

ARTICLE

<https://doi.org/10.1038/s41467-019-12955-3>

OPEN

Membrane protein megahertz crystallography at the European XFEL

Chris Gisriel  et al.[#]

The world's first superconducting megahertz repetition rate hard X-ray free-electron laser (XFEL), the European XFEL, began operation in 2017, featuring a unique pulse train structure with 886 ns between pulses. With its rapid pulse rate, the European XFEL may alleviate some of the increasing demand for XFEL beamtime, particularly for membrane protein serial femtosecond crystallography (SFX), leveraging orders-of-magnitude faster data collection. Here, we report the first membrane protein megahertz SFX experiment, where we determined a 2.9 Å-resolution SFX structure of the large membrane protein complex, Photosystem I, a >1 MDa complex containing 36 protein subunits and 381 cofactors. We address challenges to megahertz SFX for membrane protein complexes, including growth of large quantities of crystals and the large molecular and unit cell size that influence data collection and analysis. The results imply that megahertz crystallography could have an important impact on structure determination of large protein complexes with XFELs.

*email: pfromme@asu.edu; n.zatsepin@latrobe.au; [#]A full list of authors and their affiliations appears at the end of the paper.

The advent of hard X-ray free-electron lasers (XFELs) and emergence of serial femtosecond crystallography (SFX)^{1,2} has enhanced X-ray protein crystallography. In SFX, the extreme XFEL brilliance and ultrashort (fs scale) pulse duration enable atomic-resolution structure determination of proteins at room temperature (RT) from micron-sized protein crystals, while outrunning structure-altering (secondary) radiation damage.

Since the initial proof-of-principle serial crystallography experiments^{1,2} performed in 2009 at the first hard XFEL, the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory, XFEL technologies have been further developed that include novel techniques to improve the growth of nano-/microcrystals, deliver samples, and analyze SFX data. The implementation of time-resolved SFX^{3–9} garnered interest from the structural biology community and increased the demand for XFEL beamtime. To make XFEL technology available for these experiments, strategies to increase the data collection speed are needed. A prime approach to achieving this is the use of high X-ray pulse repetition rates, which have been realized at the European XFEL (EuXFEL).

Since the opening of LCLS, four more XFELs have begun user operation: SPring-8 Angstrom Compact Free-Electron Laser (SACLA) in Japan opened in 2011, Pohang Accelerator Laboratory X-ray Free-Electron Laser (PAL-XFEL) in South Korea opened in 2016, the EuXFEL in Germany opened in 2017, and SwissFEL in Switzerland began user operation recently. At PAL-XFEL, SACLA, SwissFEL, and LCLS, the data acquisition rates, limited by the XFEL pulse rate, are 30–120 Hz (Supplementary Table 1).

Powered by a superconducting linear accelerator, the EuXFEL is the first MHz repetition rate hard XFEL, with a bunch structure designed to deliver up to 27,000 pulses per second. Typically, XFELs produce uniformly spaced X-ray pulses at a constant repetition rate. However, the EuXFEL delivers XFEL pulses in 10-Hz pulse trains with up to 2700 pulses per train planned. User-assisted commissioning started in September 2017. In its first run, the EuXFEL delivered pulse trains at 10 Hz where each train contained up to 60 pulses, separated by 886 ns (~1.128 MHz). In run 2, the rate was increased to 120 pulses per train, thus delivering 1200 pulses per second, and correspondingly, at least a tenfold increase in the data collection rate compared with other hard XFELs. The term MHz crystallography was first introduced in the reporting of SFX at the EuXFEL with concanavalin A, concanavalin B, and lysozyme by Grünbein et al.¹⁰ and on lysozyme and β -lactamase by Wiedorn et al.¹¹, where MHz refers to the repetition rate within a pulse train.

MHz pulse rates bring new challenges. In SFX, the sample is delivered in a serial way to the beam, commonly in a jet of crystals in their mother liquor². It must be rapidly replenished between the XFEL pulses, removing the sample destroyed by the previous pulse. Time separation between pulses within a train at the EuXFEL is so brief that the sample must be replenished at least 9000 times faster than required at the LCLS where pulses are separated by 8.3 ms. The first two publications reporting SFX experiments at the EuXFEL demonstrated that the sample can be effectively replenished in 886 ns between the X-ray pulses^{10,11} with a 15- μ m diameter at full-width half-maximum (FWHM) beam focus by using a gas-focused micron-scale liquid stream (with a gas-dynamic virtual nozzle, or GDVN)¹² with a velocity of at least 50 m/s^{10,13}, and that molecular structures can be determined from data collected at the EuXFEL^{10,11}.

However, current sample injection technology results in substantial sample wastage between pulse trains¹¹. This poses a challenge for diminishing sample consumption, despite the MHz repetition rate within a pulse train. The volume of data collected per volume of sample consumed will, however, increase

significantly with the anticipated peak performance of the EuXFEL of up to 2700 pulses per train, and the upcoming LCLS-II, which will provide up to 1 million pulses per second¹⁴. Note that the repetition rate of the XFEL is not the only limitation for the rate of data collection at MHz XFELs. Currently, the fastest X-ray detector is the newly developed Adaptive Gain Integrating Pixel Detector (AGIPD), which can collect up to 352 images per train (potentially storing 3520 images per second)^{15,16}.

The required fast jet speed and high sample consumption was a challenge for our SFX experiment on the large membrane protein complex PSI at the EuXFEL. Previous MHz SFX experiments were performed with commercially available proteins (i.e., lysozyme and jack bean meal powder)¹⁰ or proteins that could be heterologously overexpressed in *Escherichia coli* in large amounts (e.g., BlaC)¹¹. However, it is difficult to express and purify membrane proteins in sufficient quality, quantity, and stability for X-ray crystallography, accounting for a major bottleneck in structural biology. To date, fewer than 1000 unique membrane protein structures have been solved (see <http://blanco.biomol.uci.edu/mpstruc/> for a current update on membrane protein structures). The problem is more pronounced for multi-subunit and ligand-rich membrane protein complexes, as they require specific cell machinery for membrane insertion and complex assembly. In the case of PSI, a complicated system of proteins is required for synthesis of cofactors and assembly of the complex¹⁷. This is why most structures from large membrane protein complexes, including PSI, can only be determined from native protein complexes, isolated from their natural host cells.

Here, we report the megahertz (MHz) SFX study of a large membrane protein complex at the EuXFEL, and discuss challenges that accompany the MHz repetition rates and how they were resolved. We prepared large batches of membrane proteins in sufficient amounts for the XFEL study and grew billions of uniform cyanobacterial Photosystem I (PSI) microcrystals that were of similar size to the XFEL beam focus, by crystallization at low ionic strength, a method that allowed the facilitation of fast sample delivery required for high X-ray pulse repetition rates. In addition, we describe how we overcame the challenges in the data collection and analysis of diffraction data from protein crystals with large unit cells that allowed us to determine the room-temperature (RT) structure of PSI at 2.9-Å resolution.

Results

Photosystem I isolation and crystallization. Here we briefly summarize the procedures and challenges of membrane protein isolation and crystallization for MHz SFX experiments. PSI was isolated from the natural thermophilic cyanobacterium *Thermosynechococcus elongatus*. The cells were grown in a 120-L photobioreactor (Supplementary Fig. 1a), and PSI was isolated by a 36-h multistep isolation procedure, including membrane isolation, detergent solubilization, ion exchange chromatography (Supplementary Fig. 1b), and crystallization as a final purification step (for more details see the “Methods” section). In total, 1000 mg of PSI was isolated and crystallized for XFEL experiments. PSI cannot be frozen at any step during the isolation and crystallization process, so protein was freshly isolated before the beamtime and recrystallized on-site. The first crystallization step was performed by concentrating the sample using ultrafiltration at low ionic strength (additional experimental details can be found in the Methods section). This produced a crystal suspension with a broad size distribution from ~200 nm to 100 μ m (Fig. 1a). The crystals could be sorted by size through sequential timed sedimentation, to separate the nanocrystals from larger crystals¹⁸. The nanocrystals were used as seeds for the final crystallization step. For the final crystallization, the larger crystals

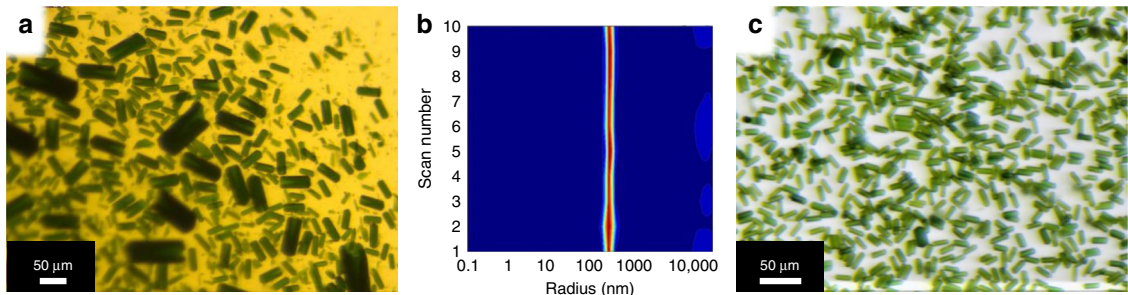


Fig. 1 Crystals of PSI. **a** Variable-size PSI crystal distribution, grown by ultrafiltration. **b** DLS (ten 18-s scans, numbered consecutively from 1 to 10, “Scan Number”) of nanocrystals with uniform size distribution used for seeding. **c** PSI crystals of uniform size of 5 × 5 × 15 μm grown by using the RAMS method

Table 1 Data collection statistics for PSI MHz SFX at the EuXFEL				
Detector distance (cm)	32.7	23.3	Combined	Dark-state data (pulses 1-10 only)
Hit rate (%)	1.0	1.0	~1	~1
Hits	7900	51,112	59,012	19,023
Indexed patterns (30 pulses/train)	7403	47,377	54,780	18,176
Indexing rate	94%	93%	93%	96%
Resolution at the edge of the detector (Å), horizontal, vertical	4.1, 3.5	3, 2.6	Various	Various
Minimum peak separation (pixels)	7.6	5.4	Various	5-7

The data were used from two different detector positions. Only the first ten pulses of a given train contributed to the dark PSI structure determined here

were dissolved by increasing the ionic strength, leading to a highly concentrated protein solution of 100 mg/mL PSI. Initial testing of optimal crystallization conditions was performed at low ionic strength (Supplementary Fig. 2).

To achieve uniform crystals for XFEL delivery, growth of crystals from the seeds must occur rapidly, thereby outrunning new nucleation events. This was achieved by a method that we term rotational agitated mixing with seeds (RAMS, see also Methods and Supplementary Fig. 3). Nanoseeds of uniform size (confirmed by dynamic light scattering (DLS), see Fig. 1b) were added to a precipitant solution consisting of the low ionic-strength buffer, before crystallization was induced by mixing of the seeded precipitant solution with small droplets of highly concentrated PSI (at 100 mg/mL). Incubation for 1 h at RT in the dark resulted in the growth of uniformly sized PSI crystals (15 × 5 × 5 μm).

Sample delivery. PSI microcrystals were delivered to the EuXFEL beam in a liquid jet produced by a GDVN¹² at a jet speed of 50 m/s, which is required to replenish the sample in 886 ns between the X-ray pulses of the EuXFEL^{10,11,13}. A high flow speed was achieved with a flow rate of 20 μL/min in glass GDVNs with an inner diameter (ID) of 50 μm. We accomplished successful sample delivery and replenishment by using 50-μm-ID nozzles due to the low-viscosity low ionic-strength buffer, high-quality manually ground glass GDVNs, and the avoidance of connections that could restrict the flow diameter and cause clogging.

Data collection and analysis. SFX data were collected at the SPB/SFX instrument of the EuXFEL^{19,20} by using X-rays of ~9.3 keV with an average pulse energy of 0.7–1 mJ, which corresponds to 4.7–6.7 × 10¹¹ photons/pulse upstream of the SPB/SFX hutch. The flux at the sample position was estimated to be reduced by ~50%, and the focal spot diameter was 16 ± 4 μm² at FWHM. The XFEL pulses were delivered in 30-pulse trains at 10 trains per second, with 886 ns between pulses within the trains (1.128 MHz), corresponding to an effective measurement rate of 300 pulses per second overall. The pulse duration was estimated from

the electron bunch length to be ~50 fs. This study focuses on the dark structure of PSI, which corresponded to the first ten images from each pulse train.

The data were collected with the AGIPD at three different distances (16.8, 23.3, and 32.7 cm from the sample) to assess indexing rate, quality (by providing sufficient peak separation), and resolution. Most data were collected at a detector distance of 23.3 cm with a resolution of 2.6 and 3.0 Å at the AGIPD edges vertically and horizontally, respectively. At 23.3-cm distance, the minimum peak separation was ~5 pixels (4 pixels for 16.8 cm).

The data acquisition was monitored²¹ in real time by using the online interface OnDA²². The data set reported here consists of 7,719,186 diffraction patterns from which 76,850 were identified as crystal diffraction patterns (hits) by Cheetah²³ corresponding to an average hit rate of ~1%. For the analysis, only runs with detector distances 23.3 and 32.7 cm were considered giving 59,012 patterns (hits, Table 1) in total. This hit rate is expected at the given crystal density of 1.68 × 10⁸ crystals/mL, which was experimentally determined by counting the crystals in a cell-counting chamber. This was the maximal PSI crystal density that could be achieved with this crystal size (5 × 5 × 15 μm); higher crystal densities led to clogging of the 50-μm-ID nozzles. With a jet speed of 50 m/s, a jet diameter of 5 μm, a beam diameter of 15 μm, and using conservative peak finder settings (i.e., only peaks consisting of at least 3 pixels), the maximal possible hit rate of 1.7% matched well to the observed hit rate of ~1%. A Gaussian unit cell distribution was observed (Fig. 2a) in the collected data. A typical diffraction pattern of PSI with close spot separation in the *a* and *c* directions is shown in Fig. 2b.

The hit rate remained relatively constant throughout the pulse sequences (Fig. 2c), indicating that the sample had been fully replenished between the pulses. The successful collection of diffraction data from PSI microcrystals was facilitated by the AGIPD’s feature of three individual gain stages (for details see Henrich et al.¹⁵). The diffraction pattern shown in Fig. 2b highlights reflections and pixels in the high- (black) and medium- (red) gain settings.

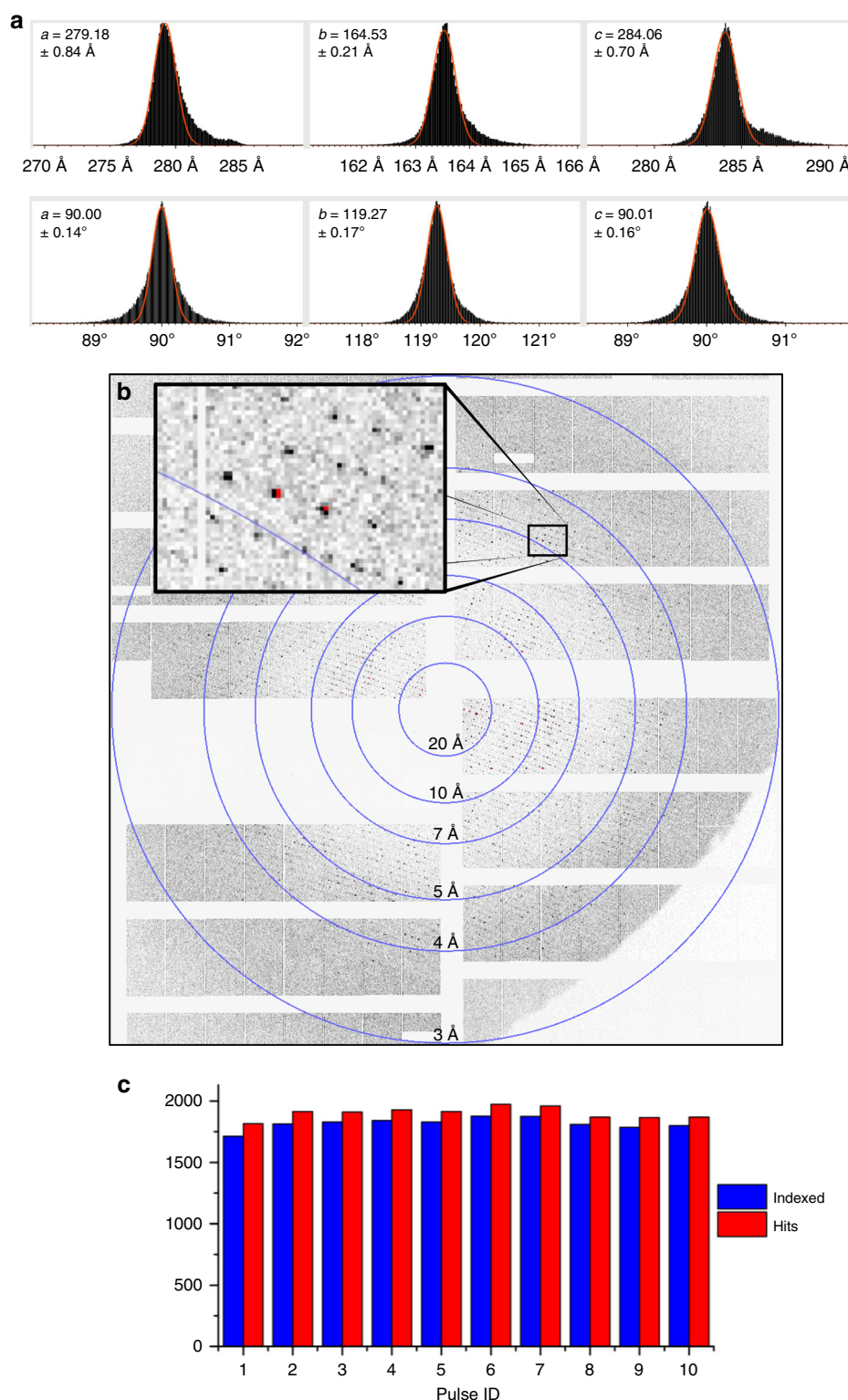


Fig. 2 Unit cell distribution, diffraction pattern, and indexing rates from PSI MHz SFX. **a** Unit cell distributions of PSI microcrystals showing narrowly distributed monoclinic unit cells. The red line shows a Gaussian function fit to the unit cell constant distribution, and the corresponding peak value is listed in each subpanel. **b** Representative X-ray diffraction pattern with pixels in high-gain mode shown in black, and medium- or low-gain mode shown in red. Resolution rings are shown and labeled accordingly. **c** The number of hits (red) and indexed patterns (blue, ~93% of hits) for each pulse

We observed 19,023 hits from the first ten pulses of the pulse train, representing diffraction images of the dark state of PSI. From these images, 18,176 (93%) were indexed in CrystFEL^{24,25} by using the indexing program Xgandalf²⁶.

Phases were obtained with CCP4's implementation of Phaser²⁷ by using the known X-ray structure of PSI from *T. elongatus* that was determined by using synchrotron radiation (PDB ID = 1JB0²⁸) as a starting model for molecular replacement. The final

Table 2 Crystallography data and refinement statistics for the XFEL and synchrotron data

Data collection	XFEL	Synchrotron
Space group	P2 ₁	P2 ₁
Dimensions: a, b, c (Å)	279.2, 164.5, 284.1	278.5, 165.1, 283.4
Dimensions: α , β , γ (°)	90, 119.3, 90.0	90, 119.4, 90
Number of hits	19,023	180
Indexed patterns	18,176	180
Resolution (Å)	64.32–2.9 (3.0–2.9)	48.9–2.9 (3.0–2.9)
I/ σ I	3.61 (0.52)	7.1 (1.9)
Completeness (%)	100 (100)	99.6 (98.3)
CC*	0.97 (0.31)	0.991 (0.890)
CC _{1/2}	0.88 (0.051)	0.633 (0.655)
Multiplicity	213 (122)	3.6 (1.9)
R _{split} (%)	22.07 (241.88)	N/A
R _{merge}	N/A	0.136 (0.328)
Wilson B-factor (Å ²)	51.3	55.38
Total number of reflections	104,759,101 (5,776,349)	1,769,616 (91,759)
Number of unique reflections	492,851 (47,459)	492,002 (47,514)
Refinement		
R _{work} /R _{free}	0.30/0.33	0.30/0.34
Number of atoms	72,533	72,738
Mean B-factor for all atoms (Å ²)	60	47
Mean B-factor for protein (Å ²)	61	46
Mean B-factor for ligands (Å ²)	60	47
Mean B-factor for solvent (Å ²)	33	32
CC _{work}	0.790 (0.161)	0.759 (0.606)
CC _{free}	0.790 (0.186)	0.697 (0.496)
RMSD		
Bond lengths (Å)	0.02	0.02
Bond angles (°)	2.38	2.44
Ramachandran allowed (%)	98.16	98.19

Where two values are quoted, these are the average overall resolution of all shells and the highest resolution shell (in parentheses)

data collection and refinement statistics for the PSI structure are shown in Table 2, and more details on the data analysis statistics can be found in the Methods and Supplementary Figs. 4–6. The generated electron density map is of high quality with important cofactors being well-resolved (see Fig. 3). For comparison of the SFX data with standard synchrotron-collected data, we additionally collected a data set from a large single PSI crystal, cutting the data at 2.9 Å to be comparable to the XFEL data set's resolution cutoff. The data statistics for both data sets are shown in Table 2, and more details on crystallization procedures, as well as data collection and analysis for the synchrotron-collected data are included in Supplementary Note 1. The processing of the cryogenic synchrotron structure also confirmed the P2₁ space group determined from the RT XFEL data (discussed further below).

The RT XFEL model and the cryogenic synchrotron model were found to be similar at this resolution (see superposition RMSD values in Supplementary Table 2). Long-range measurements on the structure showed that the RT XFEL structure is slightly expanded relative to the cryogenic synchrotron structure (Supplementary Fig. 7). Unsurprisingly, the RT XFEL structure has higher B-factors than the cryogenic synchrotron structure (Supplementary Fig. 8).

Discussion

The first structure determined at the EuXFEL was of small single-domain soluble proteins with comparatively small unit cells^{10,11}. Figure 4 shows the comparison of the trimeric PSI structure determined in this study with the structures of the four small proteins whose structures were previously solved via MHz crystallography at the EuXFEL^{10,11}. In addition, we provide a table comparing molecule size, crystal size, and the number of unit cells/crystal in Supplementary Table 3. While the structure determination of these proteins with MHz repetition rates was challenging, determining the structure of the PSI trimer, a very large membrane protein complex posed further challenges ranging from isolation and crystallization to data collection, processing, and structure refinement, which are discussed here.

PSI crystallization was performed by decreasing the ionic strength. Low ionic-strength crystallization by dialysis with micro- and macroseeding was established for the growth of large single crystals of PSI that formed the basis of the previously solved 2.5-Å synchrotron structure of PSI^{28,29}. This strategy is seldomly used in macrocrystallography as low ionic phase space is difficult to obtain by vapor diffusion, where all buffers and salts in the protein solution are concentrated. Low ionic-strength crystallization has several benefits for SFX: (i) crystallization at low ionic strength avoids harsh conditions or precipitants. (ii) It is fully reversible, so crystals can be dissolved and recrystallized multiple times to reach the desired crystal size distribution. (iii) These conditions result in a low-viscosity solution, close to that of water, allowing for high jet speeds to be reached without high back pressure. (iv) In the absence of salt and high-viscosity precipitants like polyethylene glycol (PEG), crystal slurries tend to clog less. (v) It avoids a common problem of SFX data collection, in which the impact of the X-ray beam on the sample jet leads to the accumulation of buffer components like salt or PEG, forming stalactites on the nozzle tip and stalagmites in the catcher that grow into the interaction region; these formations endanger the detector due to strong diffraction from the PEG or salt crystals and disturb the GDVN gas and sample stream. (vi) It is ideal for time-resolved studies, as the absence of PEG and salt reduces the amount of debris on the windows of the sample catcher used for sample visualization and in-coupling of the optical laser. More insight regarding low ionic-strength crystallization can be found in Supplementary Note 2. It should be noted that the settling rate of PSI crystals in low ionic-strength buffer depends on the crystal size and density. Thus, we used an anti-settling device³⁰ developed for the SPB/SFX instrument to prevent settling of the crystals in the sample reservoirs during data collection.

In SFX, crystals are ideally delivered to the interaction region at a rate equal to or greater than the X-ray pulse repetition rate whether for liquid jets, fixed targets, or other sample delivery methods³¹. The first published results from EuXFEL experiments show that a jet speed of at least 50 m/s is required to replenish the sample volume between X-ray pulses separated by ~886 ns, which can currently only be reached by injection by using the GDVN³² or the double-flow focused nozzle systems³³. Low sample consumption injection systems developed for membrane protein microcrystal delivery, like the high-viscosity injector³⁴, are currently too slow to be used at the MHz repetition rate produced at the EuXFEL. The high speed and sample flow rates required for sample injection present a challenge for proteins that are difficult to express and isolate in large amounts, with a large fraction of sample going to waste in between pulse trains. Only with exceptionally large efforts and large-volume bioreactor culture capability were we able to isolate 1000 mg of PSI from the native sources for this experiment. Ongoing research into pulse injectors^{35,36} and future developments in accelerator, detector,

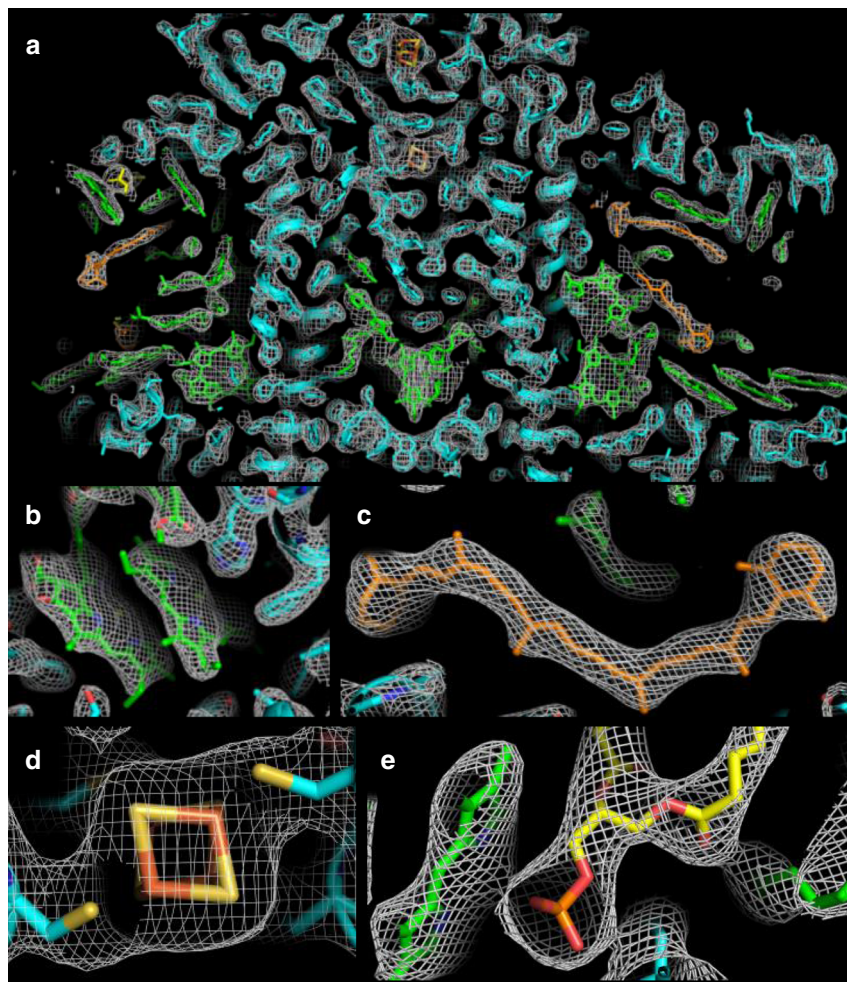


Fig. 3 Electron density map (2Fo-Fc at 1.5σ) and model of various PSI structural elements of the XFEL structure of PSI. In all images, protein is colored cyan, chlorophyll (Chl) molecules are colored green, β -carotenes are colored orange, and lipids are colored yellow. In panels **b–e**, nitrogen atoms are colored blue, oxygen atoms are colored red, and magnesium atoms are colored bright green. **a** A slice through the center of electron density of a monomer of PSI is shown, **b** the electron density of the “special pair” of Chls, P_{700} , **c** a β -carotene molecule, **d** the 4Fe-4S cluster, F_X , and **e** the phosphatidylglycerol lipid headgroup axial coordination of a Chl molecule

and computer technology may potentially alleviate this requirement in the future.

The adaptive gain switching of individual pixels provides the AGIPD with a large dynamic range, which is critical for accurate intensity measurement in crystallography, and was important for the data collection of PSI crystals that have a large unit cell. With the AGIPD’s multiple gain stages, data can be collected at full flux of the XFEL, where weaker high-resolution reflections are detected in high-gain mode, increasing their contribution to the high-resolution shell completeness. In contrast, intense low-resolution reflections are collected in medium- (or low-) gain mode, allowing for accurate determination of their intensity without the problem of saturation, leading to improved accuracy in the low-resolution shell. Figure 2b shows the distribution of Bragg peaks with those in high-gain mode in white and medium- or low-gain mode indicated in red. Supplementary Fig. 4 shows the high dynamic range of individual pixels in integrated Bragg spots from all indexed patterns. The limited size, however, of the currently installed AGIPD makes accurate spot location and indexing for samples with very large cells difficult and may preclude collection of high-resolution data. A 4-megapixel AGIPD is currently in the production stage for crystallography applications¹⁶.

Indexing crystallography data from crystals with large unit cells is not trivial. Multi-panel detector geometry optimization to subpixel accuracy becomes more critical than for samples with small unit cells because the distance between the Bragg peaks is only a few pixels (when the detector is sufficiently close to record high resolution). We used the program *geoptimiser*³⁷ to optimize the detector geometry including 16 individual distances to each 128×512 -pixel panel. Accurate peak selection and integration requires well-separated, sharp peaks. *CrystFEL’s indexamajig*²⁴ integrates peaks in 2D on each diffraction pattern. We used the three-ring integration method in which the outer annulus around each peak is used to calculate the local background and sigma. Integration radii were 2, 3, and 6 pixels to account for the small spot spacing.

Previous synchrotron-based structures of PSI from *T. elongatus*^{28,38–40} and the first SFX structure of PSI at 8-Å resolution² were determined in space group $P6_3$ with only one monomer of the trimeric PSI in the asymmetric unit (where the threefold trimer axis was a crystallographic axis). The unit cell dimensions were reported to be $a = 286$ Å, $b = 286$ Å, $c = 167$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$ from a 4-Å resolution data set collected at RT at a synchrotron⁴¹, and $a = 281.0$ Å, $b = 281.0$ Å, $c = 165.2$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$ from a 2.5-Å

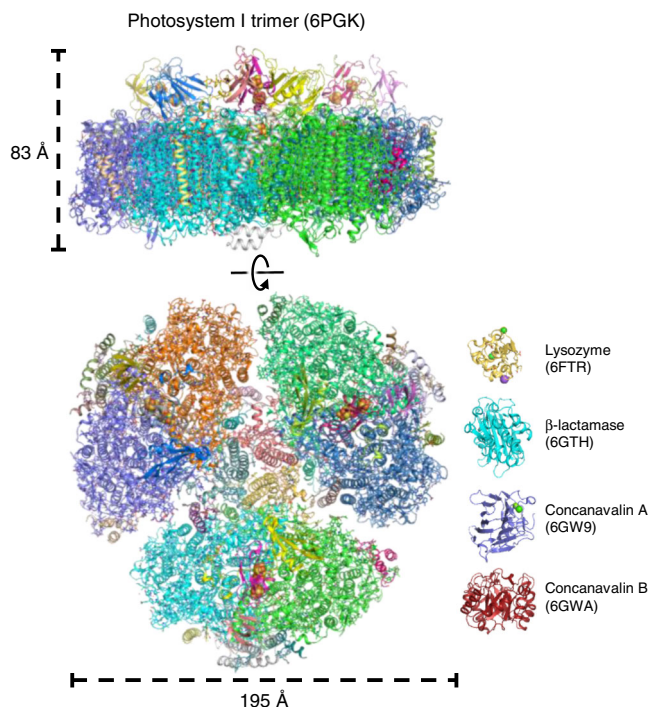


Fig. 4 Comparison of SFX structures from the EuXFEL. The structure of the trimeric PSI determined in this study is shown to scale with the four protein structures previously solved at the EuXFEL by using MHz repetition rates for comparison. Views from the membrane plane (top) and membrane normal (bottom) are shown for the PSI trimer, with major and minor axes denoted. Protein subunits are colored individually

resolution data set collected under cryogenic conditions²⁸. We originally attempted to analyze the EuXFEL data set both in space group $P6_3$ and without space group constraints. The data processing without space group constraints led to significantly improved statistics when data were merged in space group $P2_1$. By processing the SFX EuXFEL data set in the space group $P2_1$ we obtained unit cell constants of $a = 279.2$ Å, $b = 164.5$ Å, $c = 284.1$ Å, $\alpha = 90^\circ$, $\beta = 119.3^\circ$, and $\gamma = 90^\circ$.

We grew large PSI crystals by using the original crystallization and freezing procedures, and data sets were collected at the Advanced Photon Source (APS) under cryogenic conditions to 2.9-Å resolution (Table 2). The evaluation of this synchrotron data (Supplementary Text 1) confirmed the space group $P2_1$. The comparison of the data statistics of the RT EuXFEL and cryogenic synchrotron data sets is shown in Table 2. The spatial arrangement of PSI molecules in space groups $P6_3$ and $P2_1$ is very similar; the membrane-integral section of the PSI trimer is not oriented exactly parallel to the plane of the two long unit cell axes in the crystals. As the trimeric symmetry of the molecule is noncrystallographic, the space group is (pseudohexagonal) monoclinic $P2_1$. The packing of PSI in the space group $P2_1$ is shown in Supplementary Fig. 9. The quality of the data obtained from the EuXFEL experiment described here, as well as improved data evaluation software, allowed us to detect the difference between the two space groups more precisely. In space group $P2_1$, three times more indexed images are required than in space group $P6_3$, and thus the number of unique reflections is increased by a factor of 3.

The only other published PSI structure solved by SFX was determined at 8-Å resolution in one of the first experiments at the LCLS². Although this effort led to the emergence of SFX for structural biology, the limited resolution only allowed for the

identification of secondary structure elements such as α -helices and the electron-dense 4Fe-4S metal clusters. At 2.9-Å resolution, most amino acid side chains, hydrocarbon substituents, and even some well-conserved water molecules have been assigned with confidence (Fig. 3).

In the previously solved structure of PSI from *T. elongatus*²⁸, the three monomers that comprise the biologically relevant trimeric complex are perfectly equivalent, a result of the structure being solved in space group $P6_3$ that imposes crystallographic symmetry on the individual monomers, with one monomer in the asymmetric unit. In the $P2_1$ space group, the asymmetric unit contains the entire PSI trimer; this presented the opportunity to apply NCS⁴² within the refinement strategy, which has the potential to further improve the electron density map. However, when NCS was applied, the quality of the electron density map and its associated statistics did not improve. It was recently shown in the structure of PSI from another cyanobacterium, *Synechocystis* sp. PCC 6803, that processing the data without imposed symmetry in the space group $P2_1$, although quite different from the crystal packing reported here, resulted in apparent differences between the monomers of the trimer, allowing the assignment of extra lipid molecules⁴³. Supplementary Fig. 9 shows the crystal packing solved from the *T. elongatus* PSI XFEL data presented here, and Supplementary Fig. 10 shows the crystal packing reported from the *S. sp.* PCC 6803 structure. In the structures solved here, only minor differences were identified when comparing the electron density map and the resultant models for individual monomers (i.e., different rotamers for small side chains and slightly different Chl tail orientations, Supplementary Fig. 11), and no extra cofactors were identified (likely because the data are of lower resolution than the structure of PSI from *S. sp.* PCC 6803). However, long-range measurements (i.e., those spanning the entire complex) show that one of the three monomers of the trimer protrudes slightly away from the center relative to the other two monomers (Supplementary Fig. 7). Because the cryogenic synchrotron structure solved in $P2_1$ has lower B-factors (Supplementary Fig. 8) and is slightly compacted relative to the RT XFEL structure (Supplementary Fig. 7), this protrusion is less obvious in the former that speaks to the importance of understanding the structures of proteins at biologically relevant temperatures over cryogenic structures, a characteristic of XFEL-derived crystallographic data. While it may be that this observed asymmetry is an artifact of crystallization, a biological origin is also conceivable. In addition to the aforementioned asymmetry observed in PSI from *S. sp.* PCC 6803, it has recently been shown that asymmetry is observed in the *T. elongatus* PSI trimer when ferredoxin is bound (ferredoxin is the natural electron acceptor from PSI). It could be that in the assembly mechanism, two PSI monomers associate first but cause steric hindrance for the final monomer's insertion, disrupting the perfect $C3$ symmetry. We think it likely that processing higher-resolution data in the same space group will, in the future, enable the discovery of potentially important biological insights, especially with data collected from XFELs where the sample is not X-ray damaged.

Serial MHz crystallography holds potential for fast and accurate data collection at RT that is free of secondary X-ray damage. Its use, with concomitant improved methodology and tailored data analysis, may lead to the discovery of numerous structures of macromolecules in near-native states. With accelerated data collection and decreased sample consumption, it will soon be possible to obtain the high-multiplicity SFX data required for time-resolved studies at multiple time points in less time than is currently possible. The quality of merged SFX data is dependent on multiplicity, especially for larger, complex samples, for which collecting larger data sets is important. Larger data sets are also

required to resolve small or subtle structural differences in time-resolved sequences.

The challenge of producing sufficient sample volumes for proteins that are difficult to express, isolate, and crystallize for MHz serial crystallography applications will be ameliorated when the EuXFEL and the future LCLS-II reach their full pulse-rate capabilities. The EuXFEL's anticipated collection rate of 3520 frames per second with the AGIPD will enable the collection of complete SFX data sets within a few minutes as compared with hours currently. The prospect of continuous MHz pulse rates at future facilities such as the LCLS-II-HE offers even greater potential. Data acquisition and storage will then be the limitation of high-frequency XFELs, necessitating further detector development for online data processing and new avenues in high-throughput data transfer and storage to optimize their use¹⁴. With improved detector capabilities, further increase in repetition rates, and improved crystallization procedures, SFX can benefit from MHz crystallography at XFELs to determine molecular movies of various biologically relevant processes including ligand binding, substrate screening, and light-induced electron transfer, foretelling a bright future for this technique.

Methods

Large-scale cell culture of *T. elongatus*. Large-scale cultures of *T. elongatus* were grown in a 120-L cultivation photobioreactor (Photon System Instruments) that features controllable light intensity, timing, and modulation (Supplementary Fig. 1a). The starter cultures were maintained in 1-L flasks, and cells were grown in autoclaved BG-11 medium with constant agitation at 56 °C by using New Brunswick Innova shaker incubators. The small starter cultures were used to inoculate a 25-L photobioreactor (Photon System Instruments) that features the same control system as the 120-L reactor. Cell growth was monitored daily, and when the absorbance at 750 nm (A_{750}) reached 0.4, the culture was used to inoculate the 120-L photobioreactor. Initially, the cultures were dark-adapted for 6 h, and then the light intensity was linearly increased to 200 μ E (red:white = 3:1). The culture was harvested (in 4–5 days) when A_{750} reached 0.8.

For cell harvesting, the entire volume of the photobioreactor (120 L) was collected in a large container. The first stage of concentration was performed by using Cole Parmer's Masterflex Peristaltic Pump tangential flow filter system (Cat# 200-1558). Flow-through was discarded while the cells were concentrated in the filter with each flow cycle. The cell suspension was collected in a fresh beaker by flushing the system with purified water. The culture volume was reduced to 2–2.5 L, and the collected cells were centrifuged at 7000 $\times$ g at RT. After the first centrifugation round, additional cell suspension was added on top of the initial pellet and sedimented until all cells were separated from the supernatant. Cell pellets were frozen and stored at –80 °C. A typical harvest yielded ~80 g of cells per 120 L of cell culture.

PSI purification. Cells were thawed, and PSI was isolated as previously described with modifications²⁸. All steps were performed in dim green light. Cells were resuspended in buffer A (20 mM MES, pH = 6.4, 10 mM CaCl_2 , and 10 mM MgCl_2) and centrifuged at 4 °C at 7400 $\times$ g to wash cells before being resuspended in lysis buffer (20 mM MES, pH = 6.4, 10 mM CaCl_2 , 10 mM MgCl_2 , and 500 mM mannitol) and subsequently lysed by using a microfluidizer (Microfluidics Model M110-L) at 12,000 psi on ice. The cell lysate was sequentially centrifuged, and the thylakoid membranes were washed at 4 °C four times before the final resuspension in buffer A. Protein concentration was determined by performing chlorophyll (Chl) assays in 80% acetone ($\epsilon_{664} = 76780 \text{ M}^{-1} \text{ cm}^{-1}$). PSI was solubilized in buffer A with 0.75% weight per volume (w/v) β -dodecylmaltoside (β -DDM) for 45 min at a final Chl concentration of 0.75 mM. This solubilization extract was centrifuged for 2 h at 235,000 $\times$ g (Ti-45, Beckman Coulter) at 4 °C. The top layer of the supernatant (~10–15 mL/tube) was removed and discarded. The remaining supernatant was gently shaken in an orbital shaker at 4 °C for 45 min to separate PSI from the top layer of the membrane pellet. A Chl assay was performed on the supernatant, and aliquots equivalent to 80 μ mol of Chl were injected onto an XK 50/60 column (GE Healthcare) packed with DEAE anion exchange resin (Toyopearl) with a bed height of ~40 cm equilibrated with 20 mM MES, pH = 6.4, 15 mM MgSO_4 , and 0.02% w/v β -DDM. A step gradient was applied to elute first the PSI and PSII monomers and PSII dimers, before the PSI trimers were eluted from the column by a step gradient increasing the MgSO_4 concentration to 150 mM. PSI eluted at a MgSO_4 concentration of 145 mM (Supplementary Fig. 1b). The PSI fractions were pooled from multiple runs and concentrated in an Amicon 400-mL stirred ultrafiltration cell (EMD Millipore) fitted with a 100-kDa cutoff filter (Millipore) by using a headspace pressure of 30–60 psi argon. Once the target concentration of 10 mM Chl had been reached (typically ~100-fold concentration of the initial volume), the concentrated solution was diluted dropwise by addition of buffer G_0

(5 mM MES, pH = 6.4, 0.02% w/v β -DDM) to achieve a final MgSO_4 concentration of 4 mM. This solution was concentrated to 10 mM Chl. The following day, ~10 mL of buffer G_0 was applied to the mat of crystallized PSI that had formed on the membrane, and gentle pipette mixing was performed to harvest the crystals and remove them from the filter. These crystals were shipped to the EuXFEL as a crystal suspension at 4 °C and recrystallized on-site at the XFEL Biology Infrastructure (XBI) User Consortium laboratory, as crystal quality decreases during transport.

The crystal-containing aliquots could be pooled and sequentially settled for 10, 20, and 30 min, and overnight as described by Hunter¹⁸ to sort the crystals by size. For this beamtime, the small crystals were separated from the larger crystals in one settling step overnight. The supernatant of the overnight settling step contained small uniform nanocrystals, which were saved as a separate sample to be used for seeding. The crystals were stored in buffer G_0 at 4 °C. The seeds were characterized by DLS (Fig. 1b) to determine size and homogeneity profiles of the nanocrystal seeds. We also periodically performed SONICC analysis⁴⁴ to verify the crystallinity of the seed stock suspensions. All samples were stored at 4 °C in dim green light until further use in the final recrystallization experiments directly prior to the XFEL data collection.

Recrystallization. All crystals used for data collection were freshly grown at the XBI User Consortium laboratory at the EuXFEL directly prior to the experiment, which ensured size homogeneity and avoided damage during transport. The first fine screening was performed in test batches for each individual protein preparation prior to crystallization in larger sample-scale batches suitable for use in the SFX experiments. PSI was crystallized by decreasing the ionic strength with nanocrystal seeds added to the precipitant buffer (Supplementary Fig. 3). For each crystallization experiment, the pooled larger PSI crystals were dissolved by addition of MgSO_4 and subsequently recrystallized. A harvested preparation of PSI crystals was homogeneously resuspended by gentle pipette mixing before an aliquot was transferred to a preweighed microcentrifuge tube. This aliquot was then centrifuged at 13,000 $\times$ g for 2 min followed by removal of the bulk supernatant. This was repeated twice, and the final portion of supernatant was removed before weighing the crystal pellet. A μ L:mL equivalent of buffer G_{100} (100 mM MgSO_4 , 5 mM MES, pH = 6.4, and 0.02% w/v β -DDM) was added to the crystal pellet via pipette mixing, and the crystals were allowed to dissolve for at least 30 min at RT. Prior to crystallization experiments, optical microscopy by using a polarized filter was used to confirm that the solution was homogeneous and all crystals were dissolved. A Chl assay was then performed followed by dilution to 28 mM Chl (100 mg PSI/mL) by using buffer G_{50} (50 mM MgSO_4 , 5 mM MES, pH = 6.4, and 0.02% w/v β -DDM). To determine the optimal crystallization conditions, 1 μ L of this solution was placed in a 1.5-mL reaction vessel. The precipitant buffers G_X were prepared, which contained no (G_0), 1 (G_1), 2 (G_2), or 3 mM MgSO_4 (G_3) in 20 mM MES, pH = 6.4, 0.02% w/v β -DDM. Thirty seconds prior to mixing with the concentrated PSI, nanocrystal seeds were added to the G_X solution ($[\text{Chl}] = 3 \mu\text{M}$) to achieve homogeneous nucleation. The final concentration was 487 μM Chl. These test batches were then allowed to crystallize for 1–2 h before visual inspection with optical microscopy with emphasis on optimization of apparent morphology, size, and homogeneity. Small batch screening showed a large effect of MgSO_4 concentration on the size, size homogeneity, and morphology of the crystals as illustrated in Supplementary Fig. 2. Crystal macromorphology also served as an indicator; visible defects appeared more frequently at lower salt concentrations, possibly attributable to higher supersaturation conditions leading to faster crystal growth where crystallogenic growth kinetics outpace the individual molecular Gibbs energy minimization during lattice addition.

The crystal growth was typically complete in these test batches within 1–2 h, but interpretable results were realized within 30 min with no appreciable difference in ranking characteristics occurring thereafter. Supplementary Fig. 2 shows an example of the results from a crystallization test set. Once a suitable condition had been identified, scaled-up crystallogenesis was performed to obtain the larger volumes necessary for MHz crystallography as described in the “Results” section. The large-scale crystallization was achieved by using the RAMS method described in detail in Supplementary Fig. 3. Here, the PSI crystals were dissolved and diluted to 28 mM Chl as described above for the small test crystallization experiments. This concentrated solution was then added dropwise to the base of a suitably large beaker to allow for adequate surface area such that 5–20- μ L drops could be spread distinctly without consolidation. The previously optimal G_X solution was then doped with the same seed batch used during optimization and added on top of the protein drops at once to reach a Chl concentration of 487 μM Chl while simultaneous rotational mechanical mixing was performed (~120–180 rpm). Rotation was continued for at least 30 s beyond homogenization by eye. The solution was then transferred to a 50-mL conical tube and allowed to settle at RT overnight before characterization by light microscopy in both bright-field and under polarized light to check for signature birefringence. Upscaling from test batches to the large-scale RAMS method showed a general improvement in size homogeneity, possibly attributable to more rapid complete mixing, illustrated by homogeneous rod-shaped crystals depicted in Fig. 1c.

Sample delivery. In preparation for sample delivery, PSI microcrystals were allowed to settle by gravity overnight at 4 °C. The volume of the settled crystals was approximated, and most of the supernatant was removed to leave a volume of

supernatant equal to the volume of the settled crystals. The crystals were then gently resuspended in the supernatant, and the crystal suspension was prefiltered through the same 20- μm stainless-steel filter (IDEX) that was used for inline filtering during sample delivery. The sample reservoirs were mounted on an anti-settling device that maintained the temperature of the crystals at 4°C at the SBP/SFX instrument. The anti-settling device was developed by Robert Shoeman at the MPI Heidelberg³⁰. The crystals were delivered to the XFEL beam by using a GDVN liquid injection system^{12,13} with 50- μm inner-diameter glass capillaries that were individually hand-ground and tested at Arizona State University for the production of a straight jet and high sample flow rates. The small ID of the nozzle compared with the crystal size could have led to significant challenges for sample delivery. Our previous experience at XFELs indicated that ideal crystals should be maximally 1/10th the size of the ID of the nozzle to avoid clogging. Based on the size of our crystals ($5 \times 5 \times 15 \mu\text{m}^3$), 100- μm -ID nozzles could have been used. However, even when the flow rate is increased to 30 $\mu\text{L}/\text{min}$, a jet speed of only $\sim 12.5 \text{ m/s}$ is achieved with this nozzle, which is too slow to fully replenish the sample between pulses. The sample was delivered at a flow rate of 20 $\mu\text{L}/\text{min}$ to the injector by use of a syringe-reservoir system³⁰ where pressurized water from a high-performance liquid chromatography (HPLC) pump (Shimadzu) was delivered to the back side of the sample reservoir that displaces a fitted Teflon plunger separating the loaded crystal suspension from the pressurizing water. The liquid stream was focused by the sheath jet of co-propagating helium gas in the GDVN resulting in a jet speed of 50 m/s ¹¹. The gas flow was controlled by using a GP1 gas-pressure regulator (Proportion-air), and the flow rate was monitored with a gas flow meter (Bronkhorst). Required jet speeds were estimated based on the recovery of the jet as described by Wiedorn et al.¹¹.

Data collection. The PSI SFX experiments were performed by using the SPB/SFX instrument at the EuXFEL in November 2017 during the experiment P2066²⁰ in a similar manner to Wiedorn et al.¹¹. Data were collected at a photon energy of 9.3 keV with an average XFEL pulse energy of 0.7–1 mJ, and pulse duration of 50 fs. We estimate that the beam focus diameter was $16 \pm 4 \mu\text{m}$ (FWHM) based on the optical imaging of single shots by using Ce:YAG screens of various thicknesses. The X-ray diffraction data were recorded in single-shot mode by using the AGIPD 1 Mpixel with the direct beam passing through a central hole in the detector. The AGIPD consists of 2×8 application-specific integrated circuits, each with 128×512 pixels of $200 \times 200 \mu\text{m}$ size⁴⁵ and allows the collection of data at a frame rate matching the EuXFEL pulse rate of 1.1–4.5 MHz within a pulse train. This enables the measurement of up to 3520 images per second (in the 352 memory cells of each pixel per pulse train). Each pixel automatically switches between three gain modes from most to least sensitive: high, medium, and low, which allows for data collection with high dynamic range ($\sim 1 \times 10^4$ photons at 12 keV). For more details about the detector, the reader is referred to Allahgholi et al.⁴⁵.

Data analysis. Data from each AGIPD module were saved into separate files along with pulse and train ID numbers. The EuXFEL version of the hit-finding program Cheetah²³, as described in Wiedorn et al.¹¹, was used to match data from each of the 16 separate modules by train and pulse number, to process and compare data from the same X-ray pulse. Calibration of the AGIPD readout requires measurement of the pedestal, gain, and gain-switching threshold for each of the three gain stages in each memory cell of each pixel. AGIPD calibration, multi-gain-stage intensity correction, and masking of bad memory cells for each pixel were performed as described in Wiedorn et al.¹¹. Although the detector required complicated calibration¹¹, this step was essential to obtain more accurate intensities simultaneously for both the high- and low-resolution reflections, all of which are only partially measured in each snapshot.

The initial detector geometry was taken from the previous experiment¹¹ and further refined by using *geoptimiser* program³⁷. We used the three-ring integration method in *indexmajig*^{24,25}, in which the outer annulus around each peak is used to calculate the local background and sigma. Integration radii were 2, 3, and 6 pixels to account for the small spot spacing. The actual sample-to-detector distance for different detector positions was determined by the following criteria: at the correct detector distance the unit cell distribution has to be symmetrical. Small residual asymmetry (Fig. 2a) is probably due to the small shot-to-shot variation of the incident beam wavelength. The detector center was adjusted individually for the two detector positions due to the detector stage being slightly misaligned from the optical axis.

Hit finding was performed with Cheetah²³ by using the peakfinder8 algorithm with conservative parameters: minimum signal-to-noise ratio (SNR) of 6, pixel threshold of 200, minimum of 3 pixels per peak, and a minimum of 20 peaks per pattern. For the indexing (using *indexmajig* from CrystFEL version 0.8.0 + 1ccb8c35), different peak-finding parameters were used: minimum SNR of 6, pixel threshold of 50, minimum pixels per peak = 1, and a minimum of 50 peaks per pattern. To find peaks that are only one pixel in size, careful masking of unreliable regions on the detector was performed. In total (for all 30 pulses per train) 54,780 of 59,012 patterns were indexed. Of these, 99.8% of the 54,780 patterns were indexed by using *Xgandalf*²⁶ and the remaining 0.2% were indexed by using MOSFLM⁴⁶ and *DirAx*⁴⁷. The resulting unit cell distributions are very narrow with a clearly monoclinic (pseudohexagonal) lattice (Fig. 2a). *Xgandalf*²⁶ is a recently available indexing algorithm that demonstrated outstanding results, but so far only

a few experiments had employed it. Therefore, to verify the unit cell parameters, the indexing was repeated by using only MOSFLM⁴⁶ with the same CrystFEL parameters. This check resulted in very similar cell parameters (Supplementary Fig. 12) but with a wider, bimodal distribution for the *a*-axis. Also, indexing by using only MOSFLM resulted in fewer indexed patterns: 39,985 of 59,012.

Multicrystal indexing was not used in *indexmajig* because of the density of reflections from individual PSI microcrystals. The very small Bragg reflection profile radii (calculated by *indexmajig* to account for the large majority of the found reflections) are evidence of the high accuracy of indexing results from *Xgandalf* and orientation refinement in *indexmajig* (shown in Supplementary Fig. 6a–c for each detector distance). Supplementary Fig. 6d shows the diffraction resolution limits for all 30 pulses in the train, showing no evidence of sample damage in consecutive pulses. Data collection statistics for the XFEL data were calculated by using CrystFEL²⁵ and for the synchrotron data using Aimless in the CCP4 software suite⁴² (Table 2). The reflections from the first ten pulses (the dark structure) were merged with *process_hkl* in point group 2/m. The resulting Rsplit, CC*, the SNR, and completeness are shown in Supplementary Fig. 5.

Structure solution and refinement. For both the RT XFEL structure and cryogenic synchrotron structure, the phases were determined by molecular replacement with Phenix's implementation of Phaser²⁷ by using the previously determined structure of PSI from *T. elongatus*²⁸ as a starting model. All water molecules were removed from the starting model. *R*_{free} flags were assigned (5%) with phenix.refine⁴⁸. Three rounds of refinement by using phenix.refine⁴⁸, three rounds of refinement by using REFMAC5⁴⁹, and one round of find:waters by using Coot⁵⁰ were run on each data set by using the corresponding molecular replacement solution from Phaser. Refinement statistics were calculated with Phenix⁵¹. For data quality assessment of the RT XFEL structure in addition to the standard refinement output statistics, an annealing composite omit map omitting 0.5% of the atoms of the model within the asymmetric unit was generated by using the corresponding function within the Phenix software suite⁵¹ and is shown in Supplementary Fig. 13. The simulated annealing omit map appears similar to the 2Fo-Fc map (Fig. 3). We have also performed manual omission of various ligands in the XFEL structure and re-refined the map with the incomplete model. Clear density can be seen where the omitted ligand was removed (Supplementary Fig. 14).

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

Data availability

The source data underlying Fig. 2c and Supplementary Fig. 5 are provided as a Source Data File. Other data are available from the corresponding authors upon request. The XFEL structure has been deposited with the PDB accession code 6PGK, and the synchrotron structure has been deposited with the PDB accession code 6PFY. XFEL-collected diffraction data have been deposited in the Coherent X-ray Imaging Data Bank (CXIDB) under accession code 111.

Code availability

The versions of Cheetah and CrystFEL used in this work are available from the respective websites: <https://www.desy.de/~barty/cheetah> and <https://www.desy.de/~twwhite/crystfel>. The *Xgandalf* software is available at <https://stash.desy.de/users/gevorkov/repos/xgandalf> and is now implemented in the CrystFEL version 8.0 software suite.

Received: 29 January 2019; Accepted: 7 October 2019;

Published online: 04 November 2019

References

- Boutet, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 (2012).
- Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).
- Aquila, A. et al. Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt. Express* **20**, 2706–2716 (2012).
- Tenboer, J. et al. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 (2014).
- Stagno, J. R. et al. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
- Kupitz, C. et al. Structural enzymology using X-ray free electron lasers. *Struct. Dyn.* **4**, 044003 (2017).
- Olmos, J. L. et al. Enzyme intermediates captured ‘on the fly’ by mix-and-inject serial crystallography. *BMC Biol.* **16**, 59 (2018).
- Barends, T. R. M. et al. Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **350**, 445 LP–445450 (2015).
- Schmidt, M. Time-resolved macromolecular crystallography at pulsed X-ray sources. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms20061401> (2019).

10. Grünbein, M. L. et al. Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9**, 3487 (2018).
11. Wiedorn, M. O. et al. Megahertz serial crystallography. *Nat. Commun.* **9**, 4025 (2018).
12. DePonte, D. P. et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D: Appl. Phys.* **41**, 195505 (2008).
13. Wiedorn, M. O. et al. Rapid sample delivery for megahertz serial crystallography at X-ray FELs. *IUCr* **5**, 574–584 (2018).
14. Galayda, J. N. The new LCLS-II project: status and challenges. in *LINAC2014* 404–408 (2014). <https://pdfs.semanticscholar.org/4c4f/51c8414d37fae9bd3cbb70f53da2d599fc14.pdf>.
15. Henrich, B. et al. The adaptive gain integrating pixel detector AGIPD a detector for the European XFEL. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **633**, S11–S14 (2011).
16. Allahgholi, A. et al. Megapixels @ Megahertz – the AGIPD high-speed cameras for the European XFEL. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **942**, 162324 (2019).
17. Yang, H., Liu, J., Wen, X. & Lu, C. Molecular mechanism of photosystem I assembly in oxygenic organisms. *Biochimica et biophysica acta* 838–848. <https://doi.org/10.1016/j.bbabo.2014.12.011> (2015).
18. Hunter, M. S. & Fromme, P. Toward structure determination using membrane-protein nanocrystals and microcrystals. *Methods* **55**, 387–404 (2011).
19. Altarelli, M. et al. *The European X-Ray Free-Electron Laser—Technical Design Report* (2007). https://xfel.desy.de/localsExplorer_read?currentPath=/afs/desy.de/group/xfel/wof/EPT/TDR/XFEL-TDR-final.pdf.
20. Mancuso, A. P. et al. The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the European XFEL: initial installation. *J. Synchrotron Radiat.* **26**, 660–676 (2019).
21. Fangohr, H. et al. Data analysis support in Karabo at European XFEL. In *16th International Conference on Accelerator and Large Experimental Control Systems*, TUCPA01, 245–252 (Barcelona, Spain, 2017).
22. Mariani, V. et al. OnDA: online data analysis and feedback for serial X-ray imaging. *J. Appl. Cryst.* **49**, 1073–1080 (2016).
23. Barty, A. et al. Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Cryst.* **47**, 1118–1131 (2014).
24. White, T. A. et al. Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–689 (2016).
25. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).
26. Gevorkov, Y. et al. XGANDALF—extended gradient descent algorithm for lattice finding. *Acta Crystallogr. Sect. A* **75**, 694–704 (2019).
27. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
28. Jordan, P. et al. Three-dimensional structure of cyanobacterial Photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
29. Fromme, P. & Witt, H. T. Improved isolation and crystallization of photosystem I for structural analysis. *Biochim. Biophys. Acta - Bioenerg.* **1365**, 175–184 (1998).
30. Lomb, L. et al. An anti-settling sample delivery instrument for serial femtosecond crystallography. *J. Appl. Crystallogr.* **45**, 674–678 (2012).
31. Grünbein, M. L. & Nass Kovacs, G. Sample delivery for serial crystallography at free-electron lasers and synchrotrons. *Acta Crystallogr. Sect. D* **75**, 178–191 (2019).
32. Weierstall, U. Liquid sample delivery techniques for serial femtosecond crystallography. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130337 (2014).
33. Oberthuer, D. et al. Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep.* **7**, 44628 (2017).
34. Weierstall, U. et al. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5**, 1–15 (2014).
35. Echelmeier, A. et al. 3D printed droplet generation devices for serial femtosecond crystallography enabled by surface coating. *J. Appl. Crystallogr.* **52**, 997–1008 (2019).
36. Kim, D. et al. Electric triggering for enhanced control of droplet generation. *Anal. Chem.* **91**, 9792–9799 (2019).
37. Yefanov, O. et al. Accurate determination of segmented X-ray detector geometry. *Opt. Express* **23**, 28459–28470 (2015).
38. Krauss, N. et al. 3-dimensional structure of system-I of photosynthesis at 6 angstrom resolution. *Nature* **361**, 326–331 (1993).
39. Klukas, O. et al. Localization of two phytylquinones, QK and QK', in an improved electron density map of Photosystem I at 4-Å resolution. *J. Biol. Chem.* **274**, 7361–7367 (1999).
40. Klukas, O. et al. Photosystem I, an improved model of the stromal subunits Psac, Psad, and Psae. *J. Biol. Chem.* **274**, 7351–7360 (1999).
41. Krauß, N. et al. Photosystem I at 4 Å resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna system. *Nat. Struct. Biol.* **3**, 965–973 (1996).
42. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 235–242 (2011).
43. Malavath, T., Caspy, I., Netzer-El, S. Y., Klaiman, D. & Nelson, N. Structure and function of wild-type and subunit-depleted photosystem I in *Synechocystis*. *Biochim. Biophys. Acta - Bioenerg.* **1859**, 645–654 (2018).
44. Wampler, R. D. et al. Selective detection of protein crystals by second harmonic microscopy. *J. Am. Chem. Soc.* **130**, 14076–14077 (2008).
45. Allahgholi, A. et al. The adaptive gain integrating pixel detector at the European XFEL. *J. Synchrotron Radiat.* **26**, 74–82 (2019).
46. 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.* **67**, 271–281 (2011).
47. Duisenberg, A. J. M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Crystallogr.* **25**, 92–96 (1992).
48. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D.* **68**, 352–367 (2012).
49. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Crystallogr. Sect. D.* **67**, 355–367 (2011).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
51. Adams, P. D. et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 213–221 (2010).

Acknowledgements

The experiment was performed at the Scientific Instrument SPB/SFX at the EuXFEL in Schenefeld, Germany. We thank the EuXFEL for the allocation of XFEL beamtime and we thank the instrument group, facility staff, and the user office staff for their assistance. The crystallization of PSI was performed in the labs of the XBI user consortium at the EuXFEL and the support from all consortium members is highly appreciated. We also thank the XBI scientists for their support of the biochemical work in the XBI labs. We further thank Michele Zacks, Ph.D., for scientific editing and proofreading. We acknowledge funding from the Biodesign Center for Applied Structural Discovery at Arizona State University, and the following federal grants: the National Science Foundation (NSF) Science awards for Technology Center (STC) BioXFEL award no. STC-1231306 and award no. 1565180 (N.A.Z., R.A.K., S.B., and J.C.H.S.), the U.S. Department of Energy, Office of Science, Basic Energy Sciences award DE-SC0002164 and DE-SC0010575, and the National Institutes of Health grant R01GM095583. This work is also supported by the AXISIS project funded by the European Research Council under the European Union Seventh Framework Program (FP/2007–2013)/ERC Grant Agreement no. 609920. Funding was provided by the excellence cluster The Hamburg Center for Ultrafast Imaging—Structure, Dynamics, and Control of Matter at the Atomic Scale of the Deutsche Forschungsgemeinschaft (CUI, DFG-EXC1074), and the BMBF through the Roentgen-Angstrom Cluster grant 05K18CHA. This work was performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. M.A.C., M.L.S., and M.F. were supported by NIH grant R01GM117342. We acknowledge the support of the Australian Research Council through the Centre of Excellence in Advanced Molecular Imaging (CE140100011).

Author contributions

The experiment was conceived by P.F. with input from all authors. P.F., A.R., J.C.S., H.N.C., M.Sc., and R.F. led the experimental design. R.L. was the instrument scientist responsible for this experiment and he was supported by the head the SPB/SFX instrument, A.P.M., who led the development of the SPB/SFX instrument. J.M.G., J.D.Z., J.D.M., A.H., A.O., I.S., G.K., P.S., M.A.C., M.Sh., and M.F. participated in experiment planning. A.J. performed cell culturing. J.C., C.L.C., A.J., Z.D., E.D., S.R.C., and J.Y. performed PSI isolations. J.C., B.D.B., J.M., S.R.C., and M.H.A. performed PSI crystallization. J.C. worked on the crystallization and sample characterization of the crystals at the XBI lab. B.D.B. and J.M. assisted in protein crystallization, imaging, and sample preparation. M.O.W., A.E., J.C.V., C.K., J.K., V.Maz., S.Ba., G.B., S.Bo., and R.A.K. performed the sample delivery. The SPB/SFX instrument was prepared and operated by R.B., J.B., M.Me., K.D., Y.K., R.L., G.M., T.S., and B.W. Detector calibration and setup was performed by J.S.D., S.H., M.K., A.B., and V.M. AGIPD data collection was controlled by F.J., J.S.D., and M.T. Control and online analysis software was developed by V.M., S.Br., C.D., H.F., T.M., A.S., C.X., and M.Me. Logging was performed by M.S.H., V.Maz., R.F., S.S., and M.H.A. SFX data were analyzed by N.A.Z., O.N.Y., T.A.W., V.Mar., M.K., S.A., A.B., N.E.S., T.D.G., Y.G., N.A., D.B., S.Br., H.F., L.M., T.M., C.D., A.S., J.S., K.W., and M.Ma. Data analysis was performed by N.A.Z., N.A., D.B., L.M., M.Ma., G.P., and K.W. Structure determination, model building, and refinement of the XFEL structure was performed by C.G. and R.F. R.F. determined the synchrotron structure. The paper was written by C.G., J.C., P.F., and N.A.Z. with input from all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-12955-3>.

Correspondence and requests for materials should be addressed to P.F. or N.A.Z.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission 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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Chris Gisriel^{1,2,23,26}, Jesse Coe^{1,2,26}, Romain Letrun³, Oleksandr M. Yefanov⁴, Cesar Luna-Chavez^{1,2}, Natasha E. Stander^{1,2}, Stella Lisova^{1,5}, Valerio Mariani⁴, Manuela Kuhn⁴, Steve Aplin⁴, Thomas D. Grant^{6,7}, Katerina Dörner³, Tokushi Sato^{3,4}, Austin Echelmeier^{1,2}, Jorvani Cruz Villarreal^{1,2}, Mark S. Hunter⁸, Max O. Wiedorn^{4,9,10}, Juraj Knoska⁴, Victoria Mazalova⁴, Shatabdi Roy-Chowdhury^{1,2}, Jay-How Yang^{1,2}, Alex Jones^{1,2}, Richard Bean³, Johan Bielecki³, Yoonhee Kim³, Grant Mills³, Britta Weinhausen³, Jose D. Meza³, Nasser Al-Qudami³, Saša Bajt¹¹, Gerrit Brehm^{1,2,12,13}, Sabine Botha⁵, Djelloul Boukhelef³, Sandor Brockhauser^{3,14}, Barry D. Bruce^{15,16,17}, Matthew A. Coleman¹⁸, Cyril Danilevski³, Erin Discianno¹, Zachary Dobson^{1,2}, Hans Fangohr^{3,19}, Jose M. Martin-Garcia¹, Yaroslav Gevorkov^{4,20}, Steffen Hauf³, Ahmad Hosseinzadeh²¹, Friederike Januschek^{3,24}, Gihan K. Ketawala^{1,2}, Christopher Kupitz^{8,21}, Luis Maia³, Maurizio Manetti³, Marc Messerschmidt^{1,2,3}, Thomas Michelat³, Jyotirmoy Mondal¹⁵, Abbas Ourmazd²¹, Gianpietro Previtali³, Iosifina Sarrou⁴, Silvan Schön⁴, Peter Schwander²¹, Megan L. Shelby¹⁸, Alessandro Silenzi³, Jolanta Sztuk-Dambietz³, Janusz Szuba³, Monica Turcato³, Thomas A. White⁴, Krzysztof Wrona³, Chen Xu³, Mohamed H. Abdellatif⁴, James D. Zook^{1,2}, John C.H. Spence^{1,5}, Henry N. Chapman^{4,9,10}, Anton Barty⁴, Richard A. Kirian^{1,5}, Matthias Frank¹⁸, Alexandra Ros^{1,2}, Marius Schmidt²¹, Raimund Fromme^{1,2}, Adrian P. Mancuso^{3,22}, Petra Fromme^{1,2*} & Nadia A. Zatsepin^{1,5,25*}

¹Biodesign Center for Applied Structural Discovery, Arizona State University, Tempe, AZ 85287-5001, USA. ²School of Molecular Sciences, Arizona State University, Tempe, AZ 85287-1604, USA. ³European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany. ⁴Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22607 Hamburg, Germany. ⁵Department of Physics, Arizona State University, Tempe, AZ 85287-1504, USA. ⁶Hauptman-Woodward Institute, 700 Ellicott St, Buffalo, NY 14203-1102, USA. ⁷Department of Structural Biology, Jacobs School of Medicine and Biomedical Sciences, SUNY University at Buffalo, 700 Ellicott St, Buffalo, NY 14203-1102, USA. ⁸Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park 94025 CA, USA. ⁹Department of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. ¹⁰The Hamburg Centre for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. ¹¹Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22607 Hamburg, Germany. ¹²Institute for X-Ray Physics, University of Göttingen, 37077 Göttingen, Germany. ¹³Center Nanoscale Microscopy and Molecular Physiology of the Brain, Göttingen, Germany. ¹⁴Biological Research Centre, Hungarian Academy of Sciences, Temesvári krt. 62, Szeged 6726, Hungary. ¹⁵Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee at Knoxville, Knoxville, TN, USA 37996. ¹⁶Program in Energy Science and Engineering, University of Tennessee at Knoxville, Knoxville, TN, USA 37996. ¹⁷Department of Microbiology, University of Tennessee at Knoxville, Knoxville, TN, USA 37996. ¹⁸Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA. ¹⁹University of Southampton, University Rd, Southampton SO17 1BJ, UK. ²⁰Hamburg University of Technology, Vision Systems E-2, Harburger Schloßstraße 20, 21079 Hamburg, Germany. ²¹Department of Physics, University of Wisconsin-Milwaukee, 3135 N. Maryland Ave, Milwaukee, WI 53211, USA. ²²Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086 Victoria, Australia. ²³Present address: Department of Chemistry, Yale University, New Haven, CT 06520, USA. ²⁴Present address: Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22607 Hamburg, Germany. ²⁵Present address: Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne 3086 Victoria, Australia. ²⁶These authors contributed equally: Chris Gisriel, Jesse Coe.