# High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography

Sébastien Boutet[1], Lukas Lomb[2,3], Garth J. Williams[1], Thomas R.M. Barends[2,3], Andrew Aquila[4], R. Bruce Doak[5], Uwe Weierstall[5], Daniel P. DePonte[4], Jan Steinbrener[2,3], Robert L. Shoeman[2,3], Marc Messerschmidt[1], Anton Barty[4], Thomas A. White[4], Stephan Kassemeyer[2,3], Richard A. Kirian[5], M. Marvin Seibert[1], Paul A. Montanez[1], Chris Kenney[6], Ryan Herbst[6], Philip Hart[6], Jack Pines[6], Gunther Haller[6], Sol M. Gruner[7,8], Hugh T. Philipp[7], Mark W. Tate[7], Marianne Hromalik[9], Lucas J. Koerner[10], Niels van Bakel[11], John Morse[12], Wilfred Ghonsalves[1], David Arnlund[13], Michael J. Bogan[14], Carl Caleman[4], Raimund Fromme[15], Christina Y. Hampton[14], Mark S. Hunter[15], Linda Johansson[13], Gergely Katona[13], Christopher Kupitz[15], Mengning Liang[4], Andrew V. Martin[4], Karol Nass[16], Lars Redecke[17], Francesco Stellato[4], Nicusor Timneanu[18], Dingjie Wang[5], Nadia A. Zatsepin[5], Donald Schafer[1], James Defever[1], Richard Neutze[13], Petra Fromme[15], John C.H. Spence[5], Henry N. Chapman[4,16] and Ilme Schlichting[2,3]

1. Linac Coherent Light Source, LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA.

2. Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany.

3. Max Planck Advanced Study Group, Center for Free-Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany.

4. Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany.

5. Department of Physics, Arizona State University, Tempe, Arizona 85287, USA.

6. Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA.

7. Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA.

8. Wilson Laboratory, Cornell University, CHESS, Ithaca, NY 14853, USA.

9. Electrical and Computer Engineering, SUNY Oswego, Oswego, NY 13126, USA.

10. The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA.

11. Nikhef, National Institute for Subatomic Physics, Science Park 105, 1098 XG, Amsterdam, The Netherlands.

12. European Synchrotron Radiation Facility, 38043 Grenoble Cedex, France.

*13. Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden.*

*14. PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA.*

*15. Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA.*

*16. University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany.*

*17. Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at DESY, Hamburg, Germany.*

*18. Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden.*

**Abstract**
**Structure determination of proteins and other macromolecules has historically required the growth of high-quality crystals sufficiently large to diffract x-rays efficiently while withstanding radiation damage. Here we apply serial femtosecond crystallography (SFX) using an x-ray free-electron laser (XFEL) to obtain high resolution structural information from microcrystals (<1×1×3 μm$^3$) of the well-characterized model protein lysozyme. The agreement with synchrotron data demonstrates the immediate relevance of SFX for analyzing the structure of the large group of difficult-to-crystallize molecules.**

Elucidating macromolecular structures by x-ray crystallography is an important step in the quest to understand the chemical mechanisms underlying biological function. Although facilitated greatly by synchrotron x-ray sources, the method is limited by crystal quality and radiation damage *(1)*. Crystal size and radiation damage are inherently linked, as reducing radiation damage requires lowering the incident fluence. This in turn calls for large crystals that yield sufficient diffraction intensities while reducing the dose to individual molecules in the crystal. Unfortunately, growing well-ordered large crystals can be difficult in many cases, particularly for large macromolecular assemblies and membrane proteins. In contrast, micron-sized crystals are frequently observed. Although diffraction data of small crystals can be collected using micro-focus synchrotron beamlines, this remains a challenging approach due to the rapid damage suffered by these small crystals *(1)*.

Serial femtosecond crystallography (SFX) using x-ray free-electron laser (XFEL) radiation is an emerging method for 3D structure determination using crystals ranging from a few micrometers to a few hundred nanometers in size and potentially even smaller. This method relies upon x-ray pulses that are both sufficiently intense to produce high quality diffraction while of short enough

2

duration to terminate before the onset of significant radiation damage *(2-4)*. X-ray pulses of only 70 femtoseconds duration terminate before any chemical damage processes have time to occur, leaving primarily ionization and X-ray induced thermal motion as the main sources of radiation damage *(2-4)*. SFX therefore promises to break the correlation between sample size, damage and resolution in structural biology. In SFX, a liquid microjet is used to introduce fully hydrated randomly oriented crystals into the single-pulse XFEL beam *(5-8)*, as illustrated in Fig. 1. A recent low-resolution proof-of-principle demonstration of SFX performed at the Linac Coherent Light Source (LCLS) *(9)* using crystals of photosystem I ranging in size from 200 nm to 2 μm produced interpretable electron density maps *(6)*. Other demonstration experiments using crystals grown in-vivo *(7)* as well as in the lipidic sponge phase for membrane proteins *(8)* were recently published. However, in all these cases, the x-ray energy of 1.8 keV (6.9 Å) limited the resolution of the collected data to approximately 8 Å. Data collection to a resolution better than 2 Å became possible with the recent commissioning of the LCLS Coherent X-ray Imaging (CXI) instrument *(10)*. The CXI instrument provides hard x-ray pulses suitable for high-resolution crystallography and is equipped with Cornell-SLAC Pixel Array Detectors (CSPADs) consisting of 64 tiles of 192 × 185 pixels each, arranged as shown in Fig. 1 and Figs. S1 and S2. The CSPAD supports the 120 Hz readout rate required to measure each x-ray pulse from LCLS *(11)*.

Here we describe SFX experiments performed at CXI analyzing the structure of hen egg white lysozyme (HEWL) as a model system using microcrystals of approximately $1 \times 1 \times 3$ μm$^3$ *(4,11)*. HEWL is an extremely well-characterized protein that crystallizes easily. It was the first enzyme to have its structure determined by x-ray diffraction *(12)*, and has since been thoroughly characterized to very high resolution *(13)*. Lysozyme has served as a model system for many investigations, including radiation damage studies. This makes it an ideal system for the development of the SFX technique. Microcrystals of HEWL in random orientation were exposed to single 9.4 keV (1.32 Å) x-ray pulses of 5 fs or 40 fs duration focused to 10 μm$^2$ at the interaction point (Fig. 1). The average 40 fs pulse energy at the sample was 600 μJ/pulse, corresponding to an average dose of 33 MGy deposited in each crystal. This dose level represents the classical limit for damage using cryogenically-cooled crystals *(14)*, . The average 5 fs pulse energy was 53 μJ. The SFX-derived data were compared to low-dose datasets collected at room temperature using similarly prepared larger crystals *(11)*. This benchmarks the technique with a well-characterized model system.

We collected approximately 1.5 million individual "snap-shot" diffraction patterns for 40 fs duration pulses at the LCLS repetition rate of 120 Hz using the CSPAD. About 4.5 % of the patterns were classified as crystal hits, 18.4 % of which were indexed and integrated with the CrystFEL software *(15)* showing excellent statistics to 1.9 Å resolution (see Table 1 and Table S1). In addition, 2 million diffraction patterns were collected using x-ray pulses of 5 fs duration, with a 2.0 % hit rate and a 26.3 % indexing rate, yielding 10,575 indexed patterns. The structure, partially shown in Fig. 2A, was determined by molecular replacement (using PDB entry 1VDS) and using the 40 fs SFX data. No significant differences were observed in an $F_{obs}$[40 fs] − $F_{obs}$[synchrotron] difference electron density map (Fig. 2B). The electron density map shows features that were not part of the model (different conformations of amino acids, water molecules) and show no discernable signs of radiation damage. Also, when the data were phased with molecular replacement using the turkey lysozyme structure as a search model (PDB code 1LJN), the differences between the two proteins were immediately obvious from the maps (Fig. S3).

Even though the underlying radiation damage processes differ due to the different time scales of the experiments using an XFEL and a synchrotron/rotating anode (femtoseconds vs. seconds/hours), no features related to radiation damage are observed in difference maps calculated between the SFX and the low-dose synchrotron data (Fig. 2B). In addition to local structural changes, metrics like $I/I_0$ and the Wilson B-factor are most often used to characterize global radiation damage in protein crystallography ($17$). $I/I_0$ is not applicable to the SFX data. However, the Wilson-B factors of both SFX data sets show values typical for room temperature data sets and do not differ significantly from those obtained from synchrotron and rotating anode data sets collected with different doses, using similarly grown larger crystals kept at room temperature and fully immersed in solution ($11$) (Table 1 and S1). The R-factors calculated between all collected data sets do not show a dose dependent increase (Figure S4). However, higher R-factors are observed for the SFX data, indicating a systematic difference. This is not caused by non-convergence of the Monte Carlo integration since scaling the 40 fs and 5 fs data together does not affect the scaling behavior (not shown). Besides non-isomorphism, possible explanations for this difference could include suboptimal treatment of weak reflections, the difficulties associated with processing still diffraction images and other SFX-specific steps in the method.SFX is an emerging technique, and data processing algorithms, detectors and data collection methods are under continuous development.
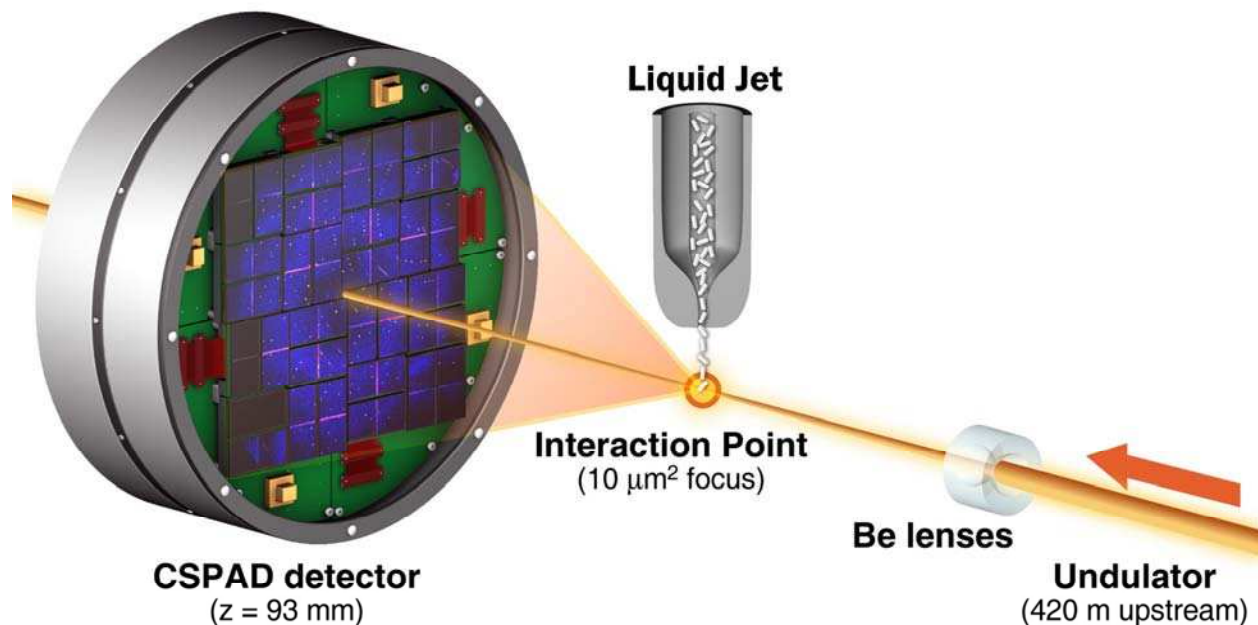
**Fig. 1:** Experimental geometry for serial femtosecond crystallography at the Coherent X-ray Imaging instrument. Single pulse diffraction patterns from single crystals flowing in a liquid jet are recorded on a CSPAD at the 120 Hz repetition rate of LCLS. Each pulse was focused at the interaction point using 9.4 keV x-rays. The sample-to-detector distance (z) was 93 mm.
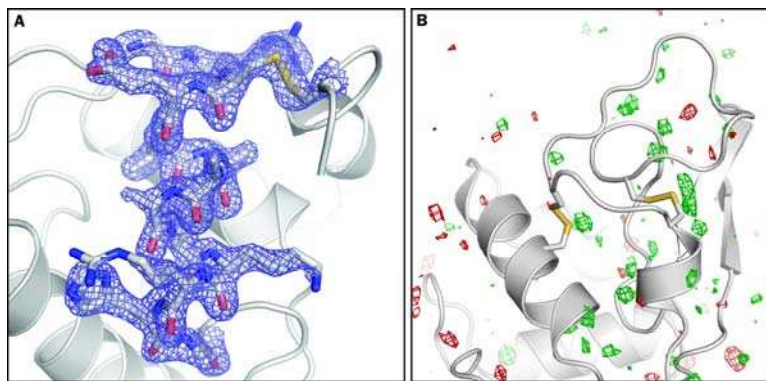


**Fig. 2: A**: Final, refined $2m F_{obs}$-$DF_{calc}$ (1.5σ) electron density map *(16)* of lysozyme at 1.9 Å resolution calculated from 40 fs pulse data.  **B**: $F_{obs}$[40 fs]-$F_{obs}$[synchrotron] difference Fourier map, contoured at +3 σ (green) and -3 σ (red). No interpretable features are apparent. The synchrotron dataset was collected with a radiation dose of 24 kGy.

**Table 1. SFX and synchrotron data and refinement statistics**

| Parameter | 40 fs* pulses | 5 fs* pulses | SLS RT data 3 **** |
|---|---|---|---|
| Wavelength | 1.32 Å | 1.32 Å | 0.9997 Å |
| X-ray focus [$\mu m^2$] | ~ 10 | ~ 10 | ~ 100 x 100 |
| Pulse energy/fluence at sample | 600 µJ/ $4 \times 10^{11}$ ph/pulse | 53 µJ/$3.5 \times 10^{10}$ ph/pulse | N.A./ $2.5 \times 10^{10}$ ph/s |
| Dose [MGy] | 33.0 per crystal | 2.9 per crystal | 0.024 total |
| Dose rate [Gy/s] | $8.3 \times 10^{20}$ | $5.8 \times 10^{20}$ | $9.6 \times 10^2$ |
| Space group | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ |
| Unit cell length [Å], $\alpha=\beta=\gamma=90°$ | $a=b=79$, $c=38$ | $a=b=79$, $c=38$ | $a=b=79.2$, $c=38.1$ |
| Oscillation range/exposure time | Still exp. / 40 fs* | Still exp. / 5 fs* | 1.0°, 0.25 s |
| # collected diffraction images | 1471615 | 1997712 | 100 |
| # of hits/indexed images | 66442 /12247 | 40115/10575 | n.a./100 |
| Number of reflections | n.a. | n.a. | 70960 |
| Number of unique reflections | 9921 | 9743 | 9297 |
| Resolution limits [Å] | 35.3-1.9 | 35.3-1.9 | 35.4-1.9 |
| Completeness** | 98.3% (96.6%) | 98.2% (91.2%) | 92.6% (95.1%) |
| I/σ(I)** | 7.4 (2.8) | 7.3 (3.1) | 18.24 (5.3) |
| $R_{split}$*** | 0.158 | 0.159 | n.a. |
| $R_{merge}$ | n.a. | n.a. | 0.075 (0.332) |
| Wilson B-factor***** | 28.3 Å$^2$ | 28.5 Å$^2$ | 19.4 Å$^2$ |
| R-factor/R-free****** | 0.196/0.229 | 0.189/0.227 | 0.166/0.200 |
| Rmsd bonds, Rmsd angles****** | 0.006 Å, 1.00° | 0.006 Å, 1.03° | 0.007 Å, 1.05° |
| PDB code | 4ET8 | 4ET9 | 4ETC |

\* Electron bunch length         ** Highest resolution shell: 2.0-1.9 Å

\*\*\* $R_{split}$ as defined in *(15)*  $$R_{split} = \left(\frac{1}{\sqrt{2}}\right) \cdot \frac{\sum_{hkl} \left| I_{hkl}^{even} - I_{hkl}^{odd} \right|}{\frac{1}{2} \sum_{hkl} \left| I_{hkl}^{even} + I_{hkl}^{odd} \right|}$$

\*\*\*\* Statistics from XDS *(18)*     \*\*\*\*\* Calculated with TRUNCATE *(19)*

\*\*\*\*\*\* Calculated with PHENIX *(20)*

A simple consideration shows the attainable velocities of atoms in the sample depend on the deposited X-ray energy versus the inertia of those atoms: $\langle v \rangle = \sqrt{3k_B T / m}$, where $m$ is the mass of a carbon atom, for example, $T$ is temperature and $k_B$ is Boltzmann's constant. For an impulse absorption of energy at the doses of our LCLS measurements we predict average velocities less than 10 Å / ps, which gives negligible displacement during the FEL pulses. On the timescale of femtoseconds, radiation damage is primarily caused by impulsive rearrangement of atoms and electron density, rather than the relatively slow processes of chemical bond breaking typical in conventional crystallography using much longer exposures at much lower dose rates (the dose rate in this experiment was approximately 0.75 MGy per femtosecond).

Neither the SFX electron density maps nor the Wilson B-factors suggest obvious signs of significant radiation damage. Very short pulses (5 fs electron bunch) are not expected to produce observable damage, according to simulations *(3)*. Furthermore, it has been reported that the actual x-ray pulses are shorter than the electron bunches for XFELs, making the pulse duration possibly shorter than the relevant Auger decays *(21)*. The agreement between the SXF results using 40 fs pulses and 5 fs pulses suggests similar damage characteristics for the two pulse durations based on the available data. Our results demonstrate that under the exposure conditions used, SFX yields high quality data suitable for structural determination. SFX reduces the requirements on crystal size and therefore the method is of immediate relevance for the large group of difficult-to-crystallize molecules, establishing SFX as a very valuable high-resolution complement to existing macromolecular crystallography techniques.

**References and Notes**

1. J. M. Holton, K. A. Frankel, *Acta Crystallogr. D***66**, 393 (2010).
2. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, *Nature* **406**, 752 (2000).
3. A. Barty *et al.*, *Nature Photonics* **6**, 35 (2012).
4. L. Lomb *et al.*, *Phys. Rev. B* **84**, 214111 (2011).
5. D. P. DePonte *et al.*, *J. Phys. D***41**, 195505 (2008).
6. H. N. Chapman *et al.*, *Nature* **470**, 73 (2011).
7. L.C. Johansson *et al.*, *Nature Methods* **9**, 263 (2012).
8. R. Koopmann *et al.*, *Nature Methods* **9**, 259 (2012).
9. P. Emma *et al.*, *Nature Photonics* **4**, 641 (2010).
10. S. Boutet, G. J. Williams, *New Journal of Physics* **12**, 035024 (2010).
11. Materials and methods are available as supporting material on *Science* Online.
12. C. C. F. Blake *et al.*, *Nature* **206**, 757 (1965).
13. J. Wang, M. Dauter, R. Alkire, A. Joachimiak, Z. Dauter, *Acta Crystallogr. D* **63** 1254 (2007)
14. R. L.Owen, E. Rudino-Pinera, E. F. Garman, *Proc. Natl. Acad. Sci.* **103**, 4912-4917 (2006)
15. T. A. White *et al.*, *J. Appl. Crystallogr.* **45**, 335-341 (2012).
16. R. J. Read, *Acta Crystallogr. A***42**, 140 (1986).
17. R.J. Southworth-Davies, M.A. Medina, I. Carmichael, E.F. Garman, *Structure* **15,** 1531–1541 (2007)
18. W. Kabsch, *J. Appl. Chrystallogr.* **26** 795-806 (1993)
19. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr. D***61**, 458 (2005).
20. P. D. Adams *et al.*, *Acta Crystallogr. D* **66**, 213 (2010).
21. L. Young *et al.*, *Nature* **466**, 56 (2010).
22. H. T. Philipp, M. Hromalik, M. Tate, L. Koerner, S. M. Gruner, *Nucl Instrum. Methods A***649**, 67 (2011).
23. The cheetah software package is available via the url: http://www.desy.de/~barty/cheetah/
24. R. A. Kirian *et al.*, *Optics Express* **18**, 5713 (2010).
25. R. A. Kirian *et al.*, *Acta Crystallogr. A* **67**, 131 (2011).
26. S. Bailey, *Acta Crystallogr. D* **50** 760-763 (1994)

**Materials and Methods**

<u>CSPAD Detector</u>

Photons are absorbed in a 500 μm thick, fully depleted layer of high-resistivity silicon for direct x-ray conversion, bump-bonded to a custom Application Specific Integrated Circuit (ASIC) *(22)*.  The CSPAD detector consists of 32 modules, with 2 ASICs each, tiled together to form a 1516 × 1516 pixel array with small gaps between tiles as shown in Fig. S1 and S2. Each detector pixel is 110×110 $\mu m^2$ wide and the total area covered is approximately 178 × 178 $mm^2$ with the gaps not sensitive to x-rays. The tiles are arranged in four identical quadrants, with a central gap variable between 1 and 9.5 mm that allows the direct beam to safely pass through. A low or high gain setting can be chosen for each individual pixel. Here, the entire detector was used in high gain mode, which provides single photon sensitivity with a signal-to-noise ratio of about 3.5 (FWHM) and a dynamic range of about ~350 photons at 9.4 keV.

<u>Sample Preparation</u>

LCLS samples: Rodshaped microcrystals ($\leq 1 \times 1 \times 3$ $\mu m^3$) of hen egg white lysozyme (Sigma) were grown as described previously *(4)* omitting the cross-linking step. After crystallization, the solution used for growing the crystals (20 % NaCl, 6 % PEG 6000, 1 M Na acetate pH 3.0) was exchanged for storage solution (10 % NaCl, 1.0 M Na acetate pH 3.4).

Synchrotron/rotating anode samples: The crystals were grown in Linbro plates in hanging or sitting drop geometry. Drops of equal volume of protein (20 mg/ml) and reservoir solution (1 M NaAc pH 3.0, 9-10 % NaCl, 6 % PEG 6000) were mixed. For the measurements, the crystals were equilibrated in 1 M NaAc pH 3.4, 10 % NaCl.

<u>SFX data collection at CXI</u>

Using a liquid microjet *(5)*, the crystals were injected into the FEL beam in their storage solution. Single shot diffraction patterns were recorded at 120 Hz while the liquid jet was flowing. All data were saved, regardless of whether a given shot contained a hit crystal, and processed offline later. A data rate of 600 MB/sec was sustained for the duration of the experiment. The sample-detector distance was 93 mm. The pulse lengths stated here correspond to the electron bunch lengths with the actual photon pulse lengths expected to be shorter *(21)*.

<u>Conventional room temperature diffraction data collection</u>

Lysozyme crystals (~ 100 × 100 × 400 $\mu m^3$) were mounted in quartz capillaries (1 mm diameter (SLS data 1,2), 0.7 mm diameter (SLS data3)) filled with storage solution (10 % NaCl in 1.0  M Na acetate pH 3.4), with the crystal fixed with pipe cleaner fibers. This setup was chosen because it resembles that of the microjet with a crystal injected in its storage solution. Apart from the influence of the quartz capillary, the location of the crystal in the solution differs. Presumably, the microcrystal is located in the center of the microjet. However, to provide good visibility of the crystal in the capillary setup using an on-axis microscope, the crystal was located close to the capillary wall facing the x-ray beam. This results in the diffracted x-rays passing through a fair amount of liquid which may have resulted in attenuation of the diffracted signal. Diffraction data were collected

1

at room temperature at the PXII beamline at the Swiss Light Source (SLS), using a PILATUS 6M detector. To restrict the dose, the x-ray beam was defocused to 50 x 50 $\mu m^2$ (SLS data1,2) and 100 x 100 $\mu m^2$ (SLS data3), and data were collected with different oscillation ranges and exposure times as listed in Table1 and Table S1. Data processing was done with XDS *(18)*. Room temperature data collected of a lysozyme crystal mounted relatively dry in a quartz capillary were non-isomorphous with the SFX data (data not shown), likely due to dehydration which might be more of an issue for the lysozyme crystals used in this study because of their non-standard, high salt crystallization conditions. For further comparison, a dataset was also collected of a lysozyme crystal cryoprotected with 15% glycerol, mounted in a standard loop, and kept at 100 K during data collection. This data is also non-isomorphous with the room temperature data of "wet" crystals (data not shown). Two room temperature datasets were collected using a Rigaku MicroMax rotating anode (40 kV/30A), equipped with Osmic VariMax HF focusing mirrors. The crystals were wedged into 0.2 mm diameter glass capillaries filled entirely with storage solution. Using a Mar345 image detector, the exposure time was 25 and 30 s/per degree. The slits were set to 0.7x 0.7 $mm^2$ and 0.5 x 0.5 $mm^2$ and 80 frames were collected, respectively, for dataset "Rotating Anode 1" and "Rotating Anode 2" (See Table S1).

The lysozyme crystals used in this study are grown under non-standard conditions to yield many small crystals instead of a few very large ones. The crystallization conditions differ from the standard conditions in pH (3.0 instead of 4.5), salt concentration (20% instead of 5% NaCl) and the addition of 6% PEG 6000. The crystals are rod-shaped instead of cubic- or bipyramidal shaped as observed conventionally. Whereas it is very easy to grow very large lysozyme crystals using the standard conditions, this is not the case for the crystals we used in our study. To obtain very large crystals (0.2-0.25 mm diameter) while keeping the essential parameters the same as for the micron-sized crystals used at the LCLS, we used 20 mg/ml protein and 1 M NaAc pH 3.0, 9-10 % NaCl, 6 % PEG 6000 as precipitant. Subsequently, the crystals were equilibrated in 1 M NaAc pH 3.4, 10 % NaCl, the storage solution for the crystals used at the LCLS. The non-conventional crystallization and storage conditions may result in a higher propensity for non-isomorphism. In fact, the room temperature data do not scale to room temperature data of conventionally grown lysozyme crystals.

Fig. S4 shows a comparison of structure factor amplitudes between all collected SFX and conventional datasets. While differences are observed in the scaling of different datasets, no clear dose-dependent increase in R-factors is observed. The 40 fs and 5 fs datasets scale with each other to within about 10% R-factor and the rotating anode data scale very well with each other. The 40 fs dataset scales best to 400 kGy, and better to "Rotating Anode 2" than to 24 kGy. Scaling seems to depend on more than one parameter. Nevertheless, it seems that both FEL datasets scale somewhat better to the 400 kGy. This does not imply that they are as damaged as the 400 kGy dataset because the 400 kGy data scale reasonably well with the 200 kGy data andalso with the rotating anode datasets, which have a dose of 2-3 kGy. This suggests that scaling is not dominated by differences caused by radiation damage but other factors, likely non-isomorphism. All conventional datasets scale approximately the same way to the SFX data, with an R-factor of ~ 10% in

the medium resolution range which is significantly higher than the scaling between the conventional data. Interestingly, the two SFX data sets also scale with about 10%. It is unlikely that this is caused by non-convergence of the Monte Carlo integration because the same behavior is observed when merging the two SFX datasets (not shown). This implies that the SFX data differ from the other data, for example due to non-isomorphism or due to some systematic feature of the data. This could be caused *e.g.* by weak reflections, indexing inaccuracies, etc. Clearly, further work is needed to resolve these issues.

Evaluation of SFX data
Single shot data where a crystal was hit giving rise to SFX diffraction patterns on the CSPAD were identified as those with more than 10 peaks using the Cheetah software package *(23)*. Different Analog-to-Digital Units (ADU) thresholds of 900 ADU (100 photons) and 500 ADU (55 photons) were necessary for peak detection for the 40 fs and 5 fs data. Detector geometry refinement, indexing and integration were carried out using CrystFEL*(15)*, using the known unit cell of lysozyme in the software. Intensities determined from indexed diffraction patterns were integrated and averaged as described in *(24,25)* to yield the intensities used for structure determination. Analysis of subsets of the data revealed that integrated intensities converged when each reflection was observed roughly 300 times on average (Fig. S5). The electron density, the signal/noise ratio and the resolution dependence of the $R_{split}$ quality measure for Monte-Carlo integrated data *(15)* confirm the resolution to be 1.9 Å (Fig. S6).

Structure determination
A lysozyme model (1VDS) was refined to excellent R-factors against the XFEL data using alternating cycles of rebuilding and simulated annealing with PHENIX *(20)*. The resulting structure is in excellent agreement with that obtained using the synchrotron data presented here.

Molecular Replacement
The data were also phased by molecular replacement with PHASER *(19)* using turkey egg white lysozyme as the search model (pdb entry 1LJN) resulting in a single peak in the rotation function (Z=11.6). After the translation search (Z=26.9), the initial R-factor was 42%. The resulting electron density clearly shows the expected differences between search model and actual structure with different amino acids as illustrated in Fig. S3.
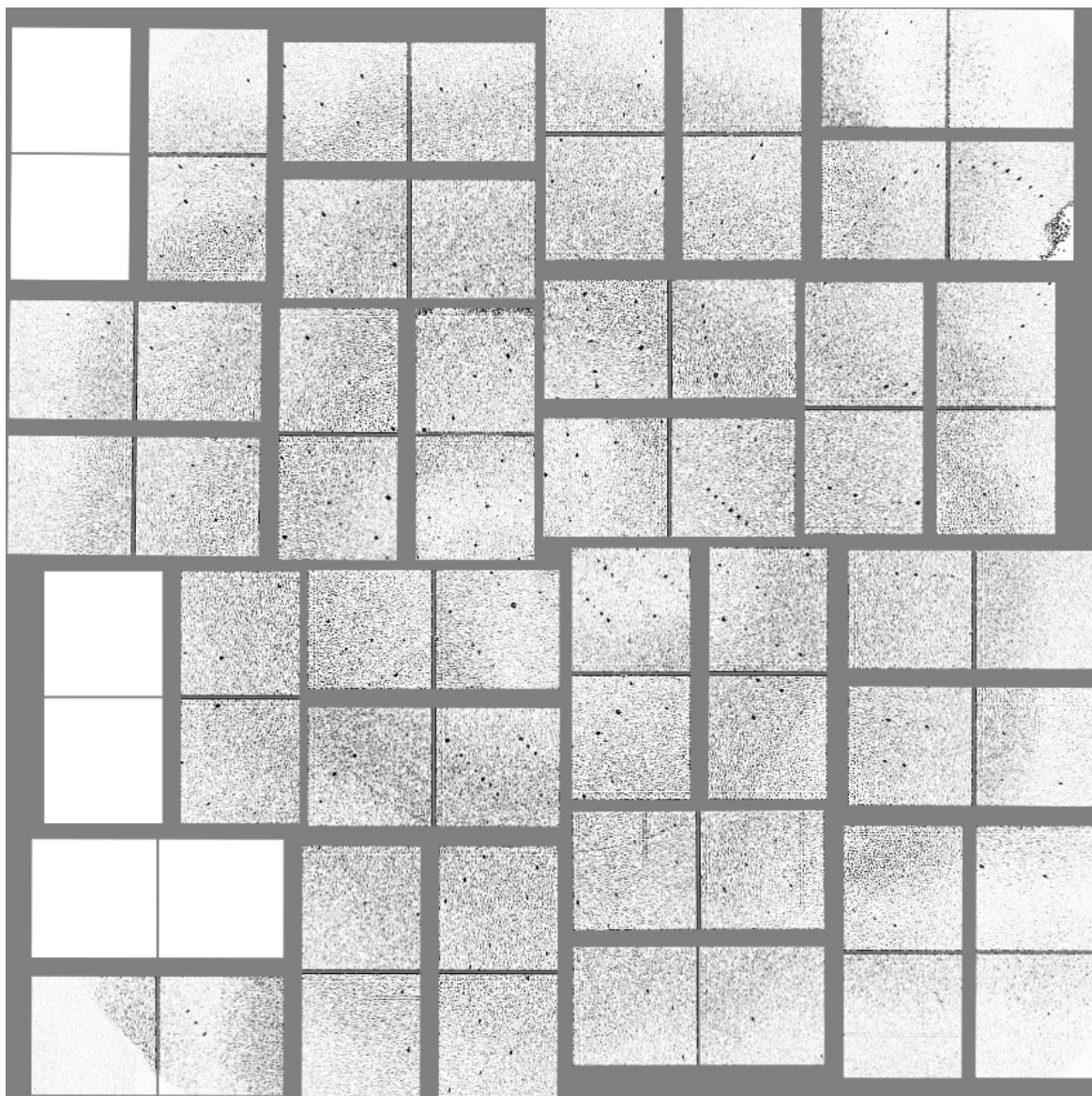
**Fig. S1.**
A typical diffraction pattern using a single 40 fs pulse showing Bragg peaks to the edge of the CSPAD detector. Note that some of the tiles of the CSPAD were not functional at the time of the measurement and appear completely white.

**Fig. S2**
Same diffraction pattern as in Fig. S1 after background subtraction, including subtraction of solvent scattering. The Bragg peaks are seen to clearly extend to the edge of the detector.
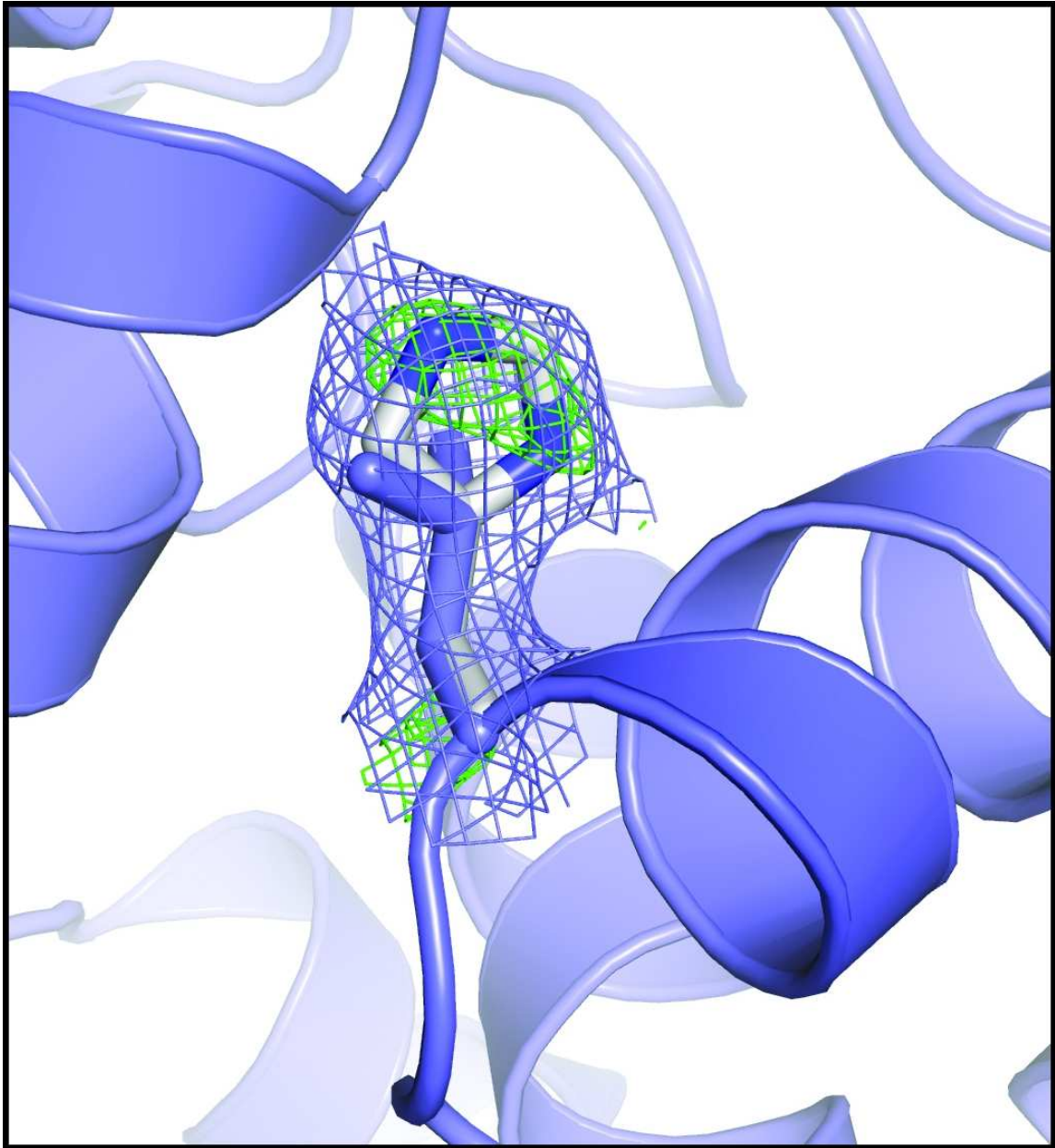
**Fig. S3**

Quality of the molecular replacement map. The region around Leu15 in the turkey egg white lysozyme model (blue carbon atoms) is shown. In the actual hen egg white lysozyme structure obtained of the presented SFX data, this is a histidine (white carbon atoms). Both the $2mF_{obs}$-$DF_{calc}$ (blue, 1.5 σ) and Fobs-DFcalc (green, +/- 3 σ) maps *(10)* clearly show the difference between the search model and the actual structure obtained from the SFX data.
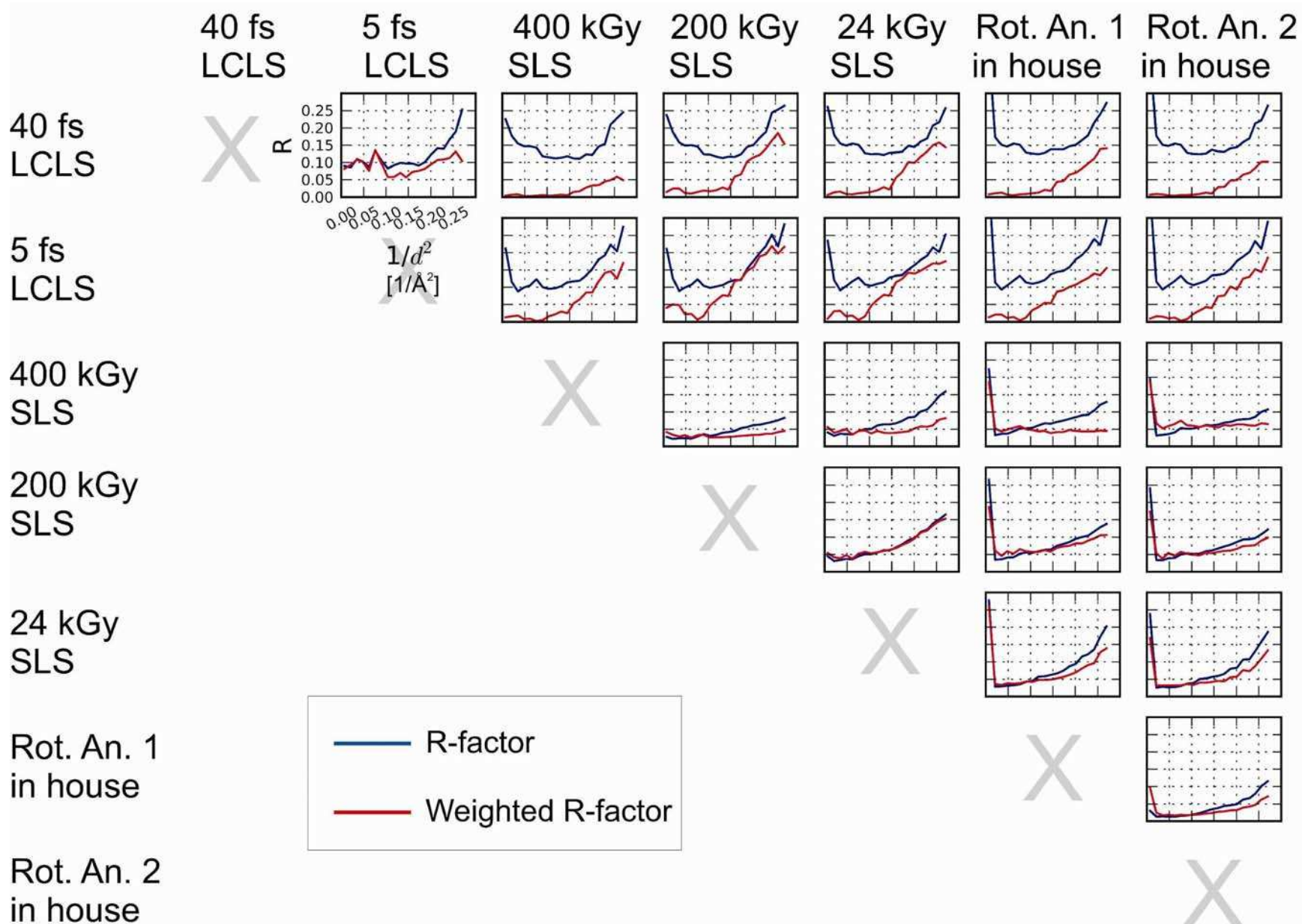
**Fig. S4** Comparison of structure factor amplitudes after scaling with SCALEIT (26) of the SFX data with room temperature datasets collected at the Swiss Light Source and at a rotating anode (see Table S1 for details). The unweighted (blue line) and variance-weighted R-factors (red line) are shown as a function of resolution.
The unweighted and weighted R-factors are calculated by SCALEIT as follows:

$$R = \frac{\sum |F_1 - F_2|}{\sum F_1}$$

$$R_{weighted} = \frac{\sum \left| \frac{F_1^2 - \langle F_1 F_2 \rangle}{\text{var}(F_1^2)} \right| + \left| \frac{F_2^2 - \langle F_1 F_2 \rangle}{\text{var}(F_2^2)} \right|}{\sum \left( \frac{F_1^2}{\text{var}(F_1^2)} \right) + \left( \frac{F_2^2}{\text{var}(F_2^2)} \right)}$$

$$\text{var}(F^2) \equiv \text{var}(F) \cdot 4F^2$$

$$\langle F_1 F_2 \rangle \equiv \frac{\left( \frac{F_1^2}{\text{var}(F_1^2)} \right) + \left( \frac{F_2^2}{\text{var}(F_2^2)} \right)}{\left( \frac{1}{\text{var}(F_1^2)} \right) + \left( \frac{1}{\text{var}(F_2^2)} \right)}$$

The systematically observed lower R-factors (blue) between conventional data sets compared to the R-factor between SFX and conventional data indicate systematic differences in the data. These differences are independent of the radiation damage induced during the measurement, as can be seen from the scaling to datasets with different radiation doses.
Moreover, the displayed R-weighted (red) has to be interpreted carefully for the SFX data. In contrast to conventional data, the errors of the intensities of the SFX data are related to the Monte Carlo integration method that was applied.
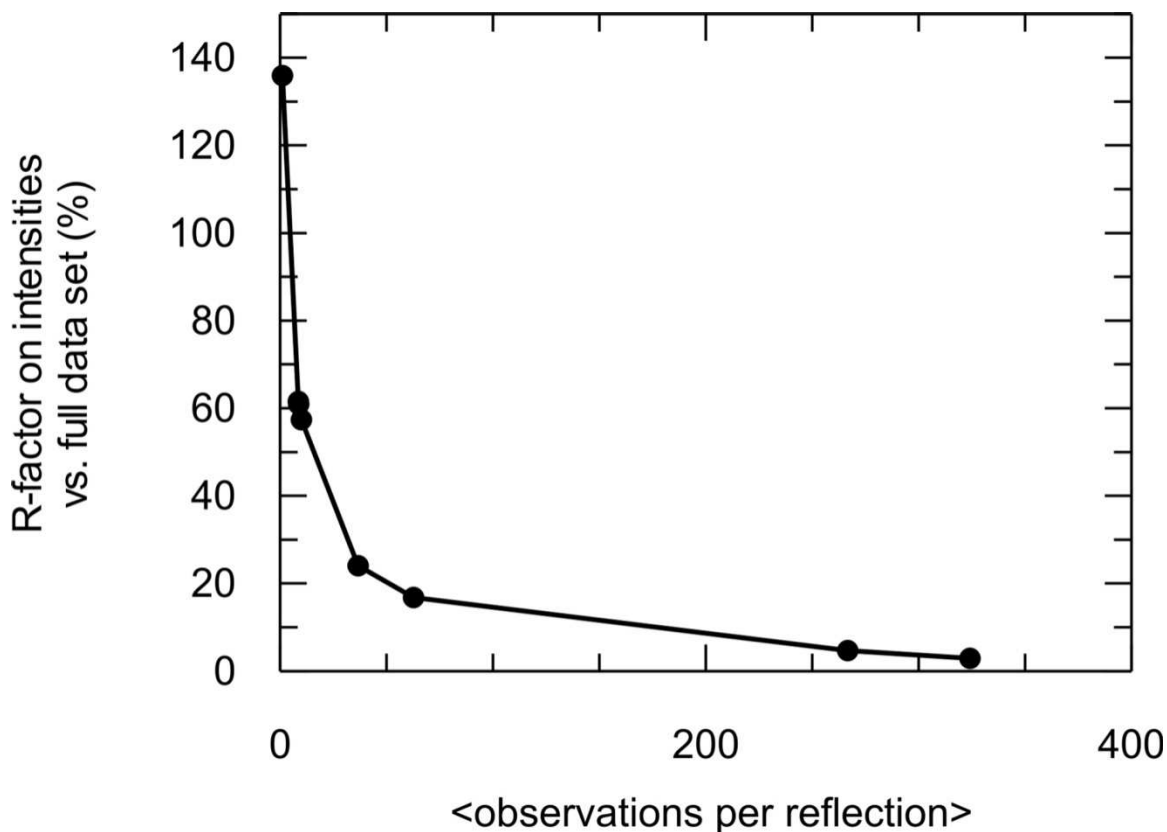
**Fig. S5**

Convergence of the integration of SFX intensities. The linear R-factor

$$R = \frac{\sum_{hkl} \left| I_{partial} - I_{final} \right|}{\frac{1}{2} \sum_{hkl} I_{partial} + I_{final}}$$

of partial 40 fs pulse length data sets versus that of the final complete data set is plotted against the average number of observations per Bragg reflection. After approximately 300 observations per reflection, the value of the R-factor does not change much anymore, suggesting this amount of observations is sufficient to produce reliable intensities. However, the R-factor is seen to still decrease slightly with more observations. The final 40 fs data set has on average 379 observations per reflection.
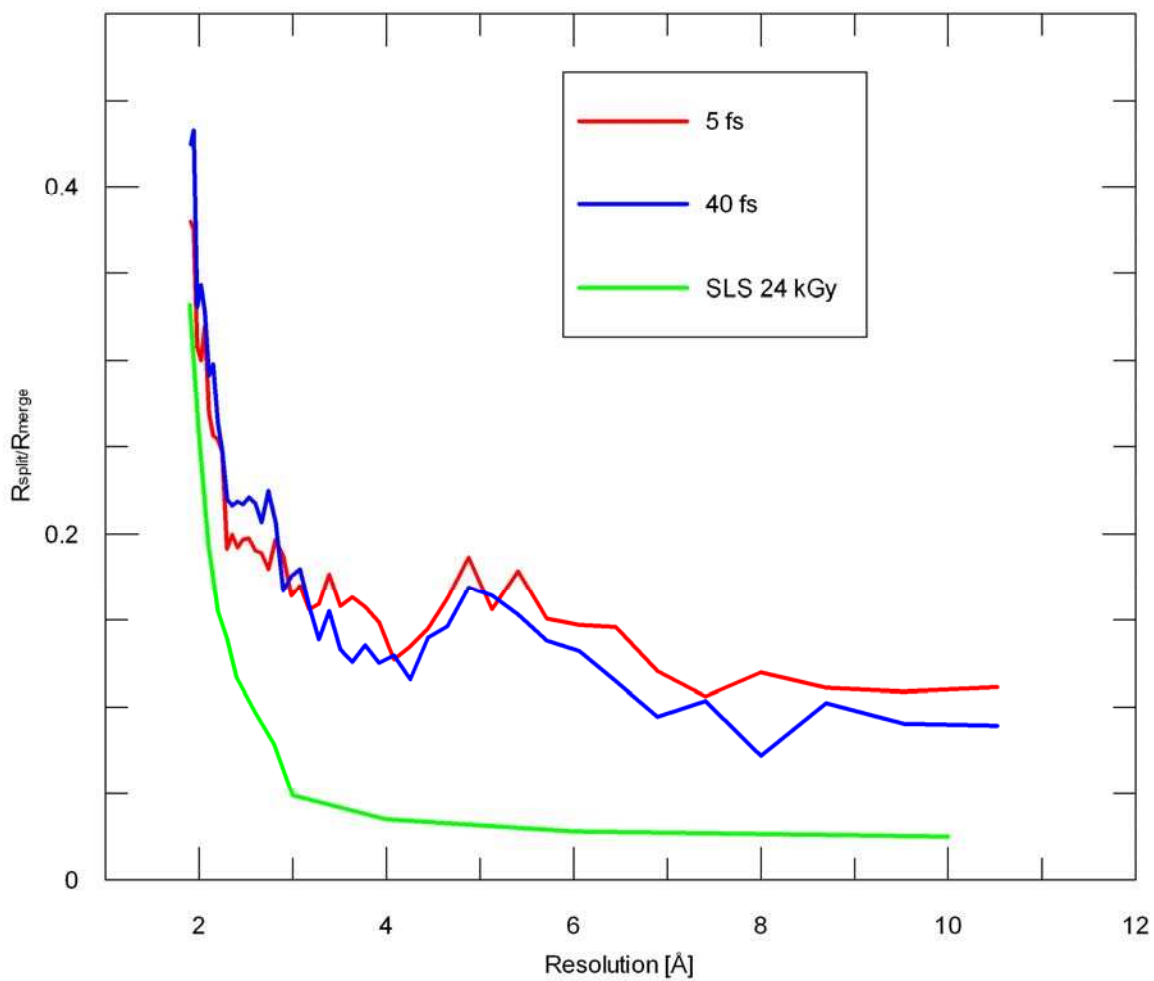
9

**Fig. S6**
Resolution dependence of the internal consistency measures $R_{split}/R_{merge}$. The steep increase at high resolution confirms the resolution of each dataset to be 1.9 Å.
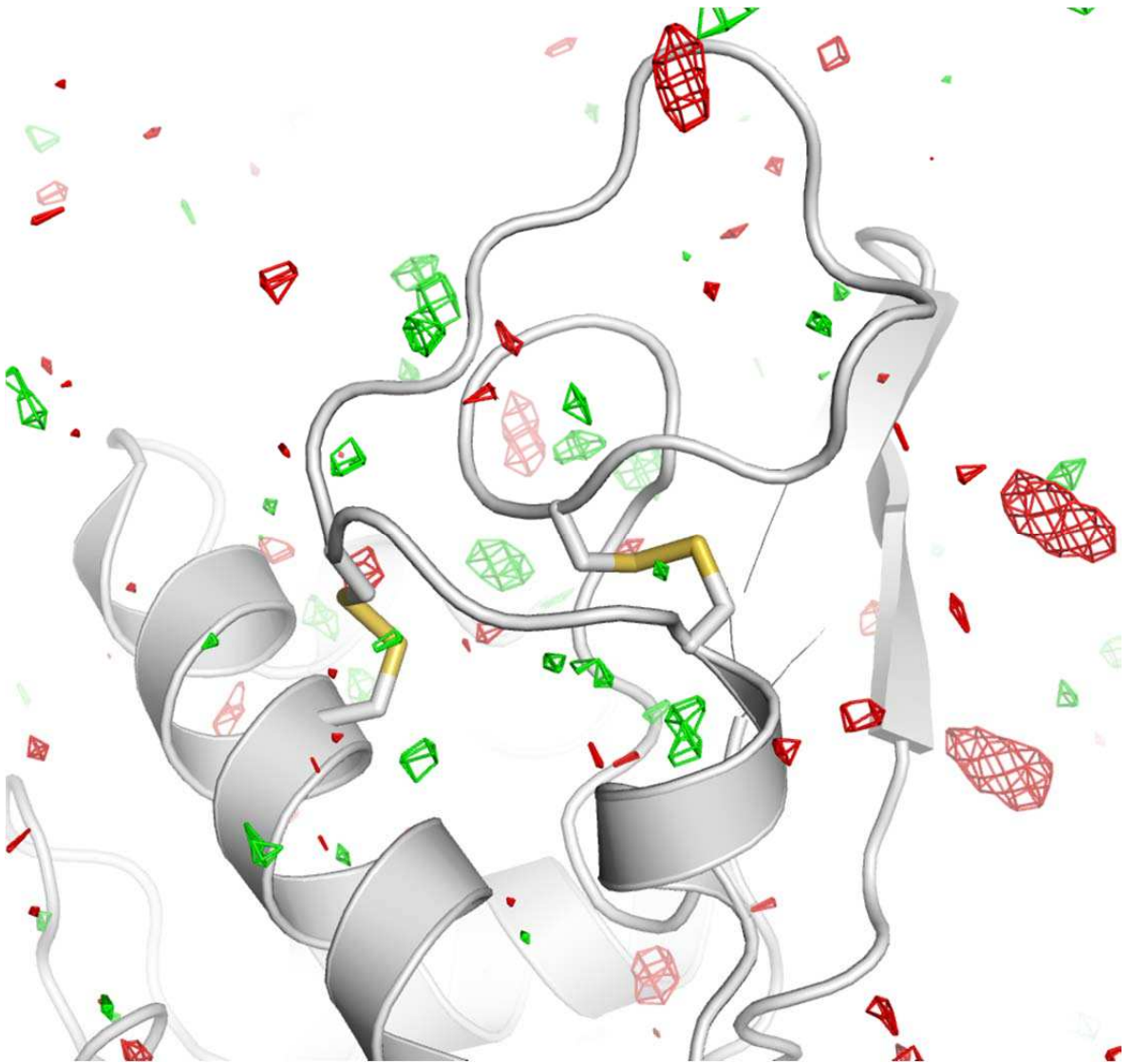
**Fig. S7**

$F_{obs}$[40 fs]-$F_{obs}$[5 fs] difference Fourier map, contoured at +3 σ (green) and -3 σ (red). No interpretable features are apparent.

## Table S1. Data and refinement statistics

| Parameter | 40 fs* pulses | 5 fs* pulses | SLS RT data 1 **** | SLS RT data 2 **** | SLS RT data 3 **** | Rotating anode RT dataset 1 | Rotating anode RT dataset 2 |
|---|---|---|---|---|---|---|---|
| Wavelength | 1.32 Å | 1.32 Å | 0.97860 Å | 0.97860 | 0.9997 Å | 1.5418 Å | 1.5418 Å |
| X-ray focus [μm$^2$] | ~ 10 | ~ 10 | ~ 50 x 50 | ~50 x 50 | ~ 100 x 100 | ~200 x 200 (FWHM) | ~ 200 x 200 (FWHM) |
| Pulse energy/fluence at sample | 600 μJ / 4x10$^{11}$ph/pulse | 53 μJ/3.5x10$^{10}$ ph/pulse | N.A./ 3.8 x10$^{10}$ph/s | N.A. 3.8 x 10$^{10}$ph/s | N.A./ 2.5 x10$^{10}$ph/s | ~10$^7$ph/s | ~10$^7$ph/s |
| Dose [MGy] | 33.0 per crystal | 2.9 per crystal | 0.4 total | 0.2 total | 0.024 total | 0.0026 | 0.0021 |
| Dose rate [Gy/s] | 8.3 x 10$^{20}$ | 5.8 x 10$^{20}$ | 5.3 x 10$^3$ | 5.7 x 10$^3$ | 9.6 x 10$^2$ | 1.1 | 1.1 |
| Space group | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ | $P4_32_12$ |
| Unit cell length [Å], $\alpha=\beta=\gamma=90°$ | $a=b=79$, $c=38$ | $a=b=79$, $c=38$ | $a=b=79.3$, $c=38.2$ | $a=b=79.3$, $c=38.1$ | $a=b=79.2$, $c=38.1$ | $a=b=79.2$, $c=38.1$ | $a=b=79.2$, $c=38.1$ |
| Oscillation range/exposure time | Still exp. / 40 fs* | Still exp. / 5 fs* | 0.25°, 0.25 s | 0.25°, 0.25 s | 1.0°, 0.25 s | 1.0°, 30 s | 1.0°, 25 s |
| # collected diffraction images | 1471615 | 1997712 | 300 | 350 | 100 | 80 | 80 |
| # of hits/indexed images | 66442 /12247 | 40115/10575 | n.a./300 | n.a./ 350 | n.a./100 | n.a./80 | n.a./80 |
| Number of reflections | n.a. | n.a. | 53131 | 62235 | 70960 | 61054 | 60945 |
| Number of unique reflections | 9921 | 9743 | 10022 | 10041 | 9297 | 9525 | 9698 |
| Resolution limits [Å] | 35.3-1.9 | 35.3-1.9 | 39.6-1.9 | 36-1.9 | 35.4-1.9 | 36-1.9 | 36-1.9 |
| Completeness** | 98.3% (96.6%) | 98.2% (91.2%) | 99.6% (99.7%) | 99.9% (99.9%) | 92.6% (95.1%) | 95.2% (92.5%) | 96.9% (94.4%) |
| I/σ(I)** | 7.4 (2.8) | 7.3 (3.1) | 33 (16) | 16.8 (6.8) | 18.24 (5.3) | 21.6 (5.9) | 27.9 (9.3) |
| R$_{split}$*** | 0.158 | 0.159 | n.a. | n.a. | n.a. | n.a. | n.a. |
| R$_{merge}$ | n.a. | n.a. | 0.032 (0.088) | 0.078 (0.251) | 0.075 (0.332) | 0.068 (0.343) | 0.048 (0.204) |
| Wilson B-factor***** | 28.3 Å$^2$ | 28.5 Å$^2$ | 25.2 Å$^2$ | 18.0 Å$^2$ | 19.4 Å$^2$ | 17.4 Å$^2$ | 17.8 Å$^2$ |
| R-factor/R-free****** | 0.196/0.229 | 0.189/0.227 | 0.160/0.180 | 0.165/0.190 | 0.166/0.200 | 0.161/0.192 | 0.160/0.195 |
| Rmsd bonds, Rmsd angles****** | 0.006 Å, 1.00° | 0.006 Å, 1.03° | 0.006 Å, 1.01° | 0.006 Å, 1.05° | 0.007 Å, 1.05° | 0.006 Å /1.02 ° | 0.007 Å /1.04 ° |
| PDB code | 4ET8 | 4ET9 | 4ETA | 4ETB | 4ETC | 4ETD | 4ETE |

* Electron bunch length    ** Highest resolution shell: 2.0-1.9 Å    *** R$_{split}$ as defined in *(14)*

$$R_{split} = \left( \frac{1}{\sqrt{2}} \right) \cdot \frac{\sum_{hkl} \left| I_{hkl}^{even} - I_{hkl}^{odd} \right|}{\frac{1}{2} \sum_{hkl} \left| I_{hkl}^{even} + I_{hkl}^{odd} \right|}$$

**** Statistics from XDS *(17)*    ***** Calculated with TRUNCATE *(23)*    ****** Calculated with PHENIX *(22)*

# References

1. J. M. Holton, K. A. Frankel, *Acta Crystallogr. D* **66**, 393 (2010).
2. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, *Nature* **406**, 752 (2000).
3. A. Barty *et al.*, *Nature Photonics* **6**, 35 (2012).
4. L. Lomb *et al.*, *Phys. Rev. B* **84**, 214111 (2011).
5. D. P. DePonte *et al.*, *J. Phys. D* **41**, 195505 (2008).
6. H. N. Chapman *et al.*, *Nature* **470**, 73 (2011).
7. L.C. Johansson *et al.*, *Nature Methods* **9**, 263 (2012).
8. R. Koopmann *et al.*, *Nature Methods* **9**, 259 (2012).
9. P. Emma *et al.*, *Nature Photonics* **4**, 641 (2010).
10. S. Boutet, G. J. Williams, *New Journal of Physics* **12**, 035024 (2010).
11. Materials and methods are available as supporting material on *Science* Online.
12. C. C. F. Blake *et al.*, *Nature* **206**, 757 (1965).
13. J. Wang, M. Dauter, R. Alkire, A. Joachimiak, Z. Dauter, *Acta Crystallogr. D* **63** 1254 (2007)
14. R. L.Owen, E. Rudino-Pinera, E. F. Garman, *Proc. Natl. Acad. Sci.* **103**, 4912-4917 (2006)
15. T. A. White *et al.*, *J. Appl. Crystallogr.* **45**, 335-341 (2012).
16. R. J. Read, *Acta Crystallogr. A* **42**, 140 (1986).
17. R.J. Southworth-Davies, M.A. Medina, I. Carmichael, E.F. Garman, *Structure* **15**, 1531–1541 (2007)
18. W. Kabsch, *J. Appl. Chrystallogr.* **26** 795-806 (1993)
19. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr. D* **61**, 458 (2005).
20. P. D. Adams *et al.*, *Acta Crystallogr. D* **66**, 213 (2010).
21. L. Young *et al.*, *Nature* **466**, 56 (2010).
22. H. T. Philipp, M. Hromalik, M. Tate, L. Koerner, S. M. Gruner, *Nucl Instrum. Methods A* **649**, 67 (2011).
23. The cheetah software package is available via the url: http://www.desy.de/~barty/cheetah/
24. R. A. Kirian *et al.*, *Optics Express* **18**, 5713 (2010).
25. R. A. Kirian *et al.*, *Acta Crystallogr. A* **67**, 131 (2011).
26. S. Bailey, *Acta Crystallogr. D* **50** 760-763 (1994)