **58**, 186–193 (2015).

126. Baskaran, Y. *et al.* An in cellulo-derived structure of PAK4 in complex with its inhibitor Inka1. *Nat. Commun.* **6**, 1–11 (2015).

127. Brandariz-Nuñez, A., Menaya-Vargas, R., Benavente, J. & Martinez-Costas, J. Avian reovirus microNS protein forms homo-oligomeric inclusions in a microtubule-independent fashion, which involves specific regions of its C-terminal domain. *J. Virol.* **84**, 4289–301 (2010).

128. Bonten, E., Van Der Spoel, A., Fornerod, M., Grosveld, G. & D'Azzo, A. Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. *Genes Dev.* **10**, 3156–3169 (1996).

129. Milner, C. M. *et al.* Identification of a Sialidase Encoded in the Human Major Histocompatibility Complex. *J. Biol. Chem.* **272**, 4549–4558 (1997).

130. Hasegawa, H. Simultaneous induction of distinct protein phase separation events in multiple subcellular compartments of a single cell. *Exp. Cell Res.* **379**, 92–109 (2019).

131. Tsukimoto, J. *et al.* In Cellulo Crystallization of Human Neuraminidase 1 and Biological Roles of N-Glycans. *ACS Appl. bio Mater.* (2021).

132. Fan, G. Y. *et al.* In vivo calcineurin crystals formed using the baculovirus expression system. *Microsc. Res. Tech.* **34**, 77–86 (1996).

133. Koopmann, R. *et al.* In vivo protein crystallization opens new routes in structural biology. *Nat. Methods* **9**, 259–262 (2012).

134. Nass, K. *et al.* In cellulo crystallization of Trypanosoma brucei IMP dehydrogenase enables the identification of genuine co-factors. *Nat. Commun.* **11**, 620 (2020).

135. Tiberti, N. & Sanchez, J. C. Sleeping Sickness in the 'Omics Era. *Proteomics - Clinical Applications* **12**, e1700041 (2018).

136. WHO. *Trypanosomiasis, human African (sleeping sickness).* (2019).

137. Shaw, E. Peptidyl sulfonium salts. A new class of protease inhibitors. *J. Biol. Chem.* **263**, 2768–2772 (1988).

138. Scory, S., Stierhof, Y.-D., Caffrey, C. R. & Steverding, D. The cysteine proteinase inhibitor Z-Phe-Ala-CHN2 alters cell morphology and cell division activity of Trypanosoma brucei bloodstream forms in vivo. *Kinetoplastid Biol. Dis.* **6**, 2 (2007).

139. Abdulla, M. H. *et al.* RNA interference of Trypanosoma brucei cathepsin B and L affects disease progression in a mouse model. *PLoS Negl. Trop. Dis.* **2**, 1–6 (2008).

140. Ferreira, L. G. & Andricopulo, A. D. Targeting cysteine proteases in trypanosomatid disease drug discovery. *Pharmacol. Ther.* **180**, 49–61 (2017).

141. Mackey, Z. B., O'Brien, T. C., Greenbaum, D. C., Blank, R. B. & McKerrow, J. H. A cathepsin B-like protease is required for host protein degradation in Trypanosoma brucei. *J. Biol. Chem.* **279**, 48426–48433 (2004).

142. Mort, J. S. & Buttle, D. J. Cathepsin B. *Int. J. Biochem. Cell Biol.* **29**, 715–720 (1997).

143. Quraishi, O. *et al.* The Occluding Loop in Cathepsin B Defines the pH Dependence of Inhibition by Its Propeptide. *Biochemistry* **38**, 5017–5023 (1999).

144. Allison, A. C. & Eugui, E. M. Mycophenolate mofetil and its mechanisms of action. *Immunopharmacology* **47**, 85–118 (2000).

145. Hedstrom, L. IMP dehydrogenase: structure, mechanism, and inhibition. *Chem. Rev.* **109**, 2903–28 (2009).

146. Shu, Q. & Nair, V. Inosine monophosphate dehydrogenase (IMPDH) as a target in drug discovery. *Med. Res. Rev.* **28**, 219–232 (2008).

147. Kabir, M. H. Bin *et al.* Repurposing existing drugs: identification of irreversible IMPDH inhibitors by high-throughput screening. *J. Enzyme Inhib. Med. Chem.* **34**, 171–178 (2018).

148. Koopmann, R. *et al.* In vivo protein crystallization opens new routes in structural biology. *Nat. Methods* **9**, 259–262 (2012).

149. Bessho, T. *et al.* Characterization of the novel Trypanosoma brucei inosine 5′-monophosphate dehydrogenase. *Parasitology* **140**, 735–45 (2013).

150. Alsford, S. *et al.* High-throughput decoding of antitrypanosomal drug efficacy and resistance. *Nature* **482**, 232–6 (2012).

151. Barrett, M. P., Boykin, D. W., Brun, R. & Tidwell, R. R. Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. *Br. J. Pharmacol.* **152**, 1155–71 (2007).

152. Abadeh, A. & Lew, R. R. Mass flow and velocity profiles in Neurospora hyphae: Partial plug flow dominates intra-hyphal transport. *Microbiol. (United Kingdom)* **159**, 2386–2394 (2013).

153. Steinberg, G., Harmer, N. J., Schuster, M. & Kilaru, S. Woronin body-based sealing of septal pores. *Fungal Genet. Biol.* **109**, 53–55 (2017).

154. Bohm, C. *et al.* Current and future implications of basic and translational research on amyloid-β peptide production and removal pathways. *Mol. Cell. Neurosci.* **66**, 3–11 (2015).

155. Xie, T. *et al.* Crystal structure of the γ-secretase component nicastrin. *Proc. Natl. Acad. Sci.* **111**, 13349–13354 (2014).

156. Bai, X. *et al.* An atomic structure of human γ-secretase. **525**, 212–217 (2015).

157. Zhou, R. *et al.* Recognition of the amyloid precursor protein by human γ-secretase. *Science (80-.).* **0930**, (2019).

158. Shah, S. *et al.* Nicastrin functions as a γ-secretase-substrate receptor. *Cell* **122**, 435–447 (2005).

159. Bolduc, D. M., Montagna, D. R., Gu, Y., Selkoe, D. J. & Wolfe, M. S. Nicastrin functions to sterically hinder γ-secretase–substrate interactions driven by substrate transmembrane domain. *Proc. Natl. Acad. Sci.* **113**, E509–E518 (2016).

160. Rohrmann, G. The baculovirus replication cycle : Effects on cells and insects. in *Baculovirus Molecular Biology* 1–32 (National Center for Biotechnology Information, 2013).

161. Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E. & Chisti, Y. Protein production using the baculovirus-insect cell expression system. *Biotechnol. Prog.* **30**, 1–18 (2014).

162. Volkman, L. E. & Summers, M. D. Autographa californica nuclear polyhedrosis virus: Comparative infectivity of the occluded, alkali-liberated, and nonoccluded forms. *J. Invertebr. Pathol.* **30**, 102–103 (1977).

163. Long, G., Pan, X., Kormelink, R. & Vlak, J. M. Functional Entry of Baculovirus into Insect and Mammalian Cells Is Dependent on Clathrin-Mediated Endocytosis. *J. Virol.* **80**, 8830–8833 (2006).

164. Jehle, J. A. *et al.* On the classification and nomenclature of baculoviruses: A proposal for revision. *Arch. Virol.* **151**, 1257–1266 (2006).

165. Skora, L., Shrestha, B. & Gossert, A. D. *Isotope Labeling of Proteins in Insect Cells. Methods in Enzymology* **565**, (Elsevier Inc., 2015).

166. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 325–338

(2010).

167. Murray, K. T. *et al.* Multilayer Laue lenses at high X-ray energies: performance and applications. *Opt. Express* **27**, 7120 (2019).

168. Li, T. *et al.* Dose-efficient scanning Compton X-ray microscopy. *Light Sci. Appl.* **12**, (2023).

169. Hendrickson, W. A. Anomalous diffraction in crystallographic phase evaluation. *Q. Rev. Biophys.* **47**, 49–93 (2014).

170. Bokhoven, C., Schoone, J. C. & Bijvoet, J. M. The Fourier synthesis of the crystal structure of strychnine sulphate pentahydrate. *Acta Crystallogr.* **4**, 275–280 (1951).

171. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

172. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

173. Evans, P. & McCoy, A. An introduction to molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 1–10 (2008).

174. Rossmann, M. G. The molecular replacement method. *Acta Crystallogr. Sect. A* **46**, 73–82 (1990).

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

176. DiMaio, F. *et al.* Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473**, 540–543 (2011).

177. Hendricksont, W. A. *et al.* Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiat(1) Hendricksont, W. A.; Pahlert, A.; Smithtt, J. L.; Satow, Y.; Merrittiitt, E. A.; Phizackerley ', R. P. Biophysics (Oxf). 1989, 86, 2190–2. *Biophysics (Oxf).* **86**, 2190–2194 (1989).

178. Pähler, A., Smith, J. L. & Hendrickson, W. A. A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Crystallogr. Sect. A* **46**, 537–540 (1990).

179. Bearden, J. A. & Burr, A. F. Reevaluation of X-ray atomic energy levels. *Rev. Mod. Phys.* **39**, 125–142 (1967).

180. Chayen, N. E. & Saridakis, E. Protein crystallization: From purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).

181. Schoenherr, R. *et al.* InCellCryst - A streamlined approach to structure elucidation using in cellulo crystallized recombinant proteins. *Nat. Commun.* (2024). doi:https://doi.org/10.1038/s41467-024-45985-7

182. Lahey-Rudolph, J. M. *et al.* Rapid screening of in cellulo grown protein crystals via a small-angle X-ray scattering/X-ray powder diffraction synergistic approach. *J. Appl. Crystallogr.* **53**, 1169–1180 (2020).

183. Lahey-Rudolph, J. M. *et al.* Fixed-target serial femtosecond crystallography using in cellulo grown microcrystals. *IUCrJ* **8**, 665–677 (2021).

184. Tang, Y., Saul, J., Nagaratnam, N. & Garcia, J. M. M. Construction of gateway - compatible baculovirus expression vectors for high - throughput protein expression and in vivo microcrystal screening. *Sci. Rep.* (2020). doi:10.1038/s41598-020-70163-2

185. Leymarie, O. *et al.* Laying the groundwork of an in vivo macromolecular crystallography platform at Synchrotron SOLEIL Laying the groundwork of an in vivo macromolecular crystallography platform at Synchrotron SOLEIL. *J. Phys. Conf. Ser.* **2380**, 012138 (2022).

186. Duden, R. ER-to-Golgi transport: COP I and COP II function (Review). *Mol. Membr. Biol.* **20**, 197–207 (2003).

187. White, T. A. *et al.* Crystallographic data processing for free-electron laser sources. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1231–1240 (2013).

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

189. Gevorkov, Y. *et al.* XGANDALF - Extended gradient descent algorithm for lattice finding. *Acta Crystallogr. Sect. A Found. Adv.* **75**, 694–704 (2019).

190. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).

191. Li, X., Li, C. & Liu, H. ClickX: a visualization-based program for preprocessing of serial crystallography data. *J. Appl. Crystallogr.* **52**, 674–682 (2019).

192. Kabsch, W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.* **21**, 916–924 (1988).

193. Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. Sect. D* **62**, 48–57 (2006).

194. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281 (2011).

195. Melnikov, I., Svensson, O., Bourenkov, G., Leonard, G. & Popov, A. The complex analysis of X-ray mesh scans for macromolecular crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 355–365 (2018).

196. Parkhurst, J. M. *et al.* Background modelling of diffraction data in the presence of ice rings. *IUCrJ* **4**, 626–638 (2017).

197. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).

198. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science (80-.).* **336**, 1030–1033 (2012).

199. Akey, D. L., Brown, W. C., Konwerski, J. R., Ogata, C. M. & Smith, J. L. Use of massively multiple merged data for low-resolution S-SAD phasing and refinement of flavivirus NS1. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 2719–2729 (2014).

200. Karplus, P. A. & Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **34**, 60–68 (2015).

201. Malý, M., Diederichs, K., Dohnálek, J. & Kolenko, P. Paired refinement under the control of {\it PAIREF}. *IUCrJ* **7**, 681–692 (2020).

202. Malý, M., Diederichs, K., Dohnálek, J. & Kolenko, P. PAIREF: paired refinement also for Phenix users. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **77**, 226–229 (2021).

203. Shi, X. & Jarvis, D. Protein N-Glycosylation in the Baculovirus-Insect Cell System. *Curr. Drug Targets* **8**, 1116–1125 (2007).

204. Zamecnik, P. C., Stephenson, M. L., Janeway, C. M. & Randerath, K. Enzymatic synthesis of diadenosine tetraphosphate and diadenosine triphosphate with a purified lysyl-sRNA synthetase. *Biochem. Biophys. Res. Commun.* **24**, 91–97 (1966).

205. Campiglio, M. *et al.* Diadenosines as FHIT-ness instructors. *J. Cell. Physiol.* **208**, 274–281 (2006).

206. Cai, H., Meignin, C. & Imler, J.-L. cGAS-like receptor-mediated immunity: the insect perspective. *Curr. Opin. Immunol.* **74**, 183–189 (2022).

207. Fros, J. J. *et al.* The dinucleotide composition of the Zika virus genome is shaped by conflicting evolutionary pressures in mammalian hosts and mosquito vectors. *PLoS Biol.* **19**, e3001201 (2021).

208. McLennan, A. G. *Diadenosinetetraphosphate (Ap4A). The Dictionary of Genomics, Transcriptomics and Proteomics* (CRC Press, 2015). doi:https://doi.org/10.1201/9781003068402

209. Vartanian, A. A., Suzuki, H. & Poletaev, A. I. The involvement of diadenosine 5′,5‴-P1,P4-tetraphosphate in cell cycle arrest and regulation of apoptosis. *Biochem. Pharmacol.* **65**, 227–235 (2003).

210. Miras-Portugal, M. T., Pintor, J. & Gualix, J. Ca2+ signalling in brain synaptosomes activated by dinucleotides. *J. Membr. Biol.* **194**, 1–10 (2003).

211. Fernández-Justel, D., Peláez, R., Revuelta, J. L. & Buey, R. M. The Bateman domain of IMP dehydrogenase is a binding target for dinucleoside polyphosphates. *J. Biol. Chem.* **294**, 14768–14775 (2019).

212. Tsutsui, H. *et al.* A diffraction-quality protein crystal processed as an autophagic cargo. *Mol. Cell* **58**, 186–93 (2015).

213. Baskaran, Y. *et al.* An in cellulo-derived structure of PAK4 in complex with its inhibitor Inka1. *Nat. Commun.* **6**, 8681 (2015).

214. Sawaya, M. R. *et al.* Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12769–74 (2014).

215. Blessing, R. H. Outlier Treatment in Data Merging. *J. Appl. Crystallogr.* **30**, 421–426 (1997).

216. Moreau, D. W., Atakisi, H. & Thorne, R. E. Ice in biomolecular cryocrystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 540–554 (2021).

217. Gevorkov, Y. *et al.* FDIP—A Fast Diffraction Image Processing Library for X-ray Crystallography Experiments. *Crystals* **14**, 1–16 (2024).

218. Gualtieri, E. J. *et al.* Detection of membrane protein two-dimensional crystals in living cells. *Biophys. J.* **100**, 207–214 (2011).

219. Casadei, C. M. *et al.* Resolution extension by image summing in serial femtosecond crystallography of two-dimensional membrane-protein crystals. *IUCrJ* **5**, 103–117 (2018).

220. Schopf, B., Howaldt, M. W. & Bailey, J. E. DNA distribution and respiratory activity of Spodoptera frugiperda populations infected with wild-type and recombinant Autographa californica nuclear polyhedrosis virus. *J. Biotechnol.* **15**, 169–185 (1990).

221. Puente-Massaguer, E., Saccardo, P., Ferrer-Miralles, N., Lecina, M. & Gòdia, F. Coupling Microscopy and Flow Cytometry for a Comprehensive Characterization of Nanoparticle Production in Insect Cells. *Cytom. Part A* **97**, 921–932 (2020).

222. Davidovich, M., Gougoutas, J. Z., Scaringe, R. P., Vitez, I. & Yin, S. Detection of polymorphism by powder X-ray diffraction: Interference by preferred orientation. *Am. Pharm. Rev.* **7**, 10–17 (2004).

223. Fili, S. *et al.* Human insulin polymorphism upon ligand binding and pH variation: the case of 4-ethylresorcinol. *IUCrJ* **2**, 534–544 (2015).

224. Karavassilia, F. & Margiolaki, I. Macromolecular Powder Diffraction: Ready for genuine biological problems. *Protein Pept. Lett.* (2016). doi:10.2174/0929866523666160120152839

225. Degen, T., Sadki, M., Bron, E., König, U. & Nénert, G. The HighScore suite. *Powder Diffr.* **29**, S13–S18 (2014).

226. Jenner, M. J. *et al.* Successful protein cryocooling for powder diffraction research papers Successful protein cryocooling for powder diffraction. 121–124 (2007). doi:10.1107/S0021889806044943

227. Kardoost, A. *et al.* Convolutional neural network approach for the automated identification of in cellulo crystals. *bioRxiv* 2023.03.28.533948 (2023). doi:10.1101/2023.03.28.533948

228. Watier, Y. Powder diffraction studies of proteins. (Université Grenoble Alpes, 2011).

229. Von Dreele, R. B. Multipattern Rietveld refinement of protein powder data: An approach to higher resolution. *J. Appl. Crystallogr.* **40**, 133–143 (2007).

230. Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. The first protein crystal structure determined from high-resolution X-ray powder diffraction data: A variant of T3R3 human insulin-zinc complex produced by grinding. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 1549–1553 (2000).

231. Margiolaki, I. *International Tables for Crystallography: Volume H: Powder Diffraction. International Tables for Crystallography Volume H* (2019).

232. Boudes, M., Garriga, D., Coulibaly, F. Microcrystallography of Protein Crystals and In

Cellulo Diffraction. *J. Vis. Exp.* **21**, e55793 (2017).

233. Norton-Baker, B. *et al.* A simple vapor-diffusion method enables protein crystallization inside the HARE serial crystallography chip. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 820–834 (2021).

234. Grünbein, M. L. & Nass Kovacs, G. Sample delivery for serial crystallography at free-electron lasers and synchrotrons. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 1–14 (2018).

235. Schlichting, I. Serial femtosecond crystallography: the first five years. *IUCrJ* **2**, 246–55 (2015).

236. Uervirojnangkoorn, M. *et al.* Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *Elife* **2015**, 1–29 (2015).

237. Batyuk, A. *et al.* Native phasing of x-ray free-electron laser data for a G protein–coupled receptor. *Sci. Adv.* **2**, e1600292 (2016).

238. Zander, U. *et al.* Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 454–466 (2016).

239. Roedig, P. *et al.* A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. *Sci. Rep.* **5**, 1–11 (2015).

240. Lieske, J. *et al.* On-chip crystallization for serial crystallography experiments and on-chip ligand-binding studies. *IUCrJ* **6**, 714–728 (2019).

241. Feiler, C. G., Wallacher, D. & Weiss, M. S. An all-in-one sample holder for macromolecular x-ray crystallography with minimal background scattering. *J. Vis. Exp.* (2019). doi:10.3791/59722

242. Roedig, P. *et al.* High-speed fixed-target serial virus crystallography. *Nat. Methods* **14**, 805–810 (2017).

243. Koiwai, K. *et al.* Improvement of Production and Isolation of Human Neuraminidase-1 in Cellulo Crystals. *ACS Appl. Bio Mater.* **2**, 4941–4952 (2019).

244. Martiel, I., Müller-Werkmeister, H. M. & Cohen, A. E. Strategies for sample delivery for femtosecond crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 160–177 (2019).

245. Zhao, F. Z. *et al.* A guide to sample delivery systems for serial crystallography. *FEBS J.* **286**, 4402–4417 (2019).

246. Ibrahim, M. *et al.* Untangling the sequence of events during the S2 → S3 transition in photosystem II and implications for the water oxidation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12624–12635 (2020).

247. Beyerlein, K. R. *et al.* Mix-and-diffuse serial synchrotron crystallography. *IUCrJ* **4**, 769–777 (2017).

248. Schneegans, S. Investigaton of the crystallization of proteins within living insect cells. (University of Lübeck, 2015).

249. Riekehr, W. M. Methodenetablierung für die Isolation und direkte Phasierung von in cellulo -Kristallen am Beispiel von humanem Nicastrin Establishment of methods for isolation and direct phasing of in cellulo crystals using human nicastrin Inhalt. (2017).

250. DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D. Appl. Phys.* **41**, (2008).

251. Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep.* **7**, 1–10 (2017).

252. Wiedorn, M. O. *et al.* Megahertz serial crystallography. *Nat. Commun.* **9**, (2018).

253. Allahgholi, A. *et al.* {AGIPD}, a high dynamic range fast detector for the European {XFEL}. *J. Instrum.* **10**, C01023--C01023 (2015).

254. Lomb, L. *et al.* An anti-settling sample delivery instrument for serial femtosecond crystallography. *J. Appl. Crystallogr.* **45**, 674–678 (2012).

255. Cheng, R. *et al.* Influence of Fixation and Permeabilization on the Mass Density of Single Cells: A Surface Plasmon Resonance Imaging Study. *Front. Chem.* **7**, (2019).

256. Yamamoto, M. *et al.* Protein microcrystallography using synchrotron radiation. *IUCrJ* **4**, 529–539 (2017).

257. de la Mora, E. *et al.* Radiation damage and dose limits in serial synchrotron crystallography at cryo- And room temperatures. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4142–4151 (2020).

258. Garman, E. F. Radiation damage in macromolecular crystallography: What is it and why should we care? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 339–351 (2010).

259. Boutet, S., Fromme, P. & Editors, M. S. H. *X-ray Free Electron Lasers*. (Springer Nature Switzerland AG, 2018). doi:10.1007/978-3-030-00551-1

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

261. Kim, S.-O., Kim, J., Okajima, T. & Cho, N.-J. Mechanical properties of paraformaldehyde-treated individual cells investigated by atomic force microscopy and scanning ion conductance microscopy. *Nano Converg.* **4**, 5 (2017).

262. Strobl, F., Duerkop, M., Palmberger, D. & Striedner, G. High shear resistance of insect cells: the basis for substantial improvements in cell culture process design. *Sci. Rep.* **11**, (2021).

263. Mueller, C. *et al.* Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography. *Struct. Dyn.* **2**, (2015).

264. Roedig, P. *et al.* High-speed fixed-target serial virus crystallography. *Nat. Methods* **14**, 805–810 (2017).

265. Lee, D. *et al.* Nylon mesh-based sample holder for fixed-target serial femtosecond crystallography. *Sci. Rep.* **9**, (2019).

266. Oghbaey, S. *et al.* Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography. *Acta Crystallogr. Sect. D* **72**, 944–955 (2016).

267. Opara, N. *et al.* Direct protein crystallization on ultrathin membranes for diffraction measurements at X-ray free-electron lasers. *J. Appl. Crystallogr.* **50**, 909–918 (2017).

268. Roedig. A New Fixed-Target Approach for Serial Crystallography at Synchrotron Light Sources and X-ray Free Electron Lasers. (2017).

269. Nogly, P. *et al.* Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography. *Nat. Commun.* **7**, (2016).

270. Ritter, C. Design of Lanthanide Binding Nanobodies for Structural Research on Membrane Proteins. (Universität zu Lübeck, 2017).

271. Klapproth, C. Tb3+-Markierung von in cellulo kristallierten Proteinen. (Universität zu Lübeck, 2017).

272. Henkel, A. Evaluierung der Tb3+-Markierung von in cellulo-Kristallen zur Lösung des kristallografischen Phasenproblems. (Universität zu Lübeck, 2018).

273. Brittain, H. G., Richardson, F. S. & Martin, R. B. Terbium(III) emission as a probe of calcium(II) binding sites in proteins. *J. Am. Chem. Soc.* **98**, 8255–8260 (1976).

274. Allen, J. E. & McLendon, G. L. Tryptophan and tyrosine to terbium fluorescence resonance energy transfer as a method to "map" aromatic residues and monitor docking. *Biochem. Biophys. Res. Commun.* **349**, 1264–1268 (2006).

275. Seffernick Hargittai, M. R. & Musier-Forsyth, K. Use of terbium as a probe of tRNA tertiary structure and folding. *RNA* **6**, 1672–1680 (2000).

276. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).

277. Cronin, C. N., Lim, K. B. & Rogers, J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci.* **16**, 2023–2029 (2007).

278. Wenzel, S., Imasaki, T. & Takagi, Y. A practical method for efficient and optimal

production of Seleno-methionine-labeled recombinant protein complexes in the insect cells. *Protein Sci.* **28**, 808–822 (2019).

279. Imasaki, T., Wenzel, S., Yamada, K., Bryant, M. L. & Takagi, Y. Titer estimation for quality control (TEQC) method: A practical approach for optimal production of protein complexes using the baculovirus expression vector system. *PLoS One* **13**, 1–21 (2018).

280. Ceballos, R. F. Untersuchung der Kristallisation der Ektodom¨ane des Glykoproteins gp130 in lebenden Insektenzellen. (2018).

281. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. Sect. A* **32**, 751–767 (1976).

282. Ledent, P. *et al.* Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. *FEBS Lett.* **413**, 194–196 (1997).

283. Zhao, D. & Huang, Z. Effect of His-Tag on Expression, Purification, and Structure of Zinc Finger Protein, ZNF191(243-368). *Bioinorg. Chem. Appl.* **2016**, 8206854 (2016).

284. Topal, M. D. & Fresco, J. R. Fluorescence of terbium ion-nucleic acid complexes: a sensitive specific probe for unpaired residues in nucleic acids. *Biochemistry* **19**, 5531–5537 (1980).

285. Ghisaidoobe, A. B. T. & Chung, S. J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).

286. Iram, S. *et al.* Biogenic terbium oxide nanoparticles as the vanguard against osteosarcoma. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **168**, 123–131 (2016).

287. Strub, M. P. *et al.* Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* (2003). doi:10.1016/j.str.2003.09.014

288. Hendrickson, W. A. & Ogata, C. M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494–523 (1997).

289. Barton, W. A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P. & Nikolov, D. B. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci.* **15**, 2008–2013 (2006).

290. Seitomer, E., Balar, B., He, D., Copeland, P. R. & Kinzy, T. G. Analysis of Saccharomyces cerevisiae null allele strains identifies a larger role for DNA damage versus oxidative stress pathways in growth inhibition by selenium. *Mol. Nutr. Food Res.* **52**, 1305–1315 (2008).

291. G Malkowski, M. *et al. Blocking S-adenosylmethionine synthesis in yeast allows selenomethionine incorporation and multiwavelength anomalous dispersion phasing. Proceedings of the National Academy of Sciences of the United States of America* **104**, (2007).

292. Bockhorn, J. *et al.* Genome-wide screen of Saccharomyces cerevisiae null allele strains identifies genes involved in selenomethionine resistance. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17682—17687 (2008).

293. Plateau, P. *et al.* Exposure to selenomethionine causes selenocysteine misincorporation and protein aggregation in Saccharomyces cerevisiae. *Sci. Rep.* **7**, 1–12 (2017).

294. Braunagel, S. C., Parr, R., Belyavskyi, M. & Summers, M. D. Autographa californica nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. *Virology* **244**, 195–211 (1998).

295. Lipson, H., Beevers, C. A. & Bragg, W. L. The crystal structure of the alums. *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **148**, 664–680 (1935).

296. Foos, N., Cianci, M. & Nanao, M. H. Choosing your (Friedel) mates wisely: Grouping data sets to improve anomalous signal. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 200–210 (2019).

297. Gromer, S. *et al.* Active sites of thioredoxin reductases: Why selenoproteins? *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12618–12623 (2003).

298. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

299. Nettleship, J. E., Assenberg, R., Diprose, J. M., Rahman-Huq, N. & Owens, R. J. Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J. Struct. Biol.* **172**, 55–65 (2010).

300. McWhirter, S. M. *et al.* Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8408–8413 (1999).

301. Liemann, S., Chandran, K., Baker, T. S., Nibert, M. L. & Harrison, S. C. Structure of the reovirus membrane-penetration protein, Mu1, in a complex with is protector protein, Sigma3. *Cell* **108**, 283–295 (2002).

302. B'Hymer, C. & Caruso, J. A. Selenium speciation analysis using inductively coupled plasma-mass spectrometry. *J. Chromatogr. A* **1114**, 1–20 (2006).

303. Wilschefski, S. C. & Baxter, M. R. Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects. *Clin. Biochem. Rev.* **40**, 115–133 (2019).

304. Hirschberg, K. *et al.* Kinetic analysis of secretory protein traffic and characterization of Golgi to plasma membrane transport intermediates in living cells. *J. Cell Biol.* **143**, 1485–1503 (1998).

305. Tsutsui, H. *et al.* A Diffraction-Quality Protein Crystal Processed as an Autophagic Cargo. *Mol. Cell* **58**, 186–193 (2015).

306. Schönherr, R. Entwicklung einer Pipeline für die Kristallisation von Proteinen in lebenden Insektenzellen. (Universität zu Lübeck, 2021).

307. Buey, R. M. *et al.* A nucleotide-controlled conformational switch modulates the activity of eukaryotic IMP dehydrogenases. *Sci. Rep.* **7**, 2648 (2017).

308. Jakobi, A. J. *et al.* In cellulo serial crystallography of alcohol oxidase crystals inside yeast cells. *IUCrJ* **3**, 88–95 (2016).

309. Sommer, B. P. Optimierung der in vivo Kristallisationsbedingungen heterolog exprimierter Proteine in Sf9 Insektenzellen und Charakterisierung der Kristalle. (2016).

310. Weinert, T. *et al.* Proton uptake mechanism in bacteriorhodopsin captured by serial synchrotron crystallography. *Science (80-.).* (2019). doi:10.1126/science.aaw8634

311. Nass, K. *et al.* Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. *Nat. Commun.* **11**, 1–9 (2020).

312. Botha, S. *et al.* Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr. Sect. D Biol. Crystallogr.* (2015). doi:10.1107/S1399004714026327

313. Owen, R. L. *et al.* Exploiting fast detectors to enter a new dimension in room-temperature crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**, 1248–1256 (2014).

314. Norton-Baker, B. *et al.* A simple vapor-diffusion method enables protein crystallization inside the HARE serial crystallography chip. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 820–834 (2021).

315. Terwilliger, T. C. *et al.* Accelerating crystal structure determination with iterative {\it AlphaFold} prediction. *Acta Crystallogr. Sect. D* **79**, 234–244 (2023).

316. Bergfors, T. & Majumdar, S. Screening cells for crystals: A synergistic approach Bergfors Terese. *J. Appl. Crystallogr.* **53**, 1414–1415 (2020).

317. Gallat, F.-X. *et al.* In vivo crystallography at X-ray free-electron lasers: the next generation of structural biology? *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130497–20130497 (2014).

318. Reed, L. J. & Muench, H. The American. *Am. J. Hyg. VOL.* **27**, 493–497 (1938).

319. Wilde, M., Klausberger, M., Palmberger, D., Ernst, W. & Grabherr, R. Tnao38, high five and Sf9-evaluation of host-virus interactions in three different insect cell lines: Baculovirus production and recombinant protein expression. *Biotechnol. Lett.* **36**, 743–749 (2014).

320. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019 F abio. *Nucleic Acids Res.* **47**, 636–641 (2019).

321. Gasteiger, E. *et al. The Proteomics Protocols Handbook - Chapter 52: Protein Identification and*

Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook (2005). doi:10.1385/1592598900

A Appendix

A.1 List of used substances according to GHS

GHS risk symbols

| GHS02 | GHS03 | GHS05 | GHS06 | GHS07 | GHS08 | GHS09 |

Hazard and precautionary statements

GHS hazard statements

Physical hazards

H225 **Highly flammable liquid or vapour**

H272 **May intensify fire; oxidizer**

Health hazards

H301 **Toxic if swallowed**

H302 **Harmful if swallowed**

H314 **Causes sever skin burns and eye damage**

H315 **Causes skin irritation**

H317 **May cause an allergic skin reaction**

H319 **Causes serious eye irritation**

H331+ **Toxic/ harmful if inhaled**
H332

H334 **May cause allergy or asthma symptoms or breathing difficulties if inhaled**

H335 **May cause respiratory irritation**

H340 **May cause genetic defects**

H350 **May cause cancer**

H360+ **May/suspected to damage fertility or the unborn child**
H361

H372, **Causes/ may cause damage to organs through prolonged or repeated exposure**
H373

Environmental hazards

H400 **Very toxic to aquatic life**

H410 **Very toxic/ toxic/ harmful to aquatic life with long-lastic effects**
- H412

GHS precautionary statements

P201 **Obtain special instructions before use**

P210 **Keep away from heat, hot surfaces, sparks, open flames and other ignition sources. No smoking.**

P240 **Ground/bond container and receiving equipment**

P261 **Avoid breathing dust/fume/gas/mist/vapours/spray**

P280 **Wear protective gloves/protective clothing/eye protection/face protection.**

P301 **If swallowed…**

P302 **If on skin…**

P304 **If inhaled…**

P304 **If in eyes…**

P310 **…immediately call a POISON CENTER/doctor**

P312 **Call a POISON CENTER/ doctor/…/if you feel unwell.**

P313 **Get medical advice**

P330 **Rinse mouth**

P333 **If skin irritation or a rash occurs:**

P338 **Remove contact lenses if present and easy to do. Continue rinsing.**

P340 **Bring person to fresh air and keep comfortable for breathing**

P342 **If experiencing respiratory symptoms:**

P351 **Rinse cautiously with water for several minutes**

P352 **Wash with plenty of water/…**

P361 **Take off immediately all contaminated clothing**

P391 **Collect spillage**

P403 **Store in a well ventilated place. Keep cool.**
+ P235

List of chemicals and GHS classification

Supplemental Table 12 Chemicals used in this work, ordered alphabetically.

Chemical	GHS hazard	Hazard statement	Precautionary Statements
Acrylamid/ bis-acrylamid	GHS06, GHS08	301, 332, 312, 315, 319, 317, 340, 350, 361, 372	201, 280, 301+310+330, 302+352+312, 304+340+312, 305+351+338
Ampicillin	GHS08	317-334	261-280-342-311
Ammoniumbicarbonate	GHS07	302	301+312, 330
Ammoniumpersulfate (APS)	GHS03, GHS07, GHS08	272, 302, 315, 317, 319, 334, 335,	210, 280, 301+312+330, 302+352, 305+351+338
Bis Tris	GHS07	315-319, 335	261, 305+351+338
CaCl$_2$	GHS07	319	305+351+338
1,4-Dithiothreitol (DTT)	GHSO7	302, 315, 319, 335	261, 302+352, 305+351+338, 501
EDTA	GHS07	319	305+338+351
Ethanol	GHS02, GHS07	225, 319	210, 240, 305+351+338, 403+235
Gentamycin	GHS08	317-334	261-280, 313
Glutaraldehyde	GHS05, GHS07- GHS09	302, 332, 314, 318, 334, 317, 335, 400, 411	261, 273, 280, 303+361+353, 304+340+310, 305+351+338+310, 342+311, 391
Kanamycin	GHS08	360	201, 308 + 313
Midori Green	-	-	261, 280
Isopropanol	GHS02, GHS07	225, 319, 336	210, 233- 305, 351, 338
Phosphoric acid	GHS05	314, 290	280, 301+330+331, 305+351+338, 309+310
SDS	GHS02, GHS05, GHS07	228, 302+332, 315, 318, 335, 412	210, 261, 273, 280, 305+351+338
Selenomethionine (SeMet)	GHS06, GHS08, GHS09	301-331-373-410	261, 273, 301+310, 311, 501
N,N,N′,N′-Tetramethylethylendiamin (TEMED)	GHS02, GHS05, GHS07	225-332-302-314	210, 233, 280, 301+330+331, 305+351+338, 309+310
Tris - HCl, Tris-Base	GHS07	315, 319, 335	261, 305+351+338

A.2   *Nc*HEX-1 and *Tb*IMPDH computational characterization

Sequences of *Nc*HEX-1 and *Tb*IMPDH were characterized using *ProtParam* of the *ExPASy* server.[321]

Computational characterization of *Nc*HEX-1

The protein *Nc*HEX-1 is 175 amino acids long, not counting the start methionine (<mark style="background-color: yellow">M</mark>). Green-highlighted glycine and alanine residues are artefacts from the blunt-end cloning strategy of the cytoplasmic construct HEX-1 cyto.

```
          10         20         30         40         50
60
MGYYDDDAHGH VEADAAPRAT TGTGTGSASQ TVTIPCHHIR LGDILILQGR PCQVIRISTS

          70         80         90        100        110
120
   AATGQHRYLG VDLFTKQLHE ESSFVSNPAP SVVVQTMLGP VFKQYRVLDM
QDGSIVAMTE

         130        140        150        160        170
   TGDVKQNLPV IDQSSLWNRL QKAFESGRGS VRVLVVSDHG REMAVDMKVV HGSRLA
```

The molecular weight (MW) is approximately 19 kDa (18995.47 Da), the theoretical isoelectric point equals pH 6.46. Extinction coefficients were calculated as described in (Gasteiger et al., 2005) with the following equation, and assumed to be reasonably accurate due to a tryptophane in the sequence.[321] Protein concentrations were estimated by application of the Lambert-Beer-law:

$$Extinction\_c_{prot} = \#_{Tyr} \cdot ext^{Tyr} + \#_{Trp} \cdot ext^{Trp} + \#_{Cystine} \cdot ext^{Cystine},$$

with $\#$ as number, (ext^{Tyr}) = 1 490; (ext^{Trp}) = 5 500; $(ext^{Cystine})$ = 125. Assuming that all pairs of cysteine residues form cysteines, the cytoplasmic HEX-1 cyto extinction coefficient equals Ext._coefficient (HEX-1 cyto, oxi) = 11585 M^{-1} cm^{-1}, at 280 nm measured in water.

<u>HEX-1 cyto, only cystines</u>: Abs 0.1% = 0.610 g/l.

Assuming that all disulphide bonds are reduced, so that no cysteines appear as half-cystines:
Ext. coefficient (HEX-1 cyto, red) = 11460 M^{-1} cm^{-1}.
<u>HEX-1 cyto, only cysteines:</u> Abs 0.1% = 0.603 g/l.

The protein has a negative net-charge, as there are 18 negatively charged (asparagine + glutamine) residues, but only 16 positively charged (arginine and lysine) ones.

Computational characterization of *Tb*IMPDH N-His

The protein *Tb*IMPDH is 540 aa long, not counting the start M. The first 28 N-terminal residues comprise a His-tag and a TEV-cutting site which were integrated by Michael Duszenko (University of Tübingen), and which are not cleaved from the protein due to the nature of the applied intracellular crystallization method. The pFastBac1 vector map including the *Tb*IMPDH N-His insert can be found in **Supplemental Figure 2**. The C-terminus constitutes of serine, lysine and leucine, which enables posttranslational transport into peroxisomes.

```
         10         20         30         40         50         60
MSYYHHHHHH DYDIPTTENL YFQGAMGSME NTNLRTKTLR DGTTAEELFS QDGLSFNDFI

         70         80         90        100        110        120
ILPGFIDFDS SKVNVSGQFT KNILLHLPLV SSPMDTVTES SMARAMALMG GIGVIHNNCT

        130        140        150        160        170        180
VEQQARMVRS VKLYRNGFIM KPKSVSPDVP VSTIRNIKSE KGISGILVTE GGKYDGKLLG

        190        200        210        220        230        240
IVCTKDIDFV KDASAPVSQY MTRRENMTVE RYPIKLEEAM DVLNRSRHGY LPVLNDKDEV

        250        260        270        280        290        300
VCLCSRRDAV RARDYPNSSL DRNGHLLCAA ATSTREADKG RVAALSEAGI DVLVLDSSQG

        310        320        330        340        350        360
NTIYQVSFIR WVKKTYPHLE VVAGNVVTQD QAKNLIDAGA DSLRIGMGSG SICITQEVLA

        370        380        390        400        410        420
CGRPQATAIY KVARYAASRG VPCVADGGLR NVGDVCKALA VGANVAMLGS MIAGTSETPG

        430        440        450        460        470        480
EYFFKDGMRL KGYRGMGSID AMLQGRESGK RYLSENETLQ VAQGVAGAVL DKGSVLKLLA

        490        500        510        520        530        540
YIHKGLQQSA QDIGEVSFDA IREKVYEGQV LFNRRSLTAQ SEGAVHSLHH YERKLFASKL
```

The MW of *Tb*IMPDH N-His is 59,066.36 Da and the theoretical pI equals pH 8.23.

Assuming that all pairs of cysteine residues form cysteines, the *Tb*IMPDH N-His extinction coefficient equals at 280 nm 35800 M^{-1} cm^{-1}, measured in water.
TbIMPDH N-His, only cystines: Abs 0.1% = 0.606 g/l.

If all disulphide bonds are reduced the ext. coefficient equals 35300 M^{-1} cm^{-1}.
TbIMPDH N-His, only cysteines: Abs 0.1% = 0.598 g/l.

The protein has a positive net-charge, with 59 negatively charged (asparagine + glutamine) residues, and 62 positively charged (arginine and lysine) residues.

A.3 Vector maps

Vector maps were created with *SnapGene.*
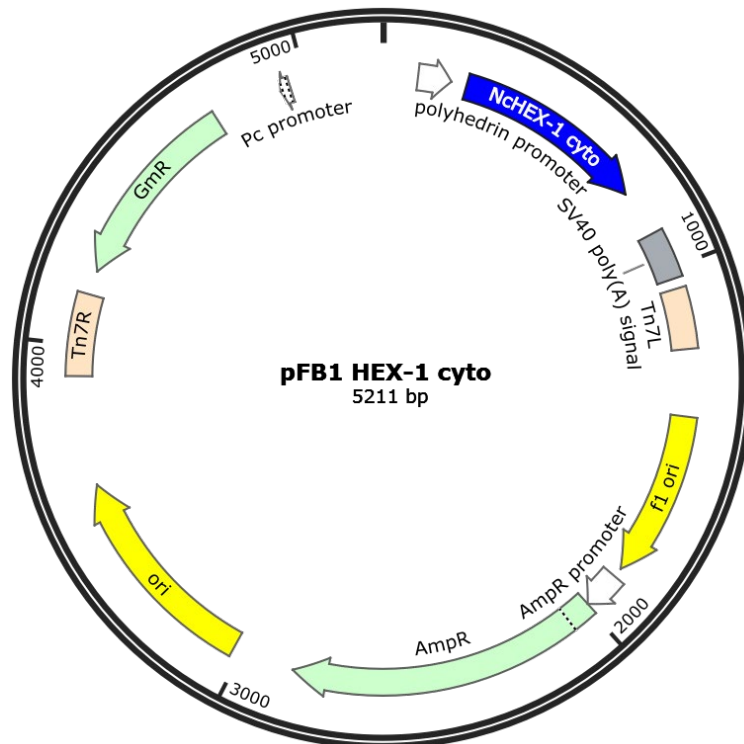
Supplemental Figure 1 A vector map of the pFastBac 1 vector with the cytoplasmic *Nc*HEX-1 insert after the polyhedrin promoter. pFB1, modified in terms of a deleted stop codon. Insert depicted in blue. Ampicellin resistency in ORF f1.
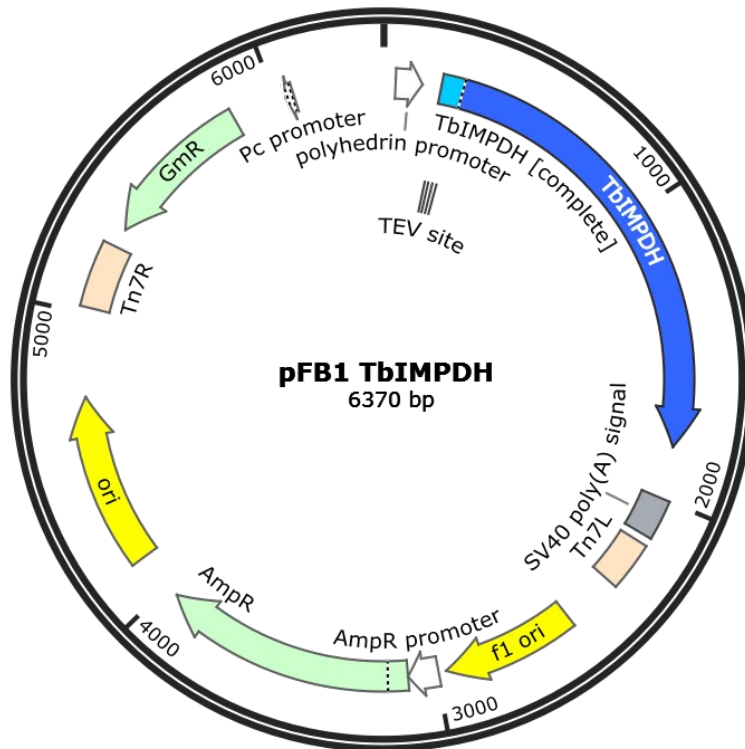
Supplemental Figure 2 Vector map of pFB1 *Tb*IMPDH with N-terminal His-tag.
Modification is a deleted stop codon, the insert the 6370 bp long IMPDH N-His, in blue.
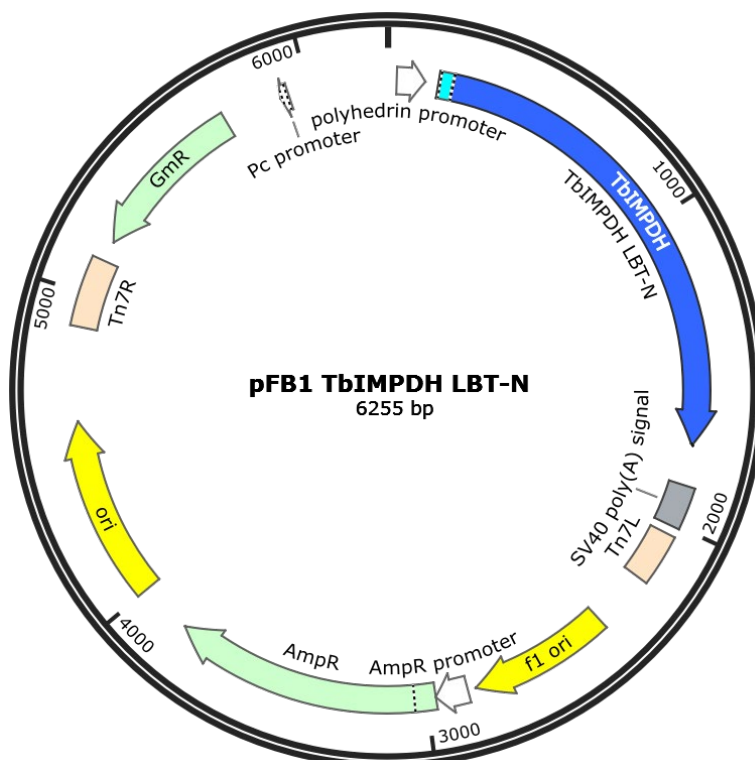
Supplemental Figure 3 Vector map of pFB1 with the *Tb*IMPDH insert sequence, including an N-terminal lanthanide binding tag.
The tag consisting of the 13 residues DTNNDGWIEGDEL was cloned using the forward primer 5'-GACACCAACAACGACGGTTGGATCGAGGGTGACGAGCTGGGC-3'.
The sequence was kindly provided by the group of Christian Löw (EMBL)
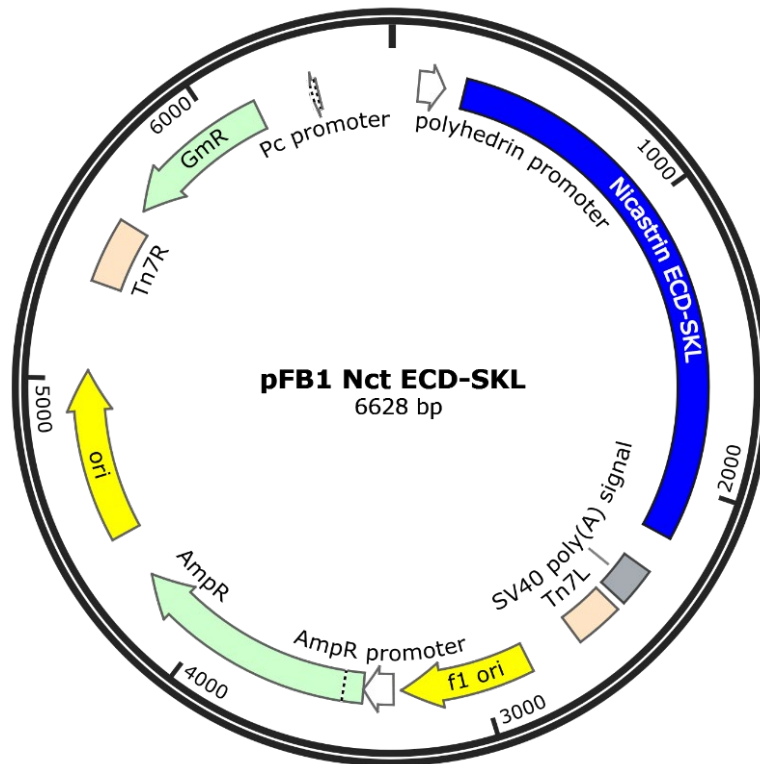
Supplemental Figure 4 Vector map of pFB1 with the human nicastrin ectodomain insert.
Includes a peroxisomal transport signal consisting of serine, lysine and leucine at the C-terminus.
The Nct ectodomain (ECD) sequence:
"SVERKIYIPLNKTAPCVRLLNATHQIGCQSSISGDTGVIHVVEKEEDLQWVLTDGPNPPYMVLLESKHFT RDLMEKLKGRTSRIAGLAVSLTKPSPASGFSPSVQCPNDGFGVYSNSYGPEFAHCREIQWNSLGNGLAYE DFSFPIFLLEDENETKVIKQCYQDHNLSQNGSAPTFPLCAMQLFSHMHAVISTATCMRRSSIQSTFSINPEI VCDPLSDYNVWSMLKPINTTGTLKPDDRVVVAATRLDSRSFFWNVAPGAESAVASFVTQLAAAEALQK APDVTTLPRNVMFVFFQGETFDYIGSSRMVYDMEKGKFPVQLENVDSFVELGQVALRTSLELWMHTDP VSQKNESVRNQVEDLLATLEKSGAGVPAVILRRPNQSQPLPPSSLQRFLRARNISGVVLADHSGAFHNK YYQSIYDTAENINVSYPEWLSPEEDLNFVTDTAKALADVATVLGRALYELAGGTNFSDTVQADPQTVTR LLYGFLIKANNSWFQSILRQDLRSYLGDGPLQHYIAVSSPTNTTYVVQYALANLTGTVVNLTREQCQDPS KVPSENKDLYEYSWVQGPLHSNETDRLPRCVRSTARLARALSPAFELSQWSSTEYSTWTESRWKDIRARI FLIASKELE" has a cloning-strategy derived N-terminal glycine and C-terminal adenine residue after/ before start and stop codon.
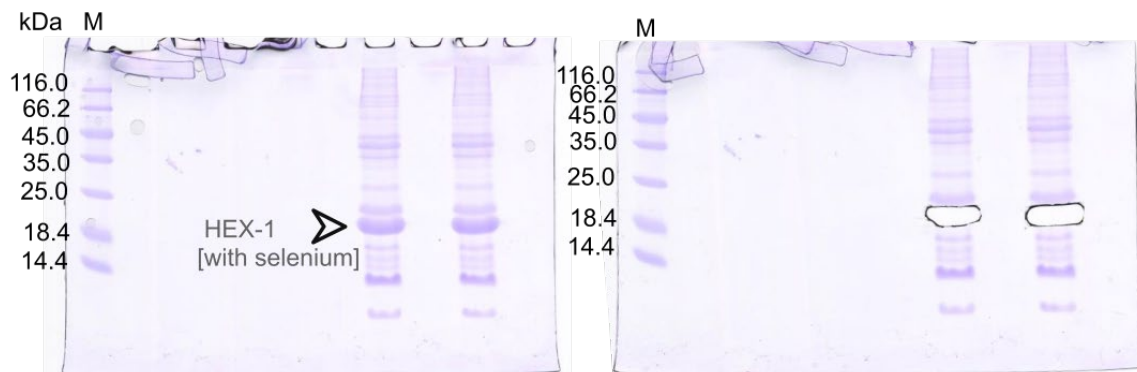
Supplemental Figure 5 Coomassie stain of SDS gel with resolubilized HEX-1 cyto crytals. Crystals were isolated from rBV HEX-1 (MOI 5) infected High Five cells four days *p.i.*. The gel in the right panel is missing the lane with cut-out protein for MS experiments.

A.4   MS spectra of HEX-1 cyto with SeMet incorporation after high-titer rBV infection

A   HEX-1 peptide (108-126) VLDMQDGSIVAMTETGDVK

B   + 1 oxidation

C   +1 Se

D   +1 Se + 1 oxidation

E   +2 Se
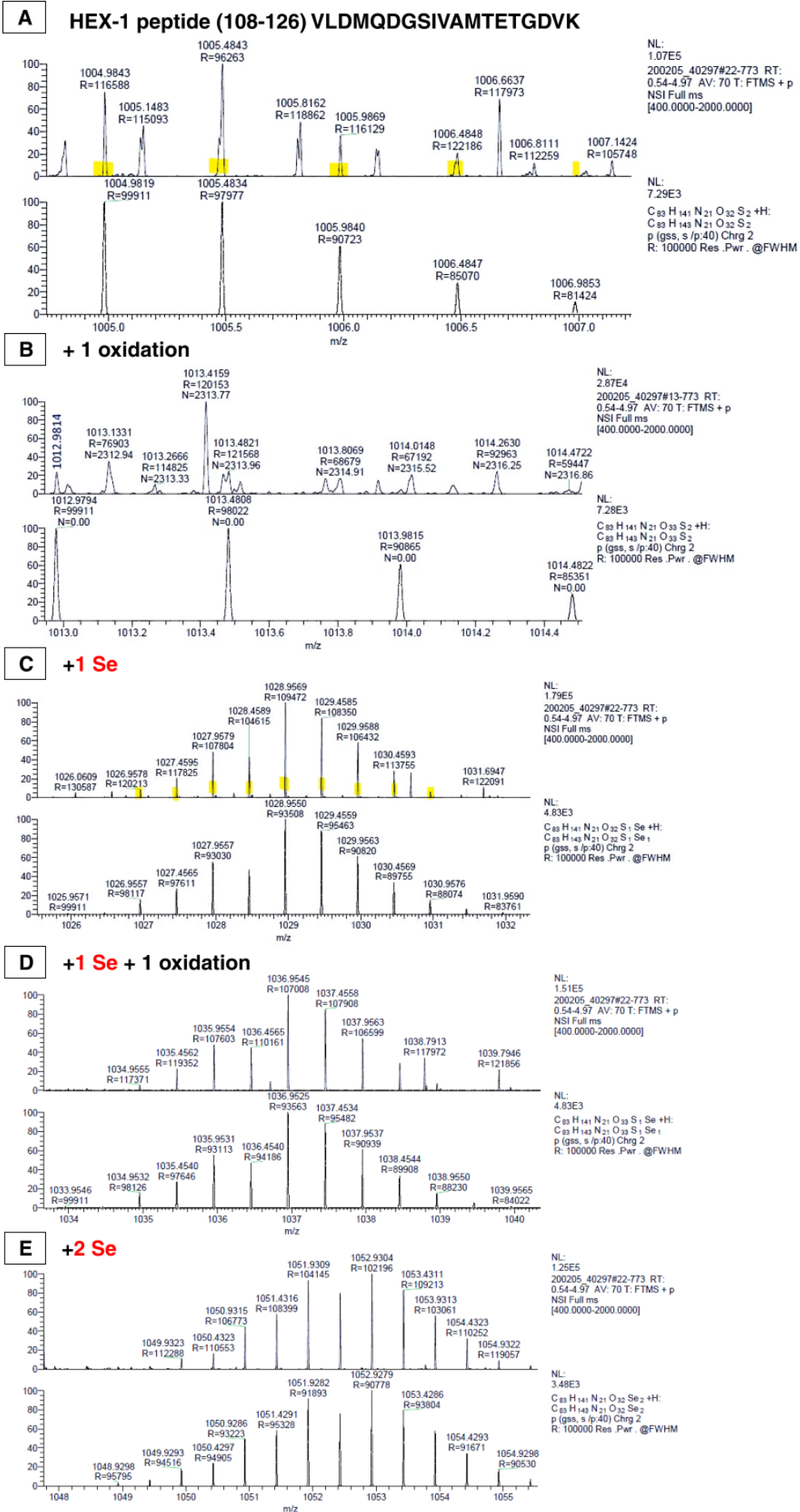
F **HEX-1 peptide (108-126) VLDMQDGSIVAMTETGDVK +2 Se +1 oxidation**
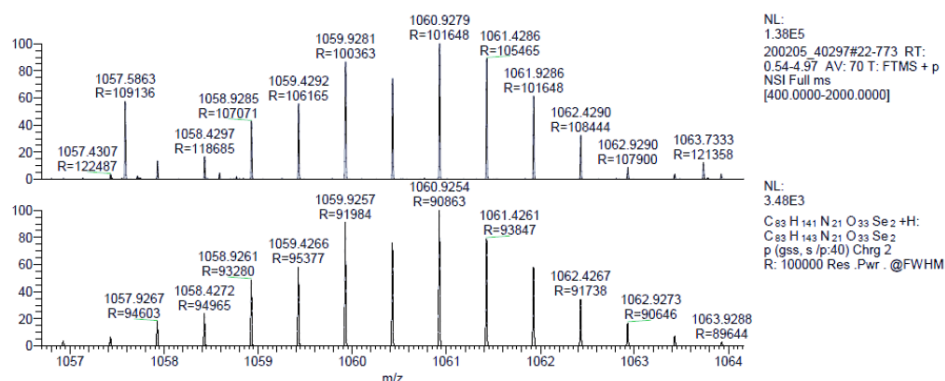
G **+2 Se + 2 oxidation**
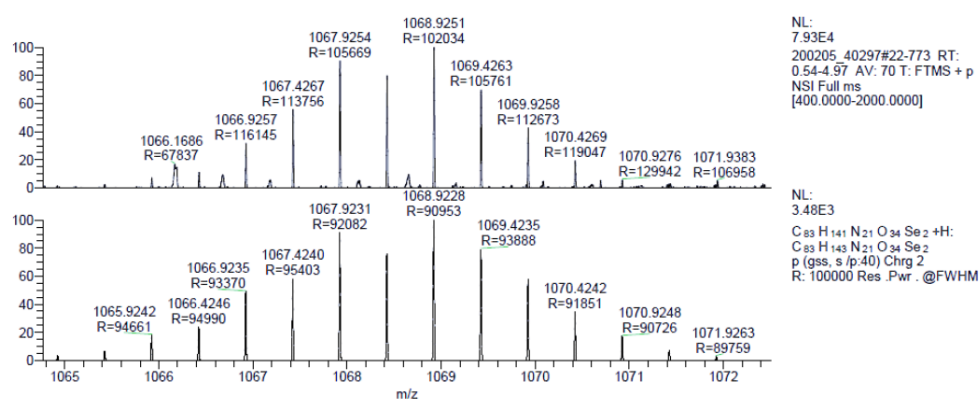
Supplemental Figure 6 MS¹ spectra of VLDMQDGSIVAMTETGDVK, a peptide of the protein *Nc*HEX-1, resolubilized from intracellular protein crystals grown in rBV infected High Five cells (MOI 5), in ESF921ΔMet medium supplemented with 90 mg/L SeMet.

The upper mass to charge (m/z) spectrum is experimentally determined, the one below is calculated, both are normalized. **A** Only the mono-isotopic peaks, highlighted in yellow, were considered for quantification of the methionine content. **B** The peptide VLDMQDGSIVAMTETGDVK with one oxidation. The abundance of this peptide was too low to be feasible for quantification of the methionine content. **C** The peptide with one selenium **D** The peptide with one selenium and one oxidation **E** MS spectrum of the peptide with two selenium incorporations. **F** Spectrum of the peptide with two selenium incorporations and one oxidation. **G** MS¹ of the peptide with two selenium incorporations and two oxidations.
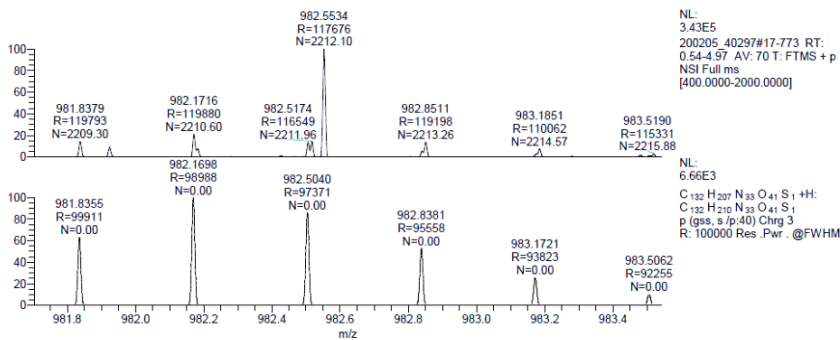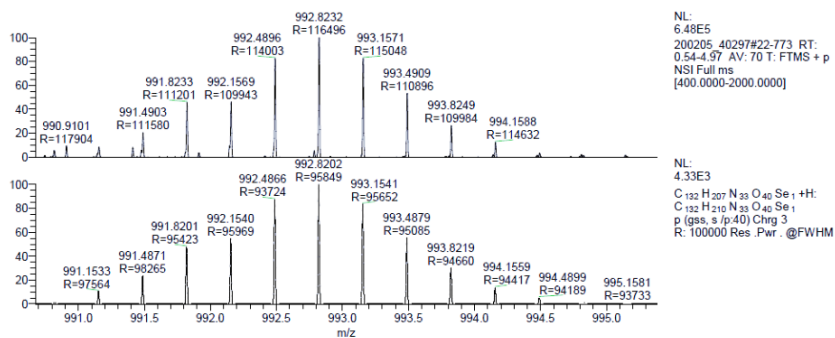
Supplemental Figure 7 MS[1] spectra of the peptide QLHEESSFVSNPAPSVVVQTMLGPVFK from *Nc*HEX-1, resolubilized from intracellular protein crystals grown in rBV infected High Five cells (MOI 5) in ESF921ΔMet medium supplemented with 90 mg/L SeMet.

The upper m/z spectrum is experimentally determined, the one below is calculated. Both are normalized. **A** Peptide without modification.; **B** + one oxidation; **C** + one incorporated selenium methionine; **D** + one selenium and one oxidation.

A   **HEX-1 peptide (163-169) EMAVDMK**

NL:
5.30E4
200205_40297#13-773 RT:
0.54-4.97 AV: 70 T: FTMS + p
NSI Full ms
[400.0000-2000.0000]

823.3706
R=132799
N=1665.49

823.3988
R=127938
N=1665.56

823.3150
R=120771
N=1665.36

823.3688
R=124888
N=0.00

823.7170
R=139200
N=1666.33

823.9939
R=139089
N=1667.00

824.3742
R=133004
N=1667.91

824.5753
R=131938
N=1668.40

825.0808
R=136676
N=1669.61

825.5557
R=136672
N=1670.76

NL:
1.39E4

EMAVDMK +H +H_2O:
$C_{33} H_{59} N_8 O_{12} S_2$
p (gss, s /p:40) Chrg 1
R: 125000 Res .Pwr . @FWHM

824.3716
R=117461
N=0.00

825.3648
R=92590
N=0.00

B   **+ 1 Se**

NL:
1.99E5
200205_40297#22-773 RT:
0.54-4.97 AV: 70 T: FTMS +
p NSI Full ms
[400.0000-2000.0000]

867.8326
R=127334

871.3150
R=125805

869.3160
R=130811

867.3180
R=128889

867.9443
R=133982

870.5435
R=129084

872.3186
R=127504

NL:
7.81E3

$C_{33} H_{58} N_8 O_{12} S_1 Se_1$ +H:
$C_{33} H_{59} N_8 O_{12} S_1 Se_1$
p (gss, s /p:40) Chrg 1
R: 17000 Res .Pwr . @FWHM

871.3136
R=16935

869.3149
R=16892

867.3161
R=16967

868.3174
R=16945

870.3173
R=16867

872.3162
R=16904

873.3146
R=16766

874.3166
R=16787

875.3165
R=16548

876.3175
R=16512

C   **+1 Se. Charged twice.**

200205_40297 #22-773 RT: 0.54-4.97 AV: 70 NL: 3.75E5
T: FTMS + p NSI Full ms [400.0000-2000.00...]

x20

435.7752
R=180653

436.3100
R=174103

434.8837
R=172511

435.1618

434.6640

436.2769
R=171560

436.4009
R=173474

436.8116
R=171805

436.6630

434.3852
R=170922

435.8775
R=176367

435.6638

436.1616
R=169596

436.8908
R=160773

437.2885
R=178027

437.9405
R=176027

C33H58N8O12S1Se1 +H: C33 H60 N8 O12 S1 Se1 p(gss, s/...

436.1604

435.1611

436.6618

437.1609

434.1617

434.6623

435.6623

437.6619

D **+1 Se + 1 oxidation. Charged twice.**

E **+2 Se. Charged twice.**
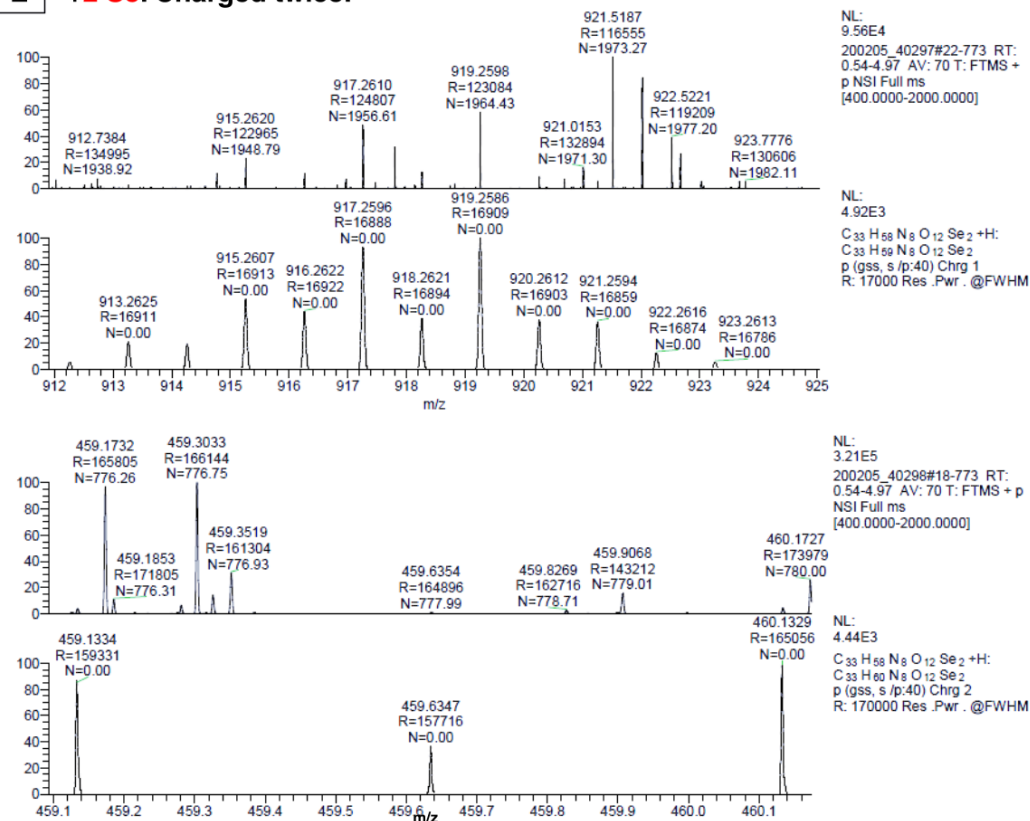
F   HEX-1 peptide (163-169) EMAVDMK +2 Se +1 oxidation. Twice charged.
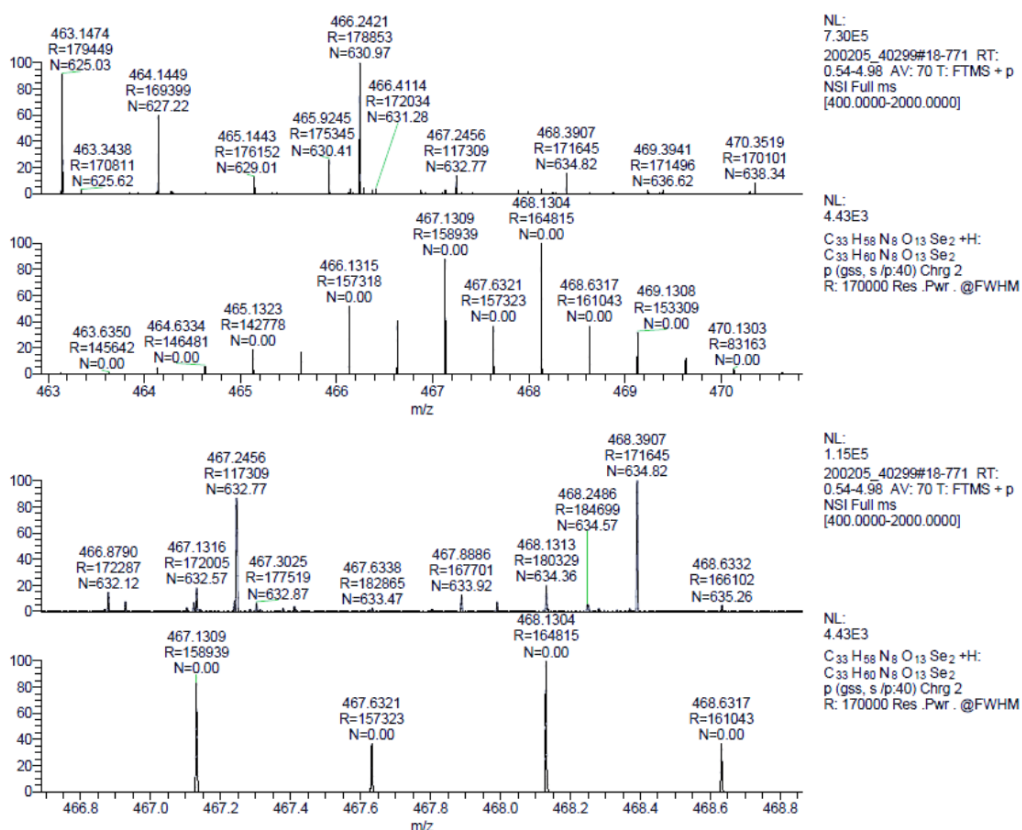
G   +2 Se +2 oxidation. Low abundance.
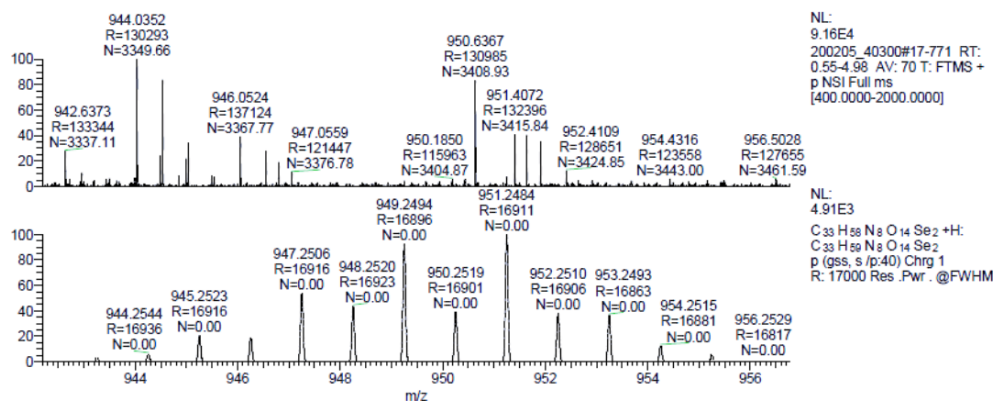
Supplemental Figure 8 MS¹ spectra of EMAVDMK, a peptide of the protein *Nc*HEX-1, resolubilized from intracellular protein crystals grown in rBV infected High Five cells (MOI 5) in ESF921ΔMet medium and supplemented with 90 mg/L SeMet.
The upper m/z spectra are experimentally determined, the ones below are calculated. Both are normalized. **A** Peptide without modification; **B** + one incorporated selenium; **C** + one incorporated selenium, charged twice; **D** + one selenium and one oxidation, twice charged. **E** + two seleniums, charged twice; **F** + two seleniums and one oxidation, charged twice; **G** + two seleniums and two oxidations. This peptide was not suitable for quantification because of low abundance.

List of figures

List of Supplemental Figures

List of tables

Acronyms

***Ac*MNPV** *Autographa californica* multicapsid nucleopolyhedrovirus.

AGIPD adaptive gain integrating pixel detector.

AMP adenosine monophosphate.

At *Arabidopsis thaliana***.**

APP amyloid precursor protein.

AP6A P1-(5'-adenosyl) P6-(5'-adenosyl) hexaphosphate.

ATP adenosine triphosphate.

BEVS baculovirus expression vector system.

Bs *Bacillus subtilis.*

bR bacteriorhodopsin.

BV budded virus.

CatB Cathepsin B of *Trypanosoma brucei.*

CLC Charcot-Leyden crystal.

CPV Cytoplasmic polyhedrosis virus.

CXI coherent x-ray imaging instrument.

cyto cytoplasmic pFB1 vector for rBV infection with N-terminal Gly and C-terminal Ala cloning-dericed artificial residues, without targeting sequence to a cellular compartment.

DXS 1-Deoxy-d-xylulose 5-phosphate synthase.

ECD ectodomain.

EM electron microscopy.

eMOI estimated effectivity of infection.

EuXFEL EuropeanXFEL.

ESI electrospray ionisation.

FACS fluorescence associated cell sorting.

GFP green fluorescent protein of the jellyfish *Aequorea victoria.*

GMP guanosine monophosphate.

GTP guanosine triphosphate.

HAT African trypanosomiasis.

***Nc*HEX-1** HEX-1 protein of *Neurospora crassa.*

IMPDH Inosine 5′-monophosphate dehydrogenase.

IspD 2-C-methyl-ᴅ-erythritol-4-phosphate cytidyltransferase.

IspE 4-diphosphocytidyl-2C-methyl-d-erythritol kinase.

IspH *lytB* gene transcript of *Plasmodium falciparum*.

FCS fetal calve serum.

FEL free-electron laser.

LB liquid broth.

LCLS Linac Coherent Light source.

LCP lipidic cubic phase.

LM light microscopy.

LS Lumazinsynthase.

MAD multi-wavelength anomalous dispersion.

MIR multiple isomorphous replacement.

MR molecular replacement.

MS mass spectrometry.

Nct human nicastrin.

NMR nuclear magnetic resonance.

NPV nuclear polyhedrosis virus.

ori blunt-end cloning dedicated pFB1 vector for rBV infection without any artificial residues.

PBS phosphate-buffered saline.

pc principal component.

PCR polymerase chain reaction.

PDB protein data bank.

PEG400 polyethylene glycol with an average mass of 400 g/mol.

PETRA III Positron-Electron Tandem Ring Accelerator 3rd generation syncrotron source.

Pf = *Plasmodium falciparum*.

PFA paraformaldehyde.

pFB1 pFastBac1 vector.

PI propidium iodide.

PIC protease inhibitor complex.

p.i. post infection.

+P/S including penicillin/ streptomycin antibiotics.

PyrA pyrimidindesaminase.

RT room temperature.

SAD single anomalous diffraction or dispersion.

SASE self-amplified spontaneous emission.

SAXS small-angle x-ray scattering.

SeMet selenomethionine.

SIR single isomorphous replacement.

SKL serine lysine leucine; if C-terminal, these residues target the protein to peroxisomes.

SFX serial femtosecond crystallography.

Sf9 *Spodoptera frugiperda.*

SNR signal-to-noise ratio.

SPB/SFX Single Particles, Clusters, and Biomolecules/
Serial Femtosecond Crystallography instrument.

TAE tris-acetate-EDTA buffer.

Tb Trypanosoma brucei, pathogenic agent of sleeping sickness.

TBS tris-buffered saline.

TCEP tris (2-carboxyethyl) phosphin.

TEQC titer estimation of quality control.

TOF time-of-flight.

TRX time-resolved X-ray crystallography.

WAXS wide-angle x-ray scattering.

XDS X-ray Detector Software.

XFEL X-ray free-electron laser.

XMP xanthosine 5′-monophosphate.

XPA crystallizable and photo-activatable, a genetically engineered coral fluorescent protein forming *in cellulo* crystals in mammalian cells.

Declaration of oath

I hereby declare on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Ich bin mit einer Einstellung meiner Abschlussarbeit in den Bestand der Bibliothek des Fachbereiches Chemie der Universität Hamburg einverstanden.

Hamburg, den 26.02.2024

[Janine Mia Lahey-Rudolph]

Janine Mia Lahey-Rudolph
Acknowledgement

I am grateful to everyone who has supported me and helped me grow with my PhD years. My supervisors, especially Lars, thank you for tutoring me in this fascinating research area! Markus, and Henry, thank you for your continuos support. Thank you to my referees and disputation commission for your valuable work at the end of my PhD time.

Most of all, I am grateful to my family. Because you strengthened my back, made me lough, believed in me, all the while helping me see that life is not only consisting of work; this is what enabled me to be who I am. Mama, Papa: thank you for everything!
Judith, I wish so much we could have celebrated the finishline of this journey. I know, you would have jumped through the phone and organized the biggest defense party ever. We all miss you. I miss you.

Theo, your laughter is the nicest think to hear after work. Zoë, you have been looking forward to tinkering my doctor hat for years. Now, I am curious about the result.
And my Thomas, your support and also your pushes mean the world to me.

Now let's have this new nameplate on our door!

Tom White, thank you for continuous CrystFEL advice and many enlightening discussions in and outside work, it's always a pleasure. Valerius Mariani, for getting me started with On- and Off-da and for valuing my superpower of finding errors in the code. Dominik Oberthür, for good discussions and advice whenever necessary during my time at CFEL.
Gleb Bourenkov, and Johanna Hakannpää, Cy Jeffries, you helped me through uncountable beamtimes at P14 and P12, thanks to you I know how a beamline scientist should be like! Dominik Schwudke and Verena Scholz at the Leibniz research center Borstel, thank you for the fruitful cooperation with mass spectrometry experiments.

To the team of the PIER Helmholtz Graduate School and the Joachim Herz Foundation, for the excellent Joachim Herz PhD scholarship and support. Not only financially, also through soft-skill courses, travel grants and the chance to organize my own conference together with the other JHS scholarship recipients.