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.
he advent of hard X-ray free-electron lasers (XFELs) and emergence of serial femtosecond crystallography (SFX)\textsuperscript{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\textsuperscript{1-9} 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\textsuperscript{1-9} 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.\textsuperscript{10} and on lysozyme and jack bean meal powder\textsuperscript{10} or proteins that could be heterologously overexpressed in Escherichia coli in large amounts (e.g., BlaC)\textsuperscript{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.biomed.uc.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 complex\textsuperscript{17}. 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 μm (Fig. 1a). The crystals could be sorted by size through sequential timed sedimentation, to separate the nanocrystals from larger crystals\textsuperscript{18}. The nanocrystals were used as seeds for the final crystallization step. For the final crystallization, the larger crystals
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 agitation 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 EuXFEL10,11,13. 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 hutch in a liquid jet produced by a GDVN12 at a jet speed of 50 m/s with an average pulse energy of 0.7–1 mJ, which corresponds 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 × 10⁹ 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.

**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 | 51112| 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.
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 by using the indexing program Xgandalf. Phases were obtained with CCP4’s implementation of Phaser by using the known X-ray structure of PSI from *T. elongatus* that was determined by using synchrotron radiation (PDB ID = 1JB0) as a starting model for molecular replacement. The final
The generated electron density map is of high quality with important SFX data with standard synchrotron-collected data, we addition 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 SFX data with standard synchrotron-collected data, we addition shown in Table 2, and more details on the data analysis statistics can be found in the Methods and Supplementary Figs. 4 – 6.

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 PSI28,29. This strategy is seldomly used in macromicroscopy as a 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 injectors35,36 and future developments in accelerator, detector,

---

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

| Data collection | XFEL | Synchrotron |
|----------------|------|-------------|
| Spacing (Å)    | 1.05 | 1.05        |
| Dimensions (Å) | 279.2, 164.5, 284.1 | 278.5, 165.1, 283.4 |
| Number of hits | 19,023 | 180 |
| Completeness (%) | 100 (100) | 99.6 (98.3) |
|Multiplicity | 213 (122) | 3.6 (1.9) |
| Rwork/Rfree | 0.30/0.33 | 0.30/0.34 |
| CC1/2 | 0.97 (0.31) | 0.991 (0.890) |
| CC1/2 | 0.88 (0.051) | 0.633 (0.655) |
| Mean B-factor for all atoms (Å²) | 492,851 (47,459) | 604,626 (58,854) |
| Mean B-factor for protein atoms (Å²) | 66 | 46 |
| Mean B-factor for ligands (Å²) | 60 | 47 |
| Mean B-factor for solvent (Å²) | 33 | 32 |
| Bond lengths (Å) | 0.02 | 0.02 |
| Bond angles (°) | 2.38 | 2.44 |
| Ramachandran allowed (%) | 98.16 | 98.39 |

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).
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 applications.

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 geooptimiser 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 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 and the first SFX structure of PSI at 8-Å resolution were determined in space group P63 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, α = 90°, β = 90°, and γ = 120° from a 4-Å resolution data set collected at RT at a synchrotron, and a = 281.0 Å, b = 281.0 Å, c = 165.2 Å, α = 90°, β = 90°, and γ = 120° from a 2.5-Å

---

**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 a–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, P700, c a β-carotene molecule, d the 4Fe–4S cluster, Fx, and e the phosphatidylglycerol lipid headgroup axial coordination of a Chl molecule.
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 *P*\(_6\)_3 that imposes crystallographic symmetry on the individual monomers, with one monomer in the asymmetric unit. In the *P*\(_2\)_1 space group, the asymmetric unit contains the entire PSI trimer; this presented the opportunity to apply NCS\(^{32}\) 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 *P*\(_2\)_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 molecules.\(^{43}\) 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 *P*\(_2\)_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

---

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.

---

resolving data set collected under cryogenic conditions\(^{28}\). We originally attempted to analyze the EuXFEL data set both in space group *P*\(_6\)_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 *P*\(_2\)_1. By processing the SFX EuXFEL data set in the space group *P*\(_2\)_1 we obtained unit cell constants of \(\alpha = 279.2\ \text{Å}, \ b = 164.5\ \text{Å}, \ c = 284.1\ \text{Å}\), \(\alpha = 90^\circ\), \(\beta = 119.3^\circ\), 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 *P*\(_2\)_1. The comparison of the data statistics of the RT EuXFEL and cryogenic data (Supplementary Text 1) confirmed the space group *P*\(_2\)_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 molecules.\(^{43}\) 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 *P*\(_2\)_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 developer development for online data processing and new avenues in high-throughput data transfer and storage to optimize their use.14 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 lipid 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 bioreactor (Photon System Instruments) control system as the 120-L reactor. Cell growth was monitored daily, and when the absorbance at 750 nm (A750) 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 µE (red:blue = 3:1). The culture was harvested after five days.

For cell harvesting, the entire volume of the photobioreactor (120 L) was filtered to remove the cells. A Chl assay was performed on the supernatant, which contained no (G0), 1 (G1), 2 (G2), or 3 mM MgSO4 (G3) in buffer G. For this beamtime, the small crystals were separated from the larger crystals in one sample. The supernatant of the overnight settling step was discarded, and the crystal fractions were washed with 40 µM MgSO4, 10 mM MES, and 6.4 M pH 6.4, and 0.02% w/v β-DDM (Supplementary Fig. 2). Crystal macromorphology also served as an indicator; visible defects appeared more frequently at lower salt concentrations, possibly attributable to better supersaturation conditions leading to faster crystal growth where crystallogenic growth kinetics outpace the individual nuclei growth kinetics.

**Rehydration**. The crystal-containing aliquots could be pooled and sequentially settled for 10, 20, and 30 min, and overnight as described by Hunter13 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 nanocrystals, which were saved for crystallization. The crystals were stored in buffer Ge 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 XFEL experiments, PSI was crystallized by decreasing the ionic strength of nanocrystal seeds added to the precipitant buffer. Supplementary Fig. 3. For each crystallization experiment, the pooled larger PSI crystals were dissolved by addition of MgSO4 and subsequently recrystallized. A harvested preparation of PSI crystals was homogeneously resuspended by gentle pipette mixing before an aliquot was transferred to a preweighed 1.5-mL centrifuge tube. This aliquot was then centrifuged at 13,000 × g for 2 min following by removal of the bulk supernatant. This was repeated twice, and the final portion of supernatant was removed before weighing the crystal pellet. A µL-mg equivalent of buffer Ge (100 mM MgSO4, 5 mM MES, pH = 6.4, and 0.02 w/v β-DDM) was added to the crystal pellet via pipette (Microfuge 15). 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 by dilution to 28 mM Chl (100 mg PSI/mL) and using buffer Ge (50 mM MgSO4, 5 mM MES, pH = 6.4, and 0.02 w/v β-DDM) to dissolve the Chl. Thirty seconds prior to mixing with the concentrated PSI, nanocrystal seeds were added to the G0 solution ([Chl] = 3 µM) to achieve homogenous nucleation. The final concentration was 487 mM 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 MgSO4 concentration on the size, size homogeneity, and morphology of the crystals, as illustrated in Supplementary Fig. 2. Crystals were seeded and also served as an indicator; visible defects appeared more frequently at lower salt concentrations, possibly attributable to better supersaturation conditions leading to faster crystal growth where crystallogenic growth kinetics outpace the individual molecular Gibbs energy minimization during lattice addition.

**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 equivalent to the volume of the settled crystals. The crystals were then gently resuspended in the supernatant, and the crystal suspension was prefixed through the same 20-µm stainless-steel filter (HEXI) that was used for inline filamentering 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 Shoemaker at the MPH Heidelberg28. The crystals were delivered to the XFEL beam by using a GDVN liquid injection system23,29 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 studies of the GDVN injection system 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_bak in point group 2/m. The resulting RepliC, 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. elongatus20 as a starting model. All water molecules were removed from the starting model. Rmerge flags were assigned (3%) with phenix. reindex. The three rounds of refinement were performed in phenix.refmac549,450 because of the density of reflections from individual PSI microcrystals. The very small Bragg reflection profile radii of the RT XFEL crystal structures (calculated by indexamajig 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 indexamajig (shown in Supplementary Fig. 6a-c for each detector distance). Supplementary Fig. 6d shows the diffraction resolution limits for the RT XFEL data, which are corrected for the two detector positions due to the detector stage being slightly misaligned from the optical axis. The resulting statistics were calculated with Phenix51. For data quality assessment of the RT XFEL structure in addition to the standard refinement output statistics, an annealing criterion (omit maps on 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).

Data availability. The data source 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 6PPY. 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/~barthi/cheetah and https://www.desy.de/~users/geworkov/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

1. Boutet, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. Science 337, 362–364 (2012).
2. Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. Nature 470, 73–78 (2011).
3. Aquila, A. et al. Time-resolved protein nanocrystallography using an X-ray free-electron laser. Opt. Express 20, 2706–2716 (2012).
4. Tenboer, J. et al. Time-resolved serial crystallography captures high-resolution intermediates of photosynthetic yellow protein. Science 346, 1242–1246 (2014).
5. Stagno, J. R. et al. Structures of riboswitch RNA reaction intermediates. Science 354, 242–246 (2017).
6. Kupitz, C. et al. Structural enzymology using X-ray free-electron lasers. Struct. Dyn. 4, 044003 (2017).
7. Olmos, J. L. et al. Enzyme intermediates captured on the fly by mix-and-inject XFEL serial crystallography. Nat. Comm. 4, 4454–44550 (2015).
8. Schmidt, M. Time-resolved macromolecular crystallography at pulsed X-ray sources. Int. J. Mol. Sci. https://doi.org/10.3390/ijms20061401 (2019).
32. Weierstall, U. Liquid sample delivery techniques for serial femtosecond crystallography enabled by surface coating. Opt. Express, 26, 6435–6442 (2018).

33. Oberthuer, D. et al. Double-strain crystallography. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 310–315 (2010).

34. Weierstall, U. et al. Megahertz serial crystallography. Philos. Trans. R. Soc. B Biol. Sci. 373, 20170138 (2018).

35. Echelmeier, A. et al. 3D printed droplet generation devices for serial crystallography at X-ray FELs. IUCr J. 5, 574–584 (2018).

36. Kim, D. et al. Electric triggering for enhanced control of droplet generation. Micromachines, 8, 149 (2017).

37. Klukas, O. et al. Localization of two phylloquinones, QK and QK’ in the photosystem I antenna complex. Biochim. Biophys. Acta - General Subjects, 1859, 1407–1416 (2016).

38. Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Cryst. Sect. D. Biol. Crystallogr. 67, 235–242 (2011).

39. Klukas, O. et al. Phylloquinone localization in Photosystem I. Acta Cryst. Sect. D. Biol. Crystallogr. 67, 346–365 (2011).

40. Klukas, O. et al. Localization and structure determination of Photosystem I antenna proteins. Acta Cryst. Sect. D. Biol. Crystallogr. 67, 182–191 (2011).

41. Klukas, O. et al. Phylloquinone localization in Photosystem I. J. Appl. Crystallogr. 45, 1151–1156 (2012).

42. Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Cryst. Sect. D. Biol. Crystallogr. 67, 235–242 (2011).

43. Malavath, T., Caspi, L., Netzer-El, S. Y., Kleinman, 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. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

45. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

46. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

47. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

48. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

49. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (2010).

50. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular structures. Acta Cryst. Sect. D. Biol. Crystallogr. 66, 352–367 (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 the 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 Michelle 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 (STC) BioXFEL award no. STC-1231306 and award no. 1156180 (N.A.Z., R.A.K., S.B., and I.C.H.S.), the U.S. Department of Energy, Office of Science, Basic Energy Sciences award DE-SC0012614 and DE-SC0010575, and the National Institutes of Health grant R01GM095583. This work is also supported by the AXSIS project funded by the European Research Council under the European Union Seventh Framework Program (FP2007–2013)/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 (DFG, EXC1074, the BBMBF through the Roentgen-Angstrom Cluster grant 03K18CHA. This work was performed, in part, under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07-LL042187, and the support of NSF grant 1053631. We acknowledge the support of the Australian Research Council through the Centre of Excellence in Advanced Molecular Imaging (CE110100011).

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., 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., O.S., G.K., P.S., M.A., and M.S., and M.F. participated in experiment planning. A.F. performed cell culturing. J.C., C.L.C., A.J., Z.D., E.D., S.R.C., and I.C.H.S., performed PSI crystallization. J.C. worked on the crystallization and sample characterization of the crystals at the EuXFEL. B.D.B. and J.M. assisted in protein crystallization, imaging, and sample preparation. M.O.W., A.E., J.C.Y., C.K., V.Maz., S.Ba., G.B., S.B., 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 T.J., J.S.D., and M.T. Control and online analysis software was developed by V.M., S.B., C.D., H.F., T.M., A.S. C.K., and M.M. 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.Maz., M.K., S.A., A.B., N.E.S., T.D.G., Y.G., N.A., D.B., S.B., H.F., T.M., C.D., A.S., J.K., and M.M. Data analysis was performed by N.A.Z., N.A., D.B., L.M., M.Ma., G.P., and K.W. Determination of macromolecular structure and model building. The paper was written by C.G., J.C.S., and P.F. with input from all other authors.

Competing interests

The authors declare no competing interests.

ARTICLE

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-12955-3

NATURE COMMUNICATIONS www.nature.com/naturecommunications

10
