

# Evaluation of serial crystallographic structure determination within megahertz pulse trains

Cite as: Struct. Dyn. 6, 064702 (2019); doi: 10.1063/1.5124387

Submitted: 14 August 2019 · Accepted: 21 October 2019 ·

Published Online: 4 December 2019



View Online



Export Citation



CrossMark

Oleksandr Yefanov,<sup>1</sup> Dominik Oberthür,<sup>1</sup> Richard Bean,<sup>2</sup> Max O. Wiedorn,<sup>1,3</sup> Juraj Knoska,<sup>1</sup> Gisel Pena,<sup>1</sup> Salah Awel,<sup>1</sup> Lars Gumprecht,<sup>1</sup> Martin Domaracky,<sup>1</sup> Iosifina Sarrou,<sup>1</sup> P. Lourdu Xavier,<sup>1</sup> Markus Metz,<sup>1</sup> Saša Bajt,<sup>4</sup> Valerio Mariani,<sup>1</sup> Yaroslav Gevorkov,<sup>1,5</sup> Thomas A. White,<sup>1</sup> Aleksandra Tolstikova,<sup>1</sup> Pablo Villanueva-Perez,<sup>1</sup> Carolin Seuring,<sup>1</sup> Steve Aplin,<sup>1</sup> Armando D. Estillore,<sup>1</sup> Jochen Küpper,<sup>1,3,6</sup> Alexander Klyuev,<sup>4</sup> Manuela Kuhn,<sup>4</sup> Torsten Laurus,<sup>4</sup> Heinz Graafsma,<sup>4</sup> Diana C. F. Monteiro,<sup>3,6</sup> Martin Trebbin,<sup>7,8</sup> Filipe R. N. C. Maia,<sup>9,10</sup> Francisco Cruz-Mazo,<sup>11</sup> Alfonso M. Gañán-Calvo,<sup>11</sup> Michael Heymann,<sup>12</sup> Connie Darmanin,<sup>13</sup> Brian Abbey,<sup>13</sup> Marius Schmidt,<sup>14</sup> Petra Fromme,<sup>15</sup> Klaus Giewekemeyer,<sup>2</sup> Marcin Sikorski,<sup>2</sup> Rita Graceffa,<sup>2</sup> Patrik Vagovic,<sup>2</sup> Thomas Kluyver,<sup>2</sup> Martin Bergemann,<sup>2,16</sup> Hans Fangohr,<sup>2</sup> Jolanta Sztuk-Dambietz,<sup>2</sup> Steffen Hauf,<sup>2</sup> Natascha Raab,<sup>2</sup> Valerii Bondar,<sup>2</sup> Adrian P. Mancuso,<sup>2,16</sup> Henry Chapman,<sup>1,3,6</sup> and Anton Barty<sup>1</sup>

## AFFILIATIONS

<sup>1</sup>Center for Free-Electron Laser Science, Deutsches Elektronen Synchrotron, Notkestrasse 85, 22607 Hamburg, Germany

<sup>2</sup>European XFEL, Holzkoppel 4, 22869 Schenefeld, Germany

<sup>3</sup>Center for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

<sup>4</sup>Deutsches Elektronen Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany

<sup>5</sup>Institute of Vision Systems, Hamburg University of Technology, Vision Systems E-2, Harburger Schloßstr. 20, 21079 Hamburg, Germany

<sup>6</sup>Department of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

<sup>7</sup>Department of Chemistry, University at Buffalo, 359 Natural Sciences Complex, Buffalo, New York 14260, USA

<sup>8</sup>Hauptman-Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, New York 14203, USA

<sup>9</sup>Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-75124 Uppsala, Sweden

<sup>10</sup>NERSC, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

<sup>11</sup>Dept. de Ingeniería Aeroespacial y Mecánica de Fluidos, ETSI, Universidad de Sevilla, 41092 Sevilla, Spain

<sup>12</sup>Intelligent Biointegrative Systems Group, Institute of Biomaterials and Biomolecular Systems, University of Stuttgart, D-70569 Stuttgart, Germany

<sup>13</sup>ARC Centre of Excellence in Advanced Molecular Imaging, La Trobe Institute for Molecular Sciences, La Trobe University, Victoria 3086, Australia

<sup>14</sup>Physics Department, University of Wisconsin-Milwaukee, 3135 N. Maryland Ave, Milwaukee, Wisconsin 53211, USA

<sup>15</sup>School of Molecular Sciences and Biodesign Center for Applied Structural Discovery, Arizona State University, Tempe, Arizona 85287-1604, USA

<sup>16</sup>Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia

## ABSTRACT

The new European X-ray Free-Electron Laser (European XFEL) is the first X-ray free-electron laser capable of delivering intense X-ray pulses with a megahertz interpulse spacing in a wavelength range suitable for atomic resolution structure determination. An outstanding but crucial question is whether the use of a pulse repetition rate nearly four orders of magnitude higher than previously possible results in unwanted structural changes due to either radiation damage or systematic effects on data quality. Here, separate structures from the first and subsequent pulses in the European XFEL pulse train were determined, showing that there is essentially no difference between structures determined from different pulses under currently available operating conditions at the European XFEL.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5124387>

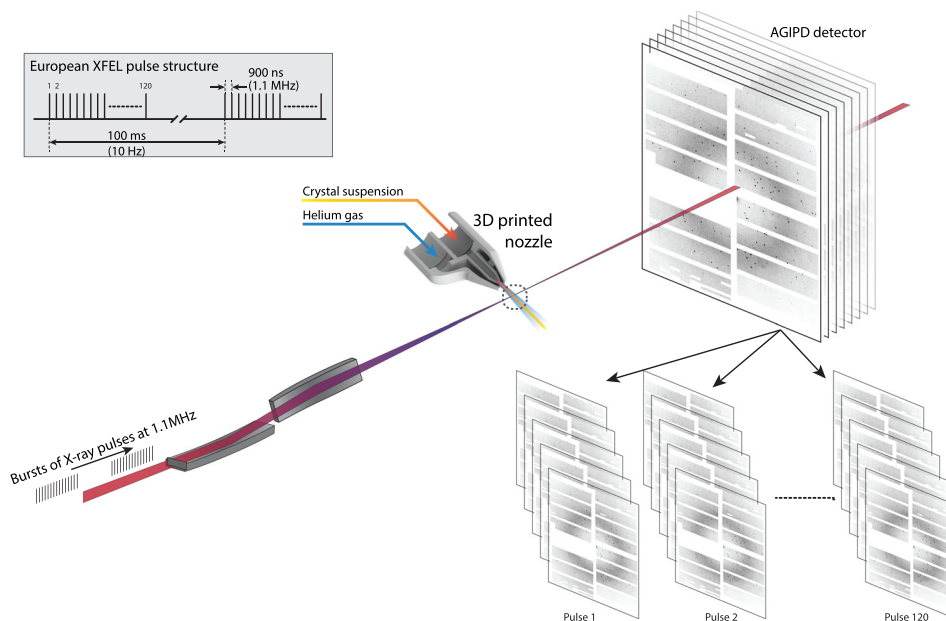
## INTRODUCTION

The development of serial femtosecond crystallography (SFX) using intense femtosecond-duration pulses from X-ray free-electron lasers has opened up new avenues for the measurement of macromolecular structures and macromolecular dynamics. Particular applications to date have been room-temperature measurements using micrometer-sized and smaller protein crystals, damage free determination of radiation sensitive structures, and time-resolved studies of biomolecular dynamics at physiologically relevant temperatures.<sup>1–8</sup> The recently opened European X-ray Free-Electron Laser (European XFEL) free electron laser is the first facility capable of delivering millijoule X-ray pulses of femtosecond duration at a megahertz pulse repetition rate in the wavelength range useful for atomic resolution structure determination.<sup>9</sup> Previous facilities were limited to a pulse repetition rate of 120 Hz by available accelerator technology. The megahertz pulse repetition rate of the European XFEL is particularly attractive because it enables the efficient measurement of the data volumes necessary for high resolution time-resolved structural studies in a comparatively short amount of data collection time. Measuring up to 3520 frames/second using currently available detector technology could enable the measurement of SFX data at almost 30 times the rate that has been possible to date, provided all pulses at megahertz pulse rates can be exploited for reliable structure determination.

The challenge in exploiting rapid data acquisition at the European XFEL lies in exploiting the 1.1 MHz repetition rate within a pulse train. The European XFEL does not deliver a uniform stream of X-ray pulses, rather it delivers 10 bursts (trains) of X-ray pulses/second with each burst containing up to 2700 X-ray pulses at an interpulse repetition rate of up to 4.5 MHz (1.1 MHz in this experiment), Fig. 1. The megahertz intrabunch repetition rate of the European XFEL poses unique challenges for SFX sample delivery and time resolved studies because the fresh sample must be delivered to the interaction point before the arrival of the next X-ray pulse and should not be pre-

exposed or damaged on its way to the peak intensity at the focus. For time resolved studies in particular, it is important to know that delivering pulses separated by only 900 ns, nearly four orders of magnitude less time between pulses than the 8.3 ms pulse separation in experiments to date using 120 Hz pulse repetition rates, does not lead to a degradation in data quality or induce artificial structural changes. It is well known that, for example, the high dose deposited in a single focused XFEL pulse causes rapid local heating of the sample and surrounding matter to hundreds of thousands of degrees.<sup>13</sup> This in turn can cause the jet to explode sending a pressure wave back up the liquid stream, creating a void that must be replenished in order to transport a fresh volume of liquid with crystals to the beam focus by the time the next pulse arrives. Photoionization in the interaction region creates high-energy electrons near the interaction region, which might be ejected from the jet. Modeling shows such electrons likely return to the (now-charged) jet column and then create a cascade of charges through collisions with atoms in the fluid.<sup>10–12</sup> Solvated electrons may further diffuse through the liquid, causing changes to the oncoming sample through radiolysis. The jet may be further disrupted after the X-ray pulse has passed. X-ray induced explosion of the jet was dramatically visualized in experiments performed by Stan *et al.* at the Linac Coherent Light Source (LCLS), which imaged the effect of LCLS X-ray illumination on relatively slow and thick liquid jets of  $\sim 20 \mu\text{m}$  diameter.<sup>13</sup> This heating and pressure wave,<sup>14</sup> and energetic products from the ionization and vaporization, may influence the structure of protein crystals still upstream of the interaction region. The effects of any pressure wave are potentially more severe at the European XFEL due to the short time duration before the arrival of the next pulse, resulting in a smaller spacing between damaged and fresh samples at current sample delivery speeds of up to 100 m/s. Therefore, it is difficult to extrapolate jet damage studies from previous experiments performed at LCLS at 120 Hz rates.

Here, we address the question of whether the use of megahertz repetition rate X-ray pulses results in any discernible or systematic



**FIG. 1.** Serial crystallography to obtain a structure on each pulse of the EuXFEL pulse train.

change in the solved macromolecule structure by determining independent structures for data accumulated from each pulse number in the European XFEL pulse train for lysozyme. The first SFX early experiments using 1.1 MHz pulse trains merged data from all pulses in the pulse train to determine a structure.<sup>15,16</sup> Here, we concentrate on whether there are any discernible differences in structures or data quality from individual pulses in the train to determine whether the high repetition rate yields any observed structural differences or degradation in data quality. A similar experimental arrangement to that in Wiedorn *et al.*<sup>15</sup> was used, however, with 120 X-ray pulses per train, reflective optics providing higher power density due to both a smaller focal spot and higher optical throughput at an average XFEL pulse energy of 1 mJ per pulse. Sample delivery was performed using liquid jets with a speed of 100 m/s and a diameter of 3  $\mu\text{m}$  as had previously been reported.<sup>15</sup> The use of 120 X-ray pulses per train and a pulse train repetition rate of 10 Hz for this experiment results in a data acquisition rate of 1200 frames/second, 10 times higher than previously available at the LCLS. Over  $1.4 \times 10^6$  crystal lattices were measured from the well-known reference test system lysozyme, enabling us to divide the data into separate datasets for each pulse number, Fig. 1, and to determine separate structures for each pulse number in the pulse train. At current experimental conditions, no significant or systematically interpretable difference in data or reconstructed structures for each pulse in the pulse train was found. Since the structures from each pulse are identical, it is possible to merge data from all 120 pulses together, at which point we demonstrate the collection of sufficient data for structure determination in approximately one minute using the currently available 1200 pulses/second.

## MATERIALS AND METHODS

The experiment was conducted at the SPB/SFX (Single Particles, Clusters, and Biomolecules and Serial Femtosecond Crystallography) instrument of the European XFEL<sup>17</sup> in September 2018 during experiment number p002120 using a similar instrument configuration as described for previous experiments.<sup>15</sup> The main difference in the instrument configuration was the use of reflective optics in Kirkpatrick-Baez like geometry for X-ray focusing instead of compound refractive lenses, resulting in a higher beamline transmission and a smaller focal spot,<sup>18</sup> Fig. 1. X-ray pulses with a mean photon energy of 9.35 keV (1.33  $\text{\AA}$  wavelength), a mean pulse energy of 1.04 mJ, and pulse length of approximately 15  $\mu\text{m}$  (50 fs duration derived from the electron bunch length) were focused by reflective Kirkpatrick-Baez optics in a 4-bounce configuration located 20 m upstream of the interaction region<sup>18</sup> into a focal spot of  $2 \times 3 \mu\text{m}^2$  full width at half maximum (FWHM) at the SPB/SFX interaction region. The European XFEL pulse structure for this experiment comprised 120 X-ray pulses at a 1.1 MHz repetition rate repeating at 10 Hz for a total of 1200 pulses/second. This pulse structure provides a factor of 10 increase in measured pulses/second than available at other currently operational hard X-ray XFELs. Diffraction from each X-ray pulse was measured using a 1 megapixel Adaptive Gain Integrating Pixel Detector (AGIPD)<sup>19</sup> located 118 mm downstream of the interaction region giving 1.9  $\text{\AA}$  resolution at the edge of the detector.

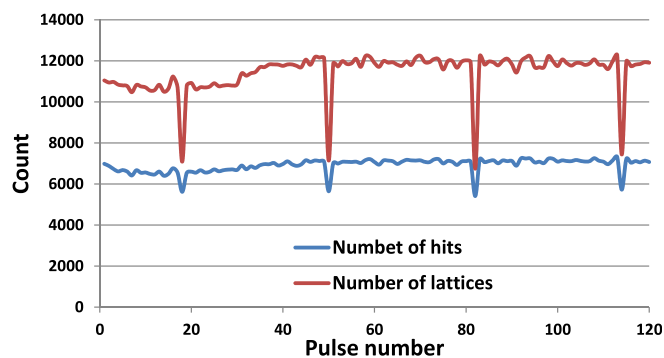
Microcrystals of hen egg white lysozyme (HEWL) of less than  $5 \times 5 \times 5 \mu\text{m}^3$  size were introduced into the X-ray interaction region in a liquid jet of 2–3  $\mu\text{m}$  diameter at speeds of 100 m/s produced by a flow focusing nozzle under similar operating conditions

as described in Wiedorn *et al.*<sup>15,20</sup> We had previously established that jets of this speed recover in time for the next X-ray pulse at megahertz repetition rates<sup>15</sup> and aimed for similar operating conditions in this experiment. Data were collected at room temperature, and the average dose for each crystal was estimated to be 14 MGy using RADDOSE-3D version 2.1.<sup>21</sup> The aim was to collect enough data to obtain a structure of lysozyme from each pulse in the train.

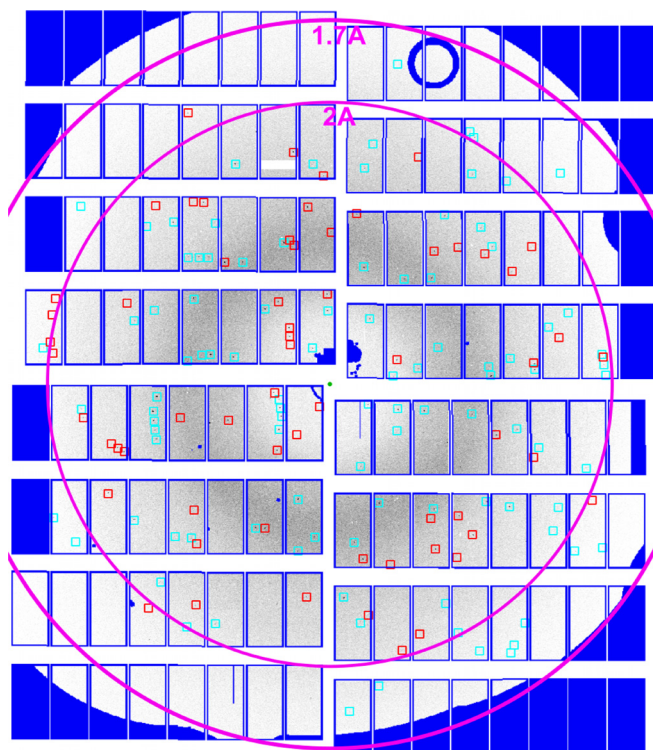
## RESULTS AND DISCUSSION

A total number of 16 654 608 data frames were collected from HEWL crystal solution at 1200 pulses/second and a peak pulse repetition rate of 1.1 MHz, of which 830 514 images (5%) were found to contain diffraction from a crystal as identified by Cheetah,<sup>22</sup> which was also used to apply the AGIPD calibration. Frames containing crystal diffraction patterns were distributed roughly evenly throughout the pulse train, Fig. 2. This is consistent with the results from previous experiments using 15 pulses per train and liquid jet speeds of 100 m/s<sup>15</sup>, demonstrating the same trend for 120 pulses per train. We note the decreased hit fraction for pulse numbers 18, 50, 82, and 114, which arise from a systematic detector artifact in these AGIPD memory cells. Some deviation from the expected straight line is explained in Methods data analysis section.

From the identified hits, 91% (782 500) could be indexed using the CrystFEL software suite,<sup>23,24</sup> yielding 1 374 785 indexed crystal lattices for structure determination when allowing multiple lattices per image (“–multi” option in CrystFEL). The presence of indexed multiple lattices in some diffraction patterns was verified by inspection of predicted peak locations for multiple lattices in frames identified by CrystFEL as containing multiple lattices. For example, Fig. 3 shows a sample diffraction pattern containing multiple lattices where spot prediction matches observed Bragg peak locations. The presence of multiple hits could be either due to random clustering of crystals in suspension or due to crystals adhering together, either of which could produce multiple lattices in a single frame consistent with the indexing results.



**FIG. 2.** Number of hits as a function of pulse ID across all data. The hit fraction as a function of pulse number indicates that crystals are hit randomly on any pulse within the megahertz European XFEL pulse train and not only on the first pulse in the pulse train. The decreased hit fraction for pulse numbers 18, 50, 82, and 114 appears to arise from a systematic detector artifact in certain AGIPD memory cells for a portion of the data.



**FIG. 3.** Multiple lattices measured in a single pulse. A single indexed diffraction pattern from HEWL shows multiple lattices indexed by XGANDALF, and red and blue squares represent predicted Bragg peaks for two crystal lattices. On average 1.6 lattices per frame containing crystal diffraction pattern was determined. Dark blue is the masked region of the detector.

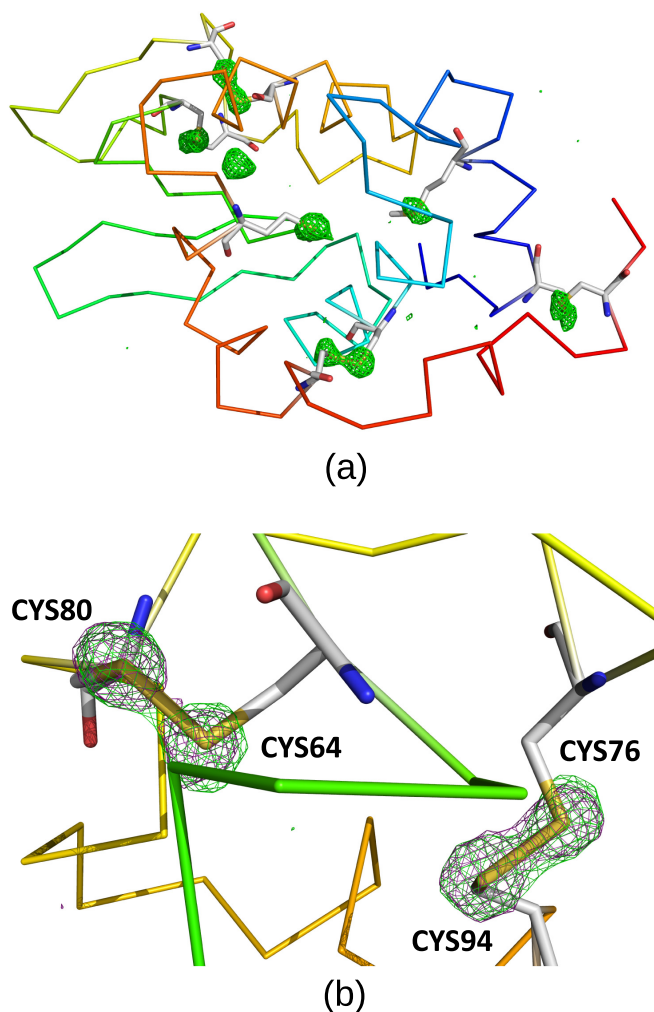
Merging all reflection intensities using the program “partialator” from CrystFEL in point group 4/mmm produced a dataset with an error metric  $R_{\text{split}}$  of 1.8 and  $CC^*$  of 0.9999 to 1.6 Å resolution. Diffraction at high resolution was limited by the exit aperture of the differential pumping catcher (see dark blue mask in Fig. 3), suggesting that a higher resolution could be achieved in the future using either shorter wavelength X-rays or adjustment of the SPB/SFX sample delivery hardware to accommodate a larger clear solid angle toward the downstream detector.

The structure from each pulse was determined by molecular replacement using Phaser within “Phenix”<sup>25</sup> software suite and structural model of lysozyme from Wiedorn *et al.*<sup>15</sup> (PDB<sup>26</sup> accession code 6FTR). Refinement with phenix.refine yields in a structural model with  $R_{\text{work}}/R_{\text{free}}$  of 14.2/16.0 to 1.6 Å resolution, Table I. Merging data in a noncentrosymmetric point group shows sufficient anomalous signals to enable *ab initio* automatic structure determination by native Sulfur SAD phasing (CFOM of 39.1 in SHELXD<sup>27</sup> using a 1.9 Å resolution cutoff). Automatic model building and refinement using the CRANK2-pipeline in CCP4<sup>28</sup> software suite followed by automatic model building and refinement using AutoBuild<sup>29</sup> in Phenix yielded a structural model with  $R_{\text{work}}/R_{\text{free}}$  of 15.4/17.3 using a resolution cutoff in a final autobuilding of 1.6 Å. The anomalous difference density map is shown in Fig. 4

**TABLE I.** SFX data and refinement statistics for combined data.

Parameter	Lysozyme (all)	
	4/mmm	422
Merged into point group	4/mmm	422
Photon energy (mean value) (eV)	9300	
X-ray focus, FWHM ( $\mu\text{m}$ )	$2 \times 3$	
Pulse energy at sample (70% beamline transmission) ( $\mu\text{J}$ )	800	
Pulse length (fs)	<40	
Space group	P 4 <sub>3</sub> 2 <sub>1</sub> 2	
Unit cell		
a, b, c (Å)	79.6, 79.6, 38.3	
$\alpha, \beta, \gamma$ (°)	90, 90, 90	
Number of hits/indexed lattices	782 500/1 374 785	
Number of unique reflections	16 454	30 336
Resolution range (Å)	15.66 – 1.6 (1.68 – 1.6)	15.37 – 1.6 (1.68 – 1.6)
Completeness (%)	99.9 (99.7)	99.8 (98.7)
Multiplicity/redundancy	9827.1 (64.9)	5327.5 (37.3)
$R_{\text{split}}$ (%)	1.80 (39.11)	2.35 (45.55)
$I/\sigma(I)$	41.5 (4.0)	30.7 (3.1)
$CC_{1/2}$	0.99970 (0.82438)	0.99953 (0.76752)
$CC^*$	0.99992 (0.95065)	0.99988 (0.93192)
$CC_{\text{ano}}$	...	0.24617 (0.096)
Wilson B-factor (Å <sup>2</sup> )	19.48	19.45
$R_{\text{work}}$	0.142 (0.239)	0.154 (0.284)
$R_{\text{free}}$	0.160 (0.258)	0.175 (0.292)
RMSD bonds (Å) / RMSD angles (deg)	0.009 / 1.30	0.010 / 1.53
Ramachandran favored	99.21	99.21
Ramachandran allowed	0.79	0.79
Ramachandran outliers	0	0
Average B-factor	25.44	25.83
Macromolecules	24.16	24.88
Ligands	49.94	...
Solvent	34.47	37.63
CXIDB data deposition	CXIDB ID-98	

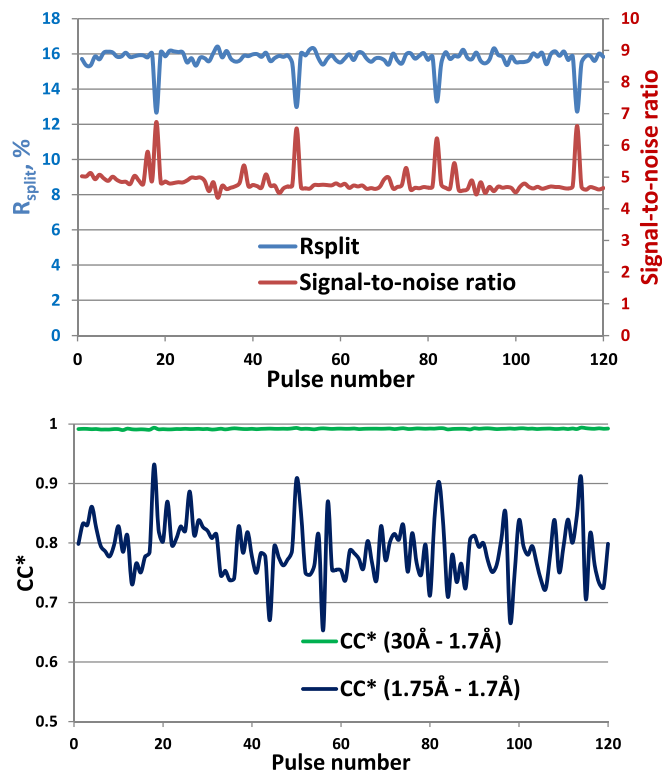
The data were then divided into separate datasets for each pulse in the pulse train to assess whether there are any systematic pulse train effects on structure quality. Merging data from each pulse separately, the data quality metrics  $R_{\text{split}}$ ,  $CC^*$ , and  $I/\sigma(I)$  show no meaningful degradation in data quality within the pulse train, Fig. 5. With almost 1.4M indexed lattices approximately evenly distributed through the pulses in the train, we obtain indexed lattices between 11 000 and 12 000 for each pulse, sufficient to obtain a structure for each pulse. The high resolution cutoff for each of the 120 datasets was set to 1.7 Å—at this resolution, all datasets are still reliable (see  $CC^*$  for the highest shell in Fig. 5). Table 3 in the Supplemental Material section shows the results of applying the same merging and refinement procedure described above to data from each individual pulse in the pulse



**FIG. 4.** Phased anomalous Fourier difference map contoured at 3.5 sigma at 9.3 keV using all data to 1.6 Å resolution. (a) Whole protein and (b) zoom in on selected residues. Green mesh is the anomalous difference map for whole data and magenta—for a half the dataset. Extra green density in (a) is probably  $\text{Cl}^-$  ions. Residues in (b): CYS80-CYS64 and CYS94-CYS76.

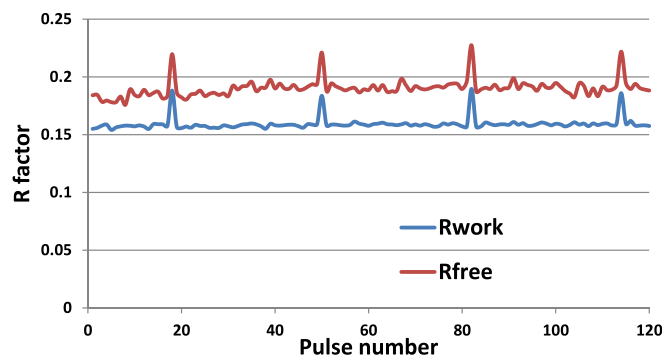
train. The data in Table 3 show that refinement statistics are consistent for all pulses in the pulse train. Figure 6 plots the  $R_{\text{work}}$  and  $R_{\text{free}}$  metrics as a function of pulse ID, which shows that the values are very consistent through the pulse train except for pulses 18, 50, 82, and 114. Excluding these pulses that contain artifacts from bad memory cells from the all-pulse dataset had no discernible impact on the structure nor were the all-pulse quality metrics degraded. Inspection of structures across pulses shows no statistically significant differences, even between the first and subsequent pulses (see the Methods structure determination section). For example, Fig. 7 shows structure and difference maps determined from the first and second pulses in the pulse train compared to the structure determined from merging all pulses together.

During one 30 min period of data collection, we were able to collect over 2 million detector frames of which 190 000 images

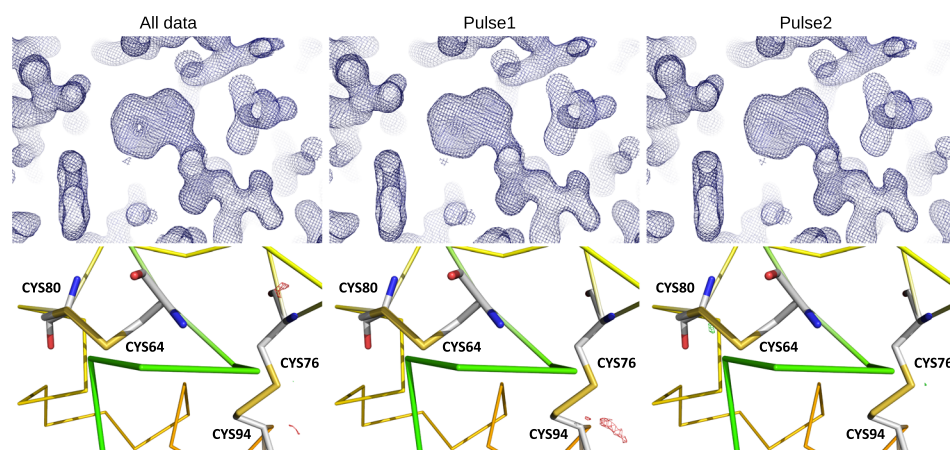


**FIG. 5.** Data quality metrics  $R_{\text{split}}$  and signal-to-noise ratio (left) and  $\text{CC}^*$  (right) as a function of pulse number in the train.

contained crystal diffraction patterns, yielding 332 482 indexed crystal lattices (runs 96–98, one run is 10 min of data collection). This raises the prospect of elucidating small changes in structure factors by the accumulation of statistics in orders of magnitude less time than previously possible, for example, in time resolved experiments for producing molecular movies, for the determination of structures from weakly diffracting crystals<sup>7</sup> in which averaging of large datasets is required to obtain a sufficient signal-to-noise ratio (SNR) for structure determination. For example, integrating data across all 120 pulses in the pulse train, a SNR of 15 was obtained from a 10 min data run containing 117 415 indexable



**FIG. 6.**  $R_{\text{free}}$  and  $R_{\text{work}}$  as a function of pulse number in the train.



**FIG. 7.** Comparison of the structure from all data to structures from the first and second pulses in the pulse train. The 2Fo-Fc map contoured at 1.5 sigma (top row) and part of the structure with Fo-Fc difference maps contoured at 3 sigma (bottom row) determined from data merged from all pulses (first column) and maps determined from data only for the first and second pulses in the pulse train (second and third columns, respectively). Structures for all pulses in the pulse train are included in the data deposition.

lattices. This increased to a SNR of 25 for a 30 min run containing 332 482 lattices and a SNR of 48 for the whole dataset consisting of 1 374 785 indexable lattices. Data metrics for different point groups as well as different resolution cutoffs and smaller portions of the data are summarized in Table II.

There is no strict requirement on the number of indexed patterns needed for successful structure determination using molecular replacement. For example, the beta-lactamase structure previously determined at European XFEL<sup>15</sup> was solved using 14 000 indexed lattices. Under current operating conditions, sufficient data could be collected in less than two minutes at an ~10% hit fraction using less than 60  $\mu$ l of crystal solution enabling substrate screening using mix-and-inject SFX as previously noted.<sup>15</sup> Naturally, data quality improves through improved statistics as more data are measured. For example, using only 10 min of data collection (run 97),  $R_{\text{work}}/R_{\text{free}}$  of 0.1399/0.1640 were obtained from 117 415 crystals in 67 090 diffraction patterns, compared to  $R_{\text{work}}/R_{\text{free}}$  of 0.1368/0.1587 from 332 482 crystals in 189 960 diffraction patterns from 30 min of data collection and  $R_{\text{work}}/R_{\text{free}}$  of 0.1347/0.1570 from 1 374 785 crystals in 830 514 diffraction patterns from the whole dataset to 1.7 Å resolution, Table II.

## CONCLUSIONS

Structures determined from each pulse number in the European XFEL pulse train result in essentially identical data quality and refinement metrics for lysozyme microcrystals under the exposure and sample delivery conditions available in this experiment. Conclusions one

can draw from this result are as follows: (1) The structures from each pulse in the pulse train show no significant differences from each other to 1.7 Å resolution; thus, it is possible to measure time-resolved structures in the future using each pulse in the pulse train without having to account for changes due to the megahertz repetition rate under the exposure conditions available in this experiment; (2) since the structures determined from each pulse are essentially identical, it is possible to merge data from all pulses in the pulse train into one dataset without having to account for changes due to the megahertz repetition rate; and (3) it is now more feasible than ever to improve data quality by averaging of large datasets for time resolved experiments or using small or weakly diffracting crystals, obtaining sufficient SNR for structure determination in reasonable experiment time;<sup>7</sup> alternatively (4) sufficient data for a molecular replacement structure determination using 1200 pulses/second can enable faster completion of datasets for structure determination by SFX as previously predicted.<sup>15</sup> Observation that no significant changes through the pulse train were observed is valid only for current experimental conditions, including reflective optics for a micrometer scale focal spot and HEWL as a test sample. One should still be careful to further verify these conclusions when performing experiments at higher peak pulse powers, using highly absorbing metal centers, or with slower liquid jet delivery than used here. Nevertheless, the results presented here form a useful baseline demonstrating negligible pulse train damage for reference in future experiments and demonstrate the potential for European XFEL to

**TABLE II.** SFX data and refinement statistics for all pulses merged together.

Dataset	Data collection		Data quality				Refinement				
	Number of hits	Number of lattices	CC*	$R_{\text{split}}$ (%)	SNR	$R_{\text{work}}$	$R_{\text{free}}$	RMSD bonds (Å)	RMSD angles (deg)	B-factor (Å <sup>2</sup> )	N water
Run 97 only (approximately 10 min)	67 090	117 415	0.9993	4.84	15.17	0.140	0.164	0.009	0.99	25.3	67
Runs 96–98 (approximately 30 min)	189 960	332 482	0.9998	2.86	25.59	0.137	0.159	0.009	0.97	25.8	64
All data to 1.7 Å	830 514	1 374 785	0.9999	1.51	48.74	0.135	0.157	0.009	0.98	25.9	70
All data to 1.6 Å	830 514	1 374 785	0.9999	1.80	41.46	0.142	0.160	0.009	1.00	25.4	64
All data to 1.6 Å merged in PG 422	830 514	1 374 785	0.9999	1.99	36.18	0.153	0.171	0.006	0.89	25.5	81

measure data using X-ray pulses with a 1.1 MHz repetition rate for the collection of datasets with higher statistics than previously possible in the same measurement time.

## ACKNOWLEDGMENTS

We acknowledge European XFEL in Schenefeld, Germany, for provision of the X-ray free-electron laser beamtime at Scientific Instrument SPB/SFX (Single Particles, Clusters, and Biomolecules and Serial Femtosecond Crystallography) and would like to thank the instrument group and facility staff for their assistance.

We acknowledge the support of funding from The Gottfried Wilhelm Leibniz Program of the DFG; the project “X-probe” funded by the European Union’s 2020 Research and Innovation Program under the Marie Skłodowska-Curie Grant Agreement No. 637295; the European Research Council, “Frontiers in Attosecond X-ray Science: Imaging and Spectroscopy (AXSIS),” No. ERC-2013-SyG 609920; the Human Frontiers Science Program Grant No. RGP0010 2017; The European Research Council under the European Union’s Seventh Framework Programme (No. FP7/2007-2013) through the Consolidator Grant COMOTION (No. ERC-614507-Küpper) and by the Deutsche Forschungsgemeinschaft through the Clusters of Excellence “Center for Ultrafast Imaging” (CUI, EXC1074, ID194651731) and “Advanced Imaging of Matter” (AIM, EXC2056, ID390715994); The Swedish Research Council, the Swedish Foundation for Strategic Research, and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT); Ministerio de Economía y Competitividad (Spain), Plan Estatal 2013-2016 Retos, Project No. DPI2016-78887-C3-1-R; the Australian Research Council Center of Excellence in Advanced Molecular Imaging (No. CE140100011), [www.imagingcoe.org](http://www.imagingcoe.org); the Australian Nuclear Science and Technology Organisation (ANSTO); the International Synchrotron Access Program (ISAP) managed by the Australian Synchrotron, part of ANSTO, and funded by the Australian Government; the Biodesign Center for Applied Structural Discovery at Arizona State University, the National Science Foundation (NSF) Science and Technology Center (STC) “BioXFEL” through Award No. STC-1231306 and NSF Award No. #1565180; and the National Institutes of Health Grant No. R01 GM095583.

## APPENDIX: METHODS

### Sample preparation

Samples were prepared in the same manner as previously reported.<sup>15</sup> Briefly, crystals of hen egg-white lysozyme (HEWL) were grown by the rapid-mixing batch method.<sup>30</sup> Tetragonal crystals with a size of each side 2–5  $\mu\text{m}$  were obtained by adding three parts of precipitant [1.2 M NaCl, 36%(v/v) ethylene glycol, 15%(w/v) PEG 4000, 50 mM sodium acetate buffer pH 3.5 filtered through a 450 nm filter] to one part of HEWL (Sigma-Aldrich; dissolved to 126 mg ml<sup>-1</sup> in 50 mM sodium acetate buffer pH 3.5 and filtered through a 100 nm filter) at 1 °C (ThermoStat C, Eppendorf, Germany). The resulting mixture was immediately subjected to rapid mixing and incubated for 30 min at 1 °C.<sup>31</sup> Crystal sizes were determined by image analysis by optical microscopy. Crystals were

resuspended before injection to yield a homogenous suspension of HEWL microcrystals.

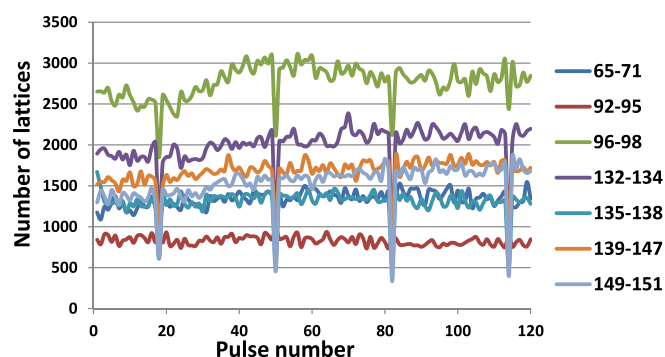
## Instrumentation

Experiments were performed at the SPB/SFX serial femtosecond crystallography instrument<sup>32</sup> at the European XFEL X-ray free electron laser in September 2018 as a part of proposal p002120 using a similar configuration as used in Wiedorn *et al.*<sup>15</sup> as described in the main text. The size of the focal spot in the interaction region was estimated to be  $2 \times 3 \mu\text{m}^2$  FWHM based on optical imaging of single shots using a 20  $\mu\text{m}$  thick Ce:YAG screen. Diffraction from the sample was measured using an AGIPD 1M located 117.7–118.6 mm downstream of the sample interaction region, with the unused direct beam passing through a central hole in the detector to a beam stop further downstream.

HEWL microcrystals were delivered to the X-ray interaction region in a similar manner as used in Wiedorn *et al.*<sup>15</sup> using a 3D-printed gasdynamic virtual nozzle<sup>33–35</sup> in which a liquid stream is focused and accelerated by the virtual orifice created by a copropagating helium gas flow or a 3D-printed double flow focusing nozzle.<sup>35</sup> The speed of the liquid jet was estimated to be 100 m/s based on measured flow rates during the experiments and offline speed measurements under the same operating conditions using similar printed 3D nozzles.<sup>20</sup> The liquid jets were positioned in the interaction region by mounting nozzles on a moveable “nozzle rod,” which held the jets just above the X-ray focal position and aligned to the X-ray beam using an in-line microscope viewing system.

## Data analysis

Experimental progress was monitored online using OnDA<sup>36</sup> for serial crystallography. Of the 16 654 608 diffraction patterns collected during HEWL data acquisition runs used for final analysis, 830 514 images (5%) were determined by Cheetah<sup>22</sup> to contain crystal diffraction patterns (peakfinder8, minSNR = 8, minADC = 50, minPix = 1, minPeaks = 20) using the calibration process described in Wiedorn *et al.*<sup>15</sup> Careful masking of shadowed and unreliable regions of the detector was performed on run-by-run basis. Independent masks were used for peakfinding to avoid false peaks, for example, due to ice formation. Indexing was performed using CrystFEL version 0.8.0, using the indexing package XGANDALF<sup>37</sup> on peaks found by Cheetah.<sup>22</sup> Indexing of multiple lattices per image resulted in a higher number of indexed lattices than the number of input images. Experimental geometry, especially the detector distance, was refined for 8 blocks of runs (with consistent parameters) using the program geoptimiser.<sup>38</sup> Merging and scaling of the Bragg peaks intensities were performed using the partialator program from CrystFEL. To ensure similar processing for each of the 120 pulses, the option custom-split was used. To avoid the integration of noise for weakly scattering patterns, reflections were included up to 0.6 nm<sup>-1</sup> above a conservative resolution estimate for each crystal (-push-res = 0.2). MTZ-files for crystallographic data-processing were generated from CrystFEL merged reflection data files using f2mtz of CCP4.<sup>28</sup> Figures of merit were calculated using compare\_hkl (Rsplit, CC<sub>1/2</sub>, CC\*) and check\_hkl (SNR, multiplicity, completeness), both a part of CrystFEL. The total number of indexed crystals 1 374 785 resulted in 11 000–12 000 crystals for



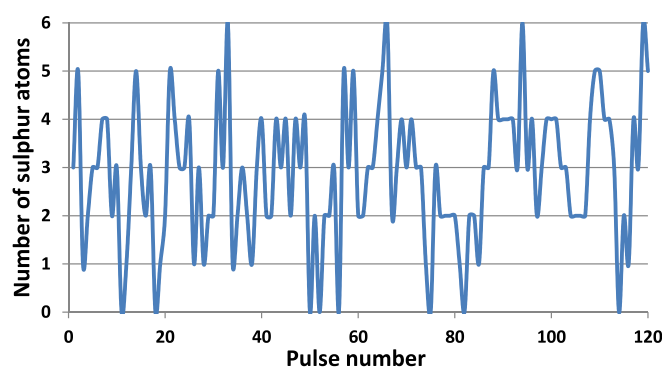
**FIG. 8.** Number of lattices per pulse for different blocks of runs. Small deviation from linear behavior for some plots might be due to different intensities for each pulse in the train. This intensity distribution was changing during the 5 days of the experiment.

each of the 120 pulses. The change in the number of measured “hits” per pulse (Fig. 2 and Table III in supplemental material) might be explained by the change of the intensity of x-ray pulses within each train. The distribution of the pulse intensity inside train was changing during the whole experiment, which was observed at the monitors in the control hut. This is reflected in the change of number of “hits” for different runs measured during the 5 days beamtime—Fig. 8(a). Unfortunately, information about per pulse intensity (or more precisely electron bunch energy) was not saved during the experiment at that time. Reconstruction of the equivalent number of patterns per pulse (6500) was also performed but showed no noticeable difference to the results obtained from all data.

For RADDOSE-3D calculations, the following parameters are used: cubic crystal of  $5 \times 5 \times 5 \mu\text{m}^3$  with 10 pixels/ $\mu\text{m}$ , providing PDB code 4ET8, Gaussian beam with  $4.8 \times 10^{24}$  photons/s in  $2 \times 3 \mu\text{m}^2$  FWHM size at 9.3 keV and  $1 \times 10^{13}$  s exposure.

### Structure determination

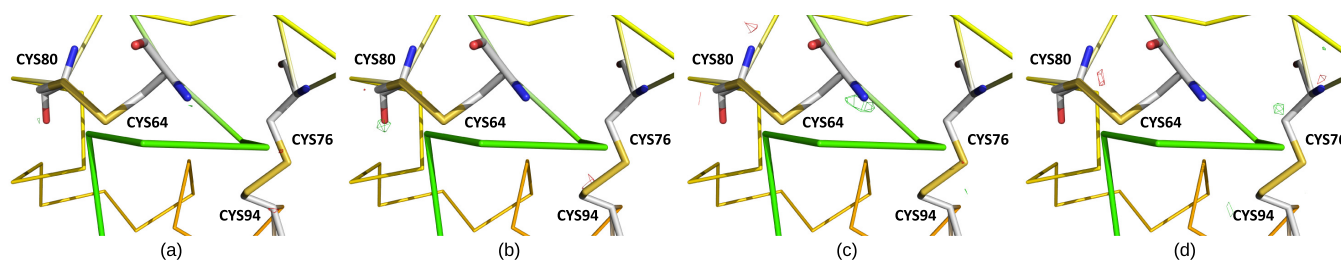
A structural model of lysozyme from Wiedorn *et al.*<sup>15</sup> (PDB accession code 6FTR) was used as a starting model for molecular replacement in Phaser.<sup>39</sup> Phaser was used to orient the model according to the orientation of the data. Using the structural model directly in phenix.refine (for which the isomorphism of the crystals would have allowed) failed. For all subsequent refinement, the same reoriented and re-refined PDB-model based on 6FTR with all solvent



**FIG. 10.** Number of sulfur anomalous difference map peaks per pulse. Only peaks with a height greater than 3 sigma were considered.

molecules was used, alternative conformations and ligands as input for refinement. A simple bash script generated the same  $R_{\text{free}}$ -flags for all datasets (copied from 6FTR and extended to full resolution) and the same refinement parameters used for the overall dataset were used for all single-pulse datasets, as well as the same starting model. Note that for this publication, we did not strive to find the best refined model, and it was for us more important to have comparable results. Hence, the structural model was not optimized in manual rebuilding cycles combined with automated maximum-likelihood refinement. The same input-model, same  $R_{\text{free}}$ -flags, and same refinement parameters in Phenix were used. Isomorphous Fo-Fo difference maps for the structures reconstructed from the first and several subsequent pulses were calculated using the Isomorphous difference map tool in Phenix and are presented in Fig. 9.

For anomalous sulfur SAD-phasing for the combined dataset, the reflection list using point group 422 in partialator was generated. Point group 422 was used to avoid merging of Friedel pairs of reflection in point group 4/mmm. MTZ files were generated as described above and phenix.french\_wilson was used to generate MTZ-files with F+/F- columns. Anomalous phasing was attempted using the CRANK2-pipeline in CCP4 employing the SHLEX/D/E pipeline and the programs Parrot, Buccaneer,<sup>40</sup> and REFMAC5.<sup>41</sup> For the successful phasing case, the MTZ-file and initial model from CRANK2 were exported to AutoBuild in Phenix for advanced automatic model building. Data metrics for 4/mmm and 422 point groups are summarized in Table II. Figure 10 also demonstrates the number of sulfur anomalous difference map peaks above 3 sigma per pulse. It shows



**FIG. 9.** Isomorphous Fo-Fo difference maps contoured at 3 sigma, comparing observed data from the first pulses only with observed data from (a) 2nd pulses, (b) 3rd pulses, (c) 60th pulses, and (d) 120th pulses.



no clear trend depending on the pulse number, which also suggests no apparent damage for different pulses in the train. The phased anomalous Fourier difference map was calculated with phenix.find\_peaks\_holes from Phenix using a custom script.

### Data deposition

Data have been deposited with the Coherent X-ray Imaging Data Bank<sup>42</sup> with deposition ID 98. These include the following:

- Crystal hits used for indexing
- Stream files for all data and for data separated into each pulse.
- MTZ and PDB files for all data and separated by pulse ID

We do not see a point of depositing 120 PDBs to the Protein Databank, and so we have deposited all resulting data for all 120 pulses to the CXIDB only.

### SUPPLEMENTARY MATERIAL

See [supplementary material](#) for SFX data and refinement statistics separated by pulse ID in the pulse train.

### REFERENCES

- <sup>1</sup>H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmeß, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhober, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, “Femtosecond X-ray protein nanocrystallography,” *Nature* **470**, 73–77 (2011).
- <sup>2</sup>S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, M. Messerschmidt, A. Barty, T. A. White, S. Kassemeier, R. A. Kirian, M. M. Seibert, P. A. Montanez, C. Kenney, R. Herbst, P. Hart, J. Pines, G. Haller, S. M. Gruner, H. T. Philipp, M. W. Tate, M. Hromalik, L. J. Koerner, N. van Bakel, J. Morse, W. Ghonsalves, D. Arnlund, M. J. Bogan, C. Caleman, R. Fromme, C. Y. Hampton, M. S. Hunter, L. C. Johansson, G. Katona, C. Kupitz, M. Liang, A. V. Martin, K. Nass, L. Redecke, F. Stellato, N. Timneanu, D. Wang, N. A. Zatsepin, D. Schafer, J. Defever, R. Neutze, P. Fromme, J. C. H. Spence, H. N. Chapman, and I. Schlichting, “High-resolution protein structure determination by serial femtosecond crystallography,” *Science* **337**, 362–364 (2012).
- <sup>3</sup>Y. Kang, X. E. Zhou, X. Gao, Y. He, W. Liu, A. Ishchenko, A. Barty, T. A. White, O. Yefanov, G. W. Han, Q. Xu, P. W. De Waal, J. Ke, M. H. E. Tan, C. Zhang, A. Moeller, G. M. West, B. D. Pascal, N. Van Eps, L. N. Caro, S. A. Vishnivetskiy, R. J. Lee, K. M. Suino-Powell, X. Gu, K. Pal, J. Ma, X. Zhi, S. Boutet, G. J. Williams, M. Messerschmidt, C. Gati, N. A. Zatsepin, D. Wang, D. James, S. Basu, S. Roy-Chowdhury, C. E. Conrad, J. Coe, H. Liu, S. Lisova, C. Kupitz, I. Grotjohann, R. Fromme, Y. Jiang, M. Tan, H. Yang, J. Li, M. Wang, Z. Zheng, D. Li, N. Howe, Y. Zhao, J. Standfuss, K. Diederichs, Y. Dong, C. S. Potter, B. Carragher, M. Caffrey, H. Jiang, H. N. Chapman, J. C. H. Spence, P. Fromme, U. Weierstall, O. P. Ernst, V. Katritch, V. V. Gurevich, P. R. Griffin, W. L. Hubbell, R. C. Stevens, V. Cherezov, K. Melcher, and H. E. Xu, “Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser,” *Nature* **523**, 561–567 (2015).
- <sup>4</sup>J. Tenboer, S. Basu, N. Zatsepin, K. Pande, D. Milathianaki, M. Frank, M. Hunter, S. Boutet, G. J. Williams, J. E. Koglin, D. Oberthuer, M. Heymann, C. Kupitz, C. Conrad, J. Coe, S. Roy-Chowdhury, U. Weierstall, D. James, D. Wang, T. Grant, A. Barty, O. Yefanov, J. Scales, C. Gati, C. Seuring, V. Srajer, R. Henning, P. Schwander, R. Fromme, A. Ourmazd, K. Moffat, J. J. Van Thor, J. C. H. Spence, P. Fromme, H. N. Chapman, and M. Schmidt, “Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein,” *Science* **346**, 1242–1246 (2014).
- <sup>5</sup>K. Pande, C. D. M. Hutchison, G. Groenhof, A. Aquila, J. S. Robinson, J. Tenboer, S. Basu, S. Boutet, D. P. DePonte, M. Liang, T. A. White, N. A. Zatsepin, O. Yefanov, D. Morozov, D. Oberthuer, C. Gati, G. Subramanian, D. James, Y. Zhao, J. Koralek, J. Brayshaw, C. Kupitz, C. Conrad, S. Roy-Chowdhury, J. D. Coe, M. Metz, P. L. Xavier, T. D. Grant, J. E. Koglin, G. Ketawala, R. Fromme, V. rاجر, R. Henning, J. C. H. Spence, A. Ourmazd, P. Schwander, U. Weierstall, M. Frank, P. Fromme, A. Barty, H. N. Chapman, K. Moffat, J. J. van Thor, and M. Schmidt, “Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein,” *Science* **352**, 725–729 (2016).
- <sup>6</sup>J. R. Stagno, Y. Liu, Y. R. Bhandari, C. E. Conrad, S. Panja, M. Swain, L. Fan, G. Nelson, C. Li, D. R. Wendel, T. A. White, J. D. Coe, M. O. Wiedorn, J. Knoska, D. Oberthuer, R. A. Tuckey, P. Yu, M. Dyba, S. G. Tarasov, U. Weierstall, T. D. Grant, C. D. Schwieters, J. Zhang, A. R. Ferré-D’Amaré, P. Fromme, D. E. Draper, M. Liang, M. S. Hunter, S. Boutet, K. Tan, X. Zuo, X. Ji, A. Barty, N. A. Zatsepin, H. N. Chapman, J. C. H. Spence, S. A. Woodson, and Y.-X. Wang, “Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography,” *Nature* **541**, 242–246 (2017).
- <sup>7</sup>C. Gati, D. Oberthuer, O. Yefanov, R. D. Bunker, F. Stellato, E. Chiu, S.-M. Yeh, A. Aquila, S. Basu, R. Bean, K. R. Beyerlein, S. Botha, S. Boutet, D. P. DePonte, R. B. Doak, R. Fromme, L. Galli, I. Grotjohann, D. R. James, C. Kupitz, L. Lomb, M. Messerschmidt, K. Nass, K. Rendek, R. L. Shoeman, D. Wang, U. Weierstall, T. A. White, G. J. Williams, N. A. Zatsepin, P. Fromme, J. C. H. Spence, K. N. Goldie, J. A. Jehle, P. Metcalf, A. Barty, and H. N. Chapman, “Atomic structure of granulin determined from native nanocrystalline granulovirus using an X-ray free-electron laser,” *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2247–2252 (2017).
- <sup>8</sup>M. Suga, F. Akita, K. Hirata, G. Ueno, H. Murakami, Y. Nakajima, T. Shimizu, K. Yamashita, M. Yamamoto, H. Ago, and J.-R. Shen, “Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses,” *Nature* **517**, 99–103 (2015).
- <sup>9</sup>M. Altarelli and A. P. Mancuso, “Structural biology at the European X-ray free-electron laser facility,” *Philos. Trans. R. Soc.* **B 369**, 20130311 (2014).
- <sup>10</sup>S. P. Hau-Riege, R. A. London, and A. Szoke, “Dynamics of biological molecules irradiated by short x-ray pulses,” *Phys. Rev. E* **69**, 051906 (2004).
- <sup>11</sup>C. Caleman, C. Ortiz, E. Marklund, F. Bultmark, M. Gabrysch, F. G. Parak, J. Hajdu, M. Klintonberg, and N. Timneanu, “Radiation damage in biological material: Electronic properties and electron impact ionization in urea,” *Europhys. Lett.* **85**, 18005 (2009).
- <sup>12</sup>C. Caleman and A. V. Martin, *X-Ray Free Electron Lasers* (Springer International Publishing, Cham, 2018), pp. 185–207.
- <sup>13</sup>C. A. Stan, D. Milathianaki, H. Laksmono, R. G. Sierra, T. A. McQueen, M. Messerschmidt, G. J. Williams, J. E. Koglin, T. J. Lane, M. J. Hayes, S. A. H. Guillet, M. Liang, A. L. Aquila, P. R. Willmott, J. S. Robinson, K. L. Gumerlock, S. Botha, K. Nass, I. Schlichting, R. L. Shoeman, H. A. Stone, and S. Boutet, “Liquid explosions induced by X-ray laser pulses,” *Nat. Phys.* **12**, 966–971 (2016).
- <sup>14</sup>G. Blaj, M. Liang, A. L. Aquila, P. R. Willmott, J. E. Koglin, R. G. Sierra, J. S. Robinson, S. Boutet, and C. A. Stan, “Generation of high-intensity ultrasound through shock propagation in liquid jets,” *Phys. Rev. Fluids* **4**, 043401 (2019).
- <sup>15</sup>M. O. Wiedorn, D. Oberthür, R. Bean, R. Schubert, N. Werner, B. Abbey, M. Aepfelbacher, L. Adriano, A. Allahgholi, N. Al-Qudami, J. Andreasson, S. Aplin, S. Awel, K. Ayyer, S. Bajt, I. Barák, S. Bari, J. Bielecki, S. Botha, D. Boukhelef, W. Brehm, S. Brockhauser, I. Cheviakov, M. A. Coleman, F. Cruz-Mazo, C. Danilevski, C. Darmanin, R. B. Doak, M. Domaracky, K. Dörner, Y. Du, H. Fangohr, H. Fleckenstein, M. Frank, P. Fromme, A. M. Gañán-Calvo, Y. Gevorkov, K. Giewekemeyer, H. M. Ginn, H. Graafsma, R. Graceffa, D. Greiffenberg, L. Gumprecht, P. Göttlicher, J. Hajdu, S. Hauf, M. Heymann, S. Holmes, D. A. Horke, M. S. Hunter, S. Inlaur, A. Kaukher, Y. Kim, A. Klyuev, J.

- Knoška, B. Kobe, M. Kuhn, C. Kupitz, J. Küpper, J. M. Lahey-Rudolph, T. Laurus, K. Le Cong, R. Letrun, P. L. Xavier, L. Maia, F. R. N. C. Maia, V. Mariani, M. Messerschmidt, M. Metz, D. Mezza, T. Michelat, G. Mills, D. C. F. Monteiro, A. Morgan, K. Mühlig, A. Munke, A. Münnich, J. Nette, K. A. Nugent, T. Nuguid, A. M. Orville, S. Pandey, G. Pena, P. Villanueva-Perez, J. Poehlsen, G. Previtali, L. Redecke, W. M. Riekehr, H. Rohde, A. Round, T. Safenreiter, I. Sarrou, T. Sato, M. Schmidt, B. Schmitt, R. Schönherr, J. Schulz, J. A. Sellberg, M. M. Seibert, C. Seuring, M. L. Shelby, R. L. Shoeman, M. Sikorski, A. Silenzi, C. A. Stan, X. Shi, S. Stern, J. Sztuk-Dambietz, J. Szuba, A. Tolstikova, M. Trebbin, U. Trunk, P. Vagovic, T. Ve, B. Weinhausen, T. A. White, K. Wrona, C. Xu, O. Yefanov, N. Zatsepin, J. Zhang, M. Perbandt, A. P. Mancuso, C. Betzel, H. Chapman, and A. Barty, "Megahertz serial crystallography," *Nat. Commun.* **9**, 4025 (2018).
- <sup>16</sup>M. L. Grünbein, J. Bielecki, A. Gorel, M. Stricker, R. Bean, M. Cammarata, K. Dörner, L. Fröhlich, E. Hartmann, S. Hauf, M. Hilpert, Y. Kim, M. Kloos, R. Letrun, M. Messerschmidt, G. Mills, G. Nass Kovacs, M. Ramilli, C. M. Roome, T. Sato, M. Scholz, M. Sliwa, J. Sztuk-Dambietz, M. Weik, B. Weinhausen, N. Al-Qudami, D. Boukhelef, S. Brockhauser, W. Ehsan, M. Emons, S. Esenov, H. Fangohr, A. Kaukher, T. Kluyver, M. Lederer, L. Maia, M. Manetti, T. Michelat, A. Münnich, F. Pallas, G. Palmer, G. Previtali, N. Raab, A. Silenzi, J. Szuba, S. Venkatesan, K. Wrona, J. Zhu, R. B. Doak, R. L. Shoeman, L. Foucar, J.-P. Colletier, A. P. Mancuso, T. R. M. Barends, C. A. Stan, and I. Schlichting, "Megahertz data collection from protein microcrystals at an X-ray free-electron laser," *Nat. Commun.* **9**, 3487 (2018).
- <sup>17</sup>A. P. Mancuso, A. Aquila, L. Batchelor, R. J. Bean, J. Bielecki, G. Borchers, K. Doerner, K. Giewekemeyer, R. Graceffa, O. D. Kelsey, Y. Kim, H. J. Kirkwood, A. Legrand, R. Letrun, B. Manning, L. Lopez Morillo, M. Messerschmidt, G. Mills, S. Raabe, N. Reimers, A. Round, T. Sato, J. Schulz, C. Signe Takem, M. Sikorski, S. Stern, P. Thute, P. Vagović, B. Weinhausen, and T. Tschentscher, "The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the European XFEL: Initial installation," *J. Synchrotron Radiat.* **26**, 660–676 (2019).
- <sup>18</sup>R. J. Bean, A. Aquila, L. Samoylova, and A. P. Mancuso, "Design of the mirror optical systems for coherent diffractive imaging at the SPB/SFX instrument of the European XFEL," *J. Opt.* **18**, 074011 (2016).
- <sup>19</sup>A. Allahgholi, J. Becker, A. Delfs, R. Dinapoli, P. Goettlicher, D. Greiffenberg, B. Henrich, H. Hirsemann, M. Kuhn, R. Klanner, A. Klyuev, H. Krueger, S. Lange, T. Laurus, A. Marras, D. Mezza, A. Mozzanica, M. Niemann, J. Poehlsen, J. Schewandt, I. Sheviakov, X. Shi, S. Smoljanin, L. Steffen, J. Sztuk-Dambietz, U. Trunk, Q. Xia, M. Zeribi, J. Zhang, M. Zimmer, B. Schmitt, and H. Graafsma, "The adaptive gain integrating pixel detector at the European XFEL," *J. Synchrotron Radiat.* **26**, 74–82 (2019).
- <sup>20</sup>M. O. Wiedorn, S. Awel, A. J. Morgan, K. Ayer, Y. Gevorkov, H. Fleckenstein, N. Roth, L. Adriano, R. Bean, K. R. Beyerlein, J. Chen, J. Coe, F. Cruz-Mazo, T. Ekeberg, R. Graceffa, M. Heymann, D. A. Horke, J. Knoška, V. Mariani, R. Nazari, D. Oberthür, A. K. Samanta, R. G. Sierra, C. A. Stan, O. Yefanov, D. Rompotis, J. Correa, B. Erk, R. Treusch, J. Schulz, B. G. Hogue, A. M. Gañán-Calvo, P. Fromme, J. Küpper, A. V. Rode, S. Bajt, R. A. Kirian, and H. N. Chapman, "Rapid sample delivery for megahertz serial crystallography at X-ray FELs," *IUCrJ* **5**, 574–584 (2018).
- <sup>21</sup>O. B. Zeldin, M. Gerstel, and E. F. Garman, "RADDOSE-3D: Time- and space-resolved modelling of dose in macromolecular crystallography," *J. Appl. Crystallogr.* **46**, 1225–1230 (2013).
- <sup>22</sup>A. Barty, R. A. Kirian, F. R. N. C. Maia, M. Hantke, C. H. Yoon, T. A. White, and H. Chapman, "Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data," *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).
- <sup>23</sup>T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty, and H. N. Chapman, "CrystFEL: A software suite for snapshot serial crystallography," *J. Appl. Crystallogr.* **45**, 335–341 (2012).
- <sup>24</sup>T. A. White, A. Barty, F. Stellato, J. M. Holton, R. A. Kirian, N. A. Zatsepin, and H. N. Chapman, "Crystallographic data processing for free-electron laser sources," *Acta Crystallogr., Sect. D* **69**, 1231–1240 (2013).
- <sup>25</sup>P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, "PHENIX: A comprehensive Python-based system for macromolecular structure solution," *Acta Crystallogr., Sect. D* **66**, 213–221 (2010).
- <sup>26</sup>H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucl. Acids Res.* **28**, 235–242 (2000).
- <sup>27</sup>G. M. Sheldrick, "A short history of SHELX," *Acta Crystallogr., Sect. A* **64**, 112–122 (2008).
- <sup>28</sup>M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, "Overview of the CCP4 suite and current developments," *Acta Crystallogr., Sect. D* **67**, 235–242 (2011).
- <sup>29</sup>T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L.-W. Hung, R. J. Read, and P. D. Adams, "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard," *Acta Crystallogr., Sect. D* **64**, 61–69 (2008).
- <sup>30</sup>F. Stellato, D. Oberthür, M. Liang, R. Bean, C. Gati, O. Yefanov, A. Barty, A. Burkhardt, P. Fischer, L. Galli, R. A. Kirian, J. Meyer, S. Panneerselvam, C. H. Yoon, F. Chervinskii, E. Speller, T. A. White, C. Betzel, A. Meents, and H. N. Chapman, "Room-temperature macromolecular serial crystallography using synchrotron radiation," *IUCrJ* **1**, 204–212 (2014).
- <sup>31</sup>K. R. Beyerlein, D. Dierksmeyer, V. Mariani, M. Kuhn, I. Sarrou, A. Ottaviano, S. Awel, J. Knoška, S. Fuglerud, O. Jönsson, S. Stern, M. O. Wiedorn, O. Yefanov, L. Adriano, R. Bean, A. Burkhardt, P. Fischer, M. Heymann, D. A. Horke, K. E. J. Jungnickel, E. Kovaleva, O. Lorbeer, M. Metz, J. Meyer, A. Morgan, K. Pande, S. Panneerselvam, C. Seuring, A. Tolstikova, J. Lieske, S. Aplin, M. Roessle, T. A. White, H. N. Chapman, A. Meents, and D. Oberthür, "Mix-and-diffuse serial synchrotron crystallography," *IUCrJ* **4**, 769–777 (2017).
- <sup>32</sup>A. P. Mancuso, N. Reimers, G. Borchers, A. Aquila, and K. Giewekemeyer, Technical Design Report No. TR-2013-004, Scientific Instrument Single Particles, Clusters, and Biomolecules (SPB), 2013.
- <sup>33</sup>D. P. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. C. H. Spence, and R. B. Doak, "Gas dynamic virtual nozzle for generation of microscopic droplet streams," *J. Phys. D: Appl. Phys.* **41**, 195505 (2008).
- <sup>34</sup>D. Oberthür, J. Knoška, M. O. Wiedorn, K. R. Beyerlein, D. A. Bushnell, E. G. Kovaleva, M. Heymann, L. Gumprecht, R. A. Kirian, A. Barty, V. Mariani, A. Tolstikova, L. Adriano, S. Awel, M. Barthelmeß, K. Dörner, P. L. Xavier, O. Yefanov, D. R. James, G. Nelson, D. Wang, G. Calvey, Y. Chen, A. Schmidt, M. Szczepiek, S. Frielingsdorf, O. Lenz, E. Snell, P. J. Robinson, B. Sarler, G. Belšák, M. Maček, F. Wilde, A. Aquila, S. Boutet, M. Liang, M. S. Hunter, P. Scheerer, J. D. Lipscomb, U. Weierstall, R. D. Kornberg, J. C. H. Spence, L. Pollack, H. N. Chapman, and S. Bajt, "Double-flow focused liquid injector for efficient serial femtosecond crystallography," *Sci. Rep.* **7**, 44628 (2017).
- <sup>35</sup>A. M. Gañán-Calvo, "Generation of steady liquid microthreads and micron-sized monodisperse sprays in gas streams," *Phys. Rev. Lett.* **80**, 285–288 (1998).
- <sup>36</sup>V. Mariani, A. Morgan, C. H. Yoon, T. J. Lane, T. A. White, C. O'Grady, M. Kuhn, S. Aplin, J. Koglin, A. Barty, and H. N. Chapman, "OnDA: Online data analysis and feedback for serial X-ray imaging," *J. Appl. Crystallogr.* **49**, 1073–1080 (2016).
- <sup>37</sup>Y. Gevorkov, O. Yefanov, A. Barty, T. A. White, V. Mariani, W. Brehm, A. Tolstikova, R.-R. Grigat, and H. N. Chapman, "XGANDALF – extended gradient descent algorithm for lattice finding," *Acta Crystallogr., Sect. A* **75**, 694–704 (2019).
- <sup>38</sup>O. Yefanov, V. Mariani, C. Gati, T. A. White, H. N. Chapman, and A. Barty, "Accurate determination of segmented X-ray detector geometry," *Opt. Express* **23**, 28459 (2015).
- <sup>39</sup>A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- <sup>40</sup>K. Cowtan, "The Buccaneer software for automated model building. 1. Tracing protein chains," *Acta Crystallogr., Sect. D* **62**, 1002–1011 (2006).
- <sup>41</sup>A. A. Vagin, R. S. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long, and G. N. Murshudov, "REFMAC5 dictionary: Organisation of prior chemical knowledge and guidelines for its use," *Acta Crystallogr., Sect. D* **60**, 2184–2195 (2004).
- <sup>42</sup>F. R. N. C. Maia, "The coherent x-ray imaging Data Bank," *Nat. Methods* **9**, 854–855 (2012).