

Long-wavelength native-SAD phasing: opportunities and challenges

Shibom Basu,^{a,‡} Vincent Olieric,^a Filip Leonarski,^a Naohiro Matsugaki,^b Yoshiaki Kawano,^c Tomizaki Takashi,^a Chia-Ying Huang,^a Yusuke Yamada,^b Laura Vera,^a Natacha Olieric,^d Jerome Basquin,^e Justyna A. Wojdyla,^a Oliver Bunk,^a Kay Diederichs,^f Masaki Yamamoto^c and Meitian Wang^{a*}

Received 21 November 2018

Accepted 22 February 2019

Edited by J. L. Smith, University of Michigan, USA

‡ Current address: EMBL Grenoble outstation, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France

Keywords: single-wavelength anomalous dispersion; native-SAD phasing; Se/S-SAD; UV-laser cutting; crystal shaping; spherical crystals; absorption correction; anomalous scattering factor; structure determination.

PDB references: T2R-TTL, 6i5c; Sen1, 6i59

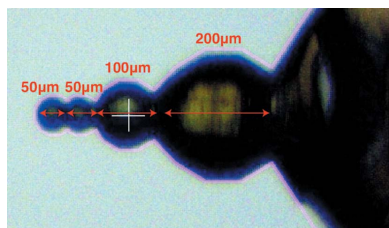
Supporting information: this article has supporting information at www.iucrj.org

^aSwiss Light Source, Paul Scherrer Institut, Villigen PSI, 5232, Switzerland, ^bStructural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization, Tsukuba, 305-0801, Japan, ^cAdvanced Photon Technology Division, RIKEN SPring-8 Center, Hyogo 679-5148, Japan, ^dLaboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institut, Villigen, PSI 5232, Switzerland, ^eDepartment of Biochemistry, Max Planck Institute of Biochemistry, Munich, Germany, and ^fDepartment of Biology, University of Konstanz, Konstanz, 78457, Germany. *Correspondence e-mail: meitian.wang@psi.ch

Native single-wavelength anomalous dispersion (SAD) is an attractive experimental phasing technique as it exploits weak anomalous signals from intrinsic light scatterers ($Z < 20$). The anomalous signal of sulfur in particular, is enhanced at long wavelengths, however the absorption of diffracted X-rays owing to the crystal, the sample support and air affects the recorded intensities. Thereby, the optimal measurable anomalous signals primarily depend on the counterplay of the absorption and the anomalous scattering factor at a given X-ray wavelength. Here, the benefit of using a wavelength of 2.7 over 1.9 Å is demonstrated for native-SAD phasing on a 266 kDa multiprotein-ligand tubulin complex (T₂R-TTL) and is applied in the structure determination of an 86 kDa helicase Sen1 protein at beamline BL-1A of the KEK Photon Factory, Japan. Furthermore, X-ray absorption at long wavelengths was controlled by shaping a lysozyme crystal into spheres of defined thicknesses using a deep-UV laser, and a systematic comparison between wavelengths of 2.7 and 3.3 Å is reported for native SAD. The potential of laser-shaping technology and other challenges for an optimized native-SAD experiment at wavelengths >3 Å are discussed.

1. Introduction

Most of our knowledge about the 3D atomic structure of biological macromolecules is derived directly or indirectly from experimental phasing (EP) in macromolecular crystallography (MX) thanks to the large success of selenomethionine incorporation and heavy-atom derivatization (Hendrickson, 2014). The emerging native single-wavelength anomalous dispersion (SAD) phasing method has distinct advantages over traditional EP using heavy atoms, but its practical use has so far been limited (Rose *et al.*, 2015). Indeed, the phasing signal from naturally present light elements (*e.g.* sulfur and phosphorus) is weak in the conventional energy range used at most MX beamlines *i.e.* between 6 and 20 keV (at wavelengths 2.07–0.62 Å) (Djinovic Carugo *et al.*, 2005). Most attempts at solving native-SAD structures in the early days, including the first native-SAD crambin structure in 1981 (Hendrickson & Teeter, 1981), were performed using a Cu source at a wavelength of 1.54 Å. Later, a dedicated laboratory X-ray system (Read *et al.*, 2001; Chen *et al.*, 2004; Gentry *et al.*, 2005; Deng *et al.*, 2005; Kitamura *et al.*, 2008; Alag *et al.*, 2009) producing a



wavelength of 2.29 Å with a chromium anode was used to solve ~31 *de novo* structures (Rose *et al.*, 2015). However, the relatively low flux and large beam size of such sources have limited their application to large and well diffracting crystals. In 2000, the first *de novo* native-SAD structure (obelin; Liu *et al.*, 2000) was solved with synchrotron radiation using a wavelength of 1.77 Å. Since then, about 60 *de novo* structures have been determined using wavelengths ranging from 1.70 to 2.07 Å (7.3–6.0 keV) at synchrotron beamlines with standard MX sample environments (Weinert *et al.*, 2015). Recent advances in data-collection methods and multi-crystal averaging have significantly improved the success rate of native SAD using 6 keV X-rays by enabling accurate diffraction-intensity measurement and effective data merging (Liu *et al.*, 2012; Weinert *et al.*, 2015; Liu & Hendrickson, 2015, 2017; Rose *et al.*, 2015; Olieric *et al.*, 2016). However, most systems with low sulfur content and/or low diffraction resolution worse than 3 Å are still out of reach.

For such challenging cases, it should be advantageous to use energy below 6 keV because the anomalous signal of S and P increases gradually towards lower energy ($\propto \lambda^2$). However, the X-ray absorption by the sample increases as well (cross-section $\propto \lambda^3$). The counterplay of these two factors produces crystal-size-dependent behavior when searching for the optimal energy, which maximizes the anomalous signal for native SAD (Fig. 1 and Appendix A) (Mueller-Dieckmann *et al.*, 2005; Liu *et al.*, 2013; Wagner *et al.*, 2016; Liebschner *et al.*,

2016). A wavelength of 3 Å appears to be optimal for a 100 µm sized ‘naked’ crystal, *i.e.* one without any surrounding solvent or a loop, in an ideal experiment [Fig. 1(a)]. However, in addition to the crystal itself, any material in the X-ray beam-path contributes to absorption and background scattering. Assuming 50 µm of solvent around the crystal and 100 mm of air between the crystal and the detector [Figs. 1(b) and 1(c), and S1 in the Supporting information], the optimum for native SAD shifts towards ~2 Å [Fig. 1(d)], a wavelength that has been used successfully at standard MX beamlines (Mueller-Dieckmann *et al.*, 2007; Liu *et al.*, 2012; Weinert *et al.*, 2015). Recent attempts towards using longer wavelengths were made at beamline BL-1A at the Photon Factory (PF), Japan and beamline I23 at the Diamond Light Source (DLS), UK, with helium and a vacuum environment, respectively. At beamline BL-1A, the advantage of a wavelength of 2.7 Å over a wavelength of 1.9 Å was demonstrated for both ferredoxin reductase and lysozyme crystals of 100 µm or smaller, while the advantage of using wavelengths of 3.0 Å (4.13 keV energy) and 3.3 Å (3.75 keV energy) remained elusive (Liebschner *et al.*, 2016). Very recently, a proof-of-principle native-SAD experiment using a wavelength of 4.96 Å (*i.e.* just above the sulfur *K* edge) with a thaumatin crystal was demonstrated for the first time (Aurelius *et al.*, 2017). In addition to model systems, only a few new structures have been determined using wavelengths ranging from 2.5 to 3.1 Å: death receptor 6 or DR6 (2.7 Å; Ru *et al.*, 2012) and Lili-Mip (2.7 Å; Banerjee

et al., 2016) at BL-1A in PF; PETase (2.5 Å; Austin *et al.*, 2018), SSeK3 (Se/S-SAD at 2.77 Å; Esposito *et al.*, 2018) and ThcOX (3.1 Å; Bent *et al.*, 2016) at I23 in DLS; and Cdc23^{Nterm} (2.7 Å; Cianci *et al.*, 2016) at P13 at PETRAIII. These results provide a glimpse into how X-rays with wavelengths longer than 2 Å can be exploited for native-SAD phasing. Indeed, the prospect of native SAD at wavelengths closer to the sulfur *K* edge is very appealing but the technical challenges caused by the increased absorption and scattering, as well as detector efficiency, could be impediments to its wide adoption at synchrotron beamlines. Therefore, we set out a systematic study to identify factors that currently limit the optimal use of longer wavelengths for native SAD and to propose approaches to overcome such limitations.

Both sample thickness – the crystal itself, as well as the loop and surrounding solvent – and absorption correction have to be considered at long wavelengths. While crystal size and morphology are difficult to control precisely during crystallization, crystal ablation using UV-laser-shaping technology (Murakami *et al.*, 2004) may be used to reduce the

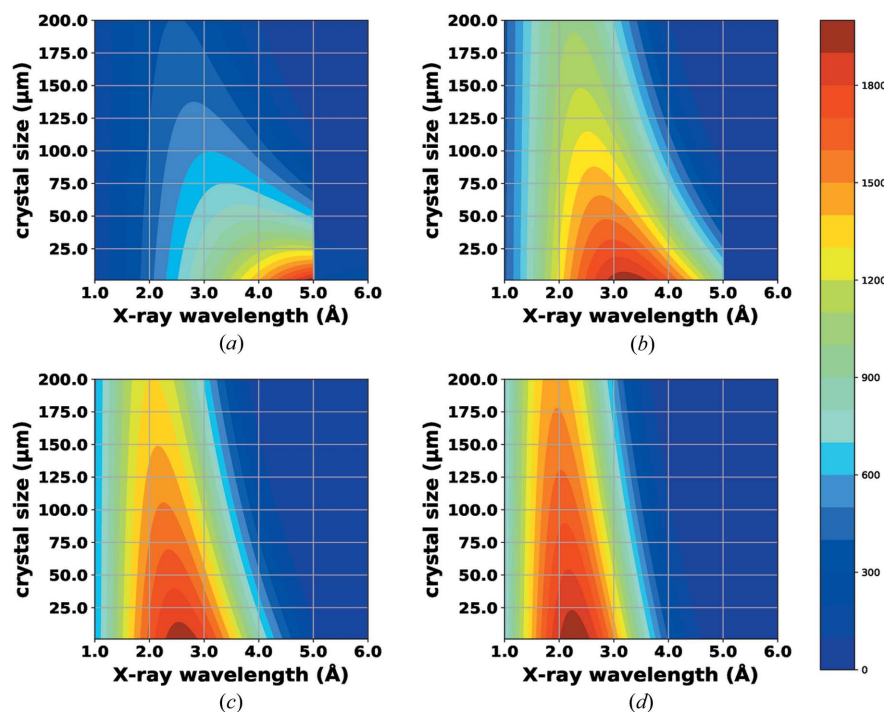
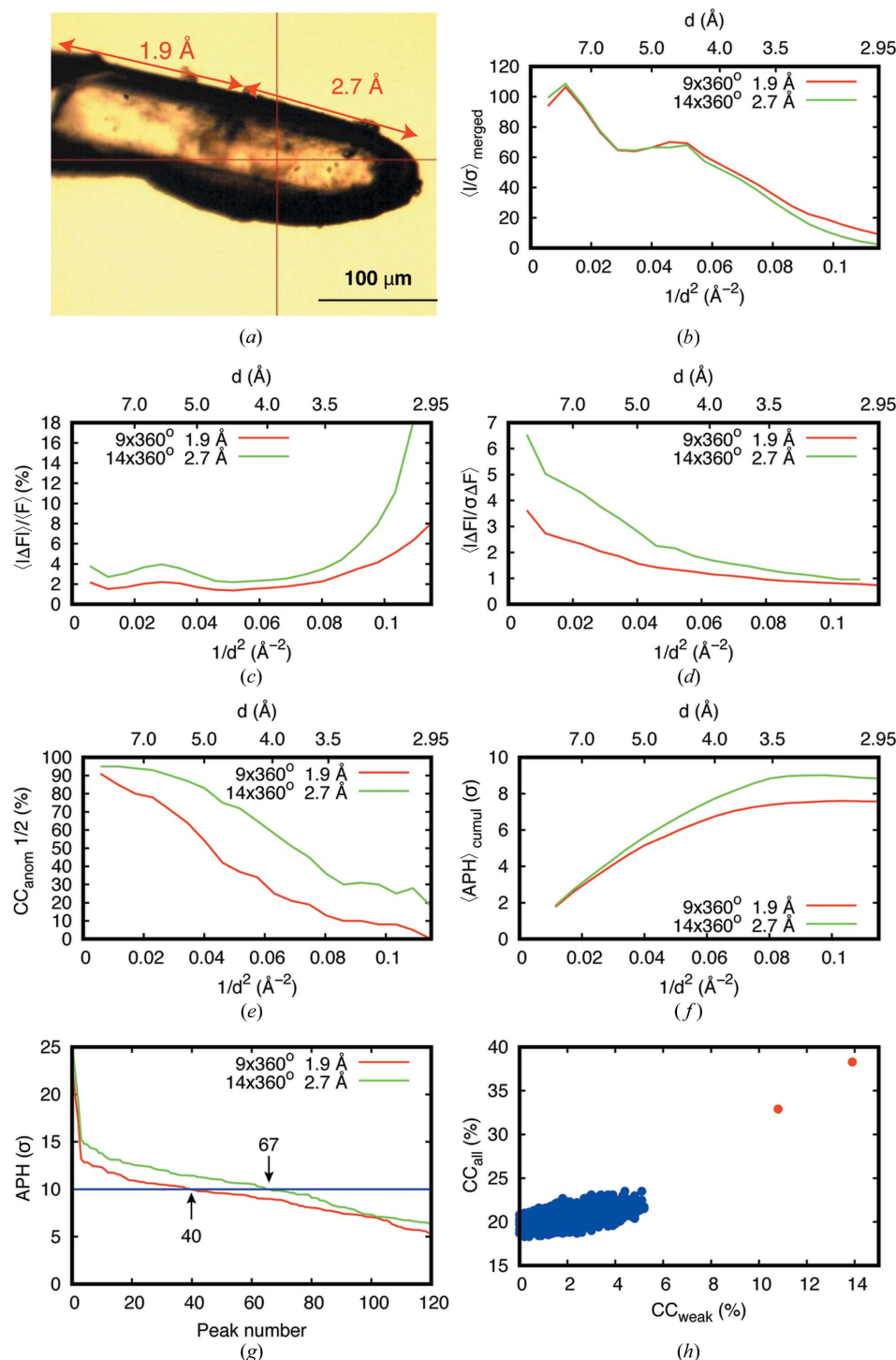


Figure 1
2D contour plots of theoretical anomalous diffraction efficiency for S atoms (shown as a heat map) as a function of X-ray wavelength (*x* axis) and crystal thickness (*y* axis). (a) In an ideal experimental condition with ‘naked’ crystals. (b) The absorption of 50 µm of solvent around the crystal is included. (c) The absorption of 100 mm of air in the scattering path between the crystal and the detector surface is included. (d) Both the 50 µm solvent layer and the 100 mm of air are included.


Figure 2

Measurement and comparison of T₂R-TTL crystal collected at 1.9 and 2.7 Å. (a) The crystal mounted on an elliptical Actiloop and the presence of minimum solvents around the crystal. The data-collection positions for each wavelength are marked with red lines with arrows. (b) $I/\sigma(I)$ values plotted against resolution for datasets collected at both wavelengths. (c)–(f) $\langle |\Delta F| \rangle / \langle F \rangle$, $\langle |\Delta F| \rangle / \sigma(\Delta F)$, $CC_{\text{anom}}(1/2)$ and average anomalous peak height ($\langle \text{APH} \rangle$) values are plotted against the diffraction resolution. (g) The anomalous peak heights (APH) are plotted for both wavelengths with dose-equivalent datasets. (h) CC_{all} versus CC_{weak} plot from the *SHELXD* solution for the 14 × 360° datasets at 2.7 Å.

sample thickness and to remove extra materials. A deep-UV laser – which can cut polymers such as proteins and fibers and can break chemical bonds by photochemical reactions – was developed for such applications at RIKEN, SPring-8, Japan. It was used to trim crystals from both lysozyme and the

membrane protein AcrB mounted on nylon loops under a cryogenic temperature of 100 K (Kitano *et al.*, 2005). The crystal integrity was shown to be preserved after laser irradiation (Kitano *et al.*, 2005), which spreads damage only within ~3 μm of the beam footprint (*Materials and methods*). These

results clearly show that this technique can shape fragile protein crystals in a more controlled way than mechanical actions such as manual cutting or sonication (de la Cruz *et al.*, 2017), and is effective in producing various geometric shapes including spheres.

In addition to sample absorption, other factors such as detector performance at low energy, and inaccuracy in data reduction and correction, could result in reduced data quality and compromise the gain in anomalous signal for native SAD at long wavelengths. In this study, the absorption effect for wavelengths $>2 \text{ \AA}$ in native-SAD phasing experiments was assessed systematically. We used a challenging 266 kDa tubulin complex to show the advantages of a 2.7 \AA wavelength over a 1.9 \AA wavelength for native SAD. We then successfully applied a 2.7 \AA wavelength to solve a 86 kDa helicase Sen1 protein using a multi-orientation data-collection protocol (Weinert *et al.*, 2015). Finally, we exploited the potential of a wavelength of 3.3 \AA using spherical lysozyme crystals that have been shaped by the laser (Kitano *et al.*, 2005). The conditions required to perform an optimal native-SAD experiment at X-ray wavelengths $>2 \text{ \AA}$, in particular with regard to both sample absorption and detector technology, are discussed.

2. Materials and methods

2.1. Sample preparation and crystallization

2.1.1. Lysozyme. Lysozyme crystals were produced by the vapor-diffusion method. The protein concentration was 50 mM at 50 mg ml⁻¹. The lysozyme was solubilised at 50 mg ml⁻¹ in 50 mM sodium acetate at pH 4.5. The crystals were obtained by mixing 1 μl of the protein with 1 μl of reservoir solution, consisting of 50 mM sodium acetate at pH 4.5, 5% PEG MME 5000 and 25% ethylene glycol. Lysozyme crystals of average size 800 \times 500 \times 400 μm in space group $P4_32_12$ grew within two days. They were harvested in MiTeGen MicroLoops E and snap-cooled in liquid nitrogen prior to the laser-shaping experiment.

2.1.2. T₂R-TTL. Tubulin-TTL is a multi-ligand globular protein complex (PDB code 4wbn; Weinert *et al.*, 2015) of size 266 kDa, containing 118 S, 13 P, 2 Cl⁻ and 3 Ca²⁺, which crystallizes in space group $P2_12_12_1$. The protein was expressed, purified and crystallized as described elsewhere (Prota *et al.*, 2013). The needle-like crystals were harvested in Molecular Dimensions ActiLoops with minimum surrounding solvent ($<10 \mu\text{m}$) and snap-cooled in liquid nitrogen.

2.1.3. Sen1. Sen1 is a superfamily 1B (SF1B) helicase protein of 85.7 kDa size (PDB code 5mzn), containing 32 S. Sen1 was expressed and purified from *Escherichia coli* as described elsewhere (Leonaitė *et al.*, 2017). The Sen1 protein was concentrated to 3 mg ml⁻¹ and crystallized at 4°C using the vapor-diffusion method by mixing an equal volume of protein with reservoir solution. The crystallization solution consists of 6% PEG 8000, 8% ethylene glycol and 0.1 M HEPES at pH 7.5 buffer. The crystals, in space group $P2_12_12$, were harvested in Molecular Dimensions ActiLoops with

Table 1

Data collection and refinement statistics for T₂R-TTL and Sen1 native-SAD experiments at a wavelength of 2.7 \AA .

Values in parenthesis represent the highest resolution shell.

Protein	T ₂ R-TTL	Sen1
PDB entry	6i5c	6i59
Data collection		
Photon energy (keV)	4.6	4.6
Beam size (μm^2)	40 \times 40	40 \times 40
Flux (photons ⁻¹)	4.1 \times 10 ⁹	4.1 \times 10 ⁹
Space group	$P2_12_12_1$	$P2_12_12$
Unit-cell dimensions (\AA)	$a = 104.24, b = 156.83,$ $c = 179.54$	$a = 69.11, b = 91.05,$ $c = 172.07$
Oscillation angle (°)	0.2	0.2
Exposure time (s)	0.1	0.1
Total range (°)	14 \times 360	4 \times 360
Detector distance (mm)	60	60
Total dose (MGy)	3.9	1.8
κ angles (°)	0–65; $\Delta = 5^\circ$	0–30; $\Delta = 10^\circ$
No. of crystal positions	1	1
Structure		
Crystal size (μm^3)	500 \times 70 \times 50	220 \times 100 \times 50
Molecular weight (kDa)/ No. of residues	266/2363	85.7/720
Monomer/ asymmetric unit	1	1
No. of scatterers	118 S, 13 P, 2 Cl, 3 Ca	32 S
Bijvoet ratio (%)	1.53	1.39
Phasing		
Resolution (\AA)	50–2.95 (3.03–2.95)	50–2.95 (3.03–2.95)
No. of unique reflections	62640 (4564)	23557 (1653)
No. of total reflections	11135360 (625357)	1157226 (62626)
Multiplicity	177.8 (137.0)	49.1 (37.9)
Completeness (%)	99.9 (99.5)	99.1 (95.2)
R_{meas} (%)	15.3 (289.1)	3.18 (42.79)
CC _{1/2} (%)	100 (82.3)	99.9 (69.3)
$\langle I/\sigma(I) \rangle$	52.23 (3.29)	31.08 (1.66)
Mosaicity (°)	0.17	0.17
SHELXD resolution cut-off (\AA)	3.5	3.3
CC _{all} /CC _{weak}	38.3/13.9	36.0/18.9
Solvent content (%)	56.3	60.9
Refinement		
$R_{\text{work}}/R_{\text{free}}$ (%)	17.0/20.8	16.9/21.3
RMS deviations		
Bond lengths (\AA)	0.003	0.003
Bond angles (°)	0.692	0.658
Wilson B factor (\AA^2)	77.9	88.2
Average B factor (\AA^2)		
All atoms	73.5	88.8
Macromolecules	73.7	83.4
Ligands	72.0	130.2
Solvent	66.3	102.8
Clashscore	4.0	6.5
Ramachandran plot		
Favored (%)	98.0	97.2
Allowed (%)	2.0	2.8
Outliers (%)	0	0

minimum surrounding solvent (10–20 μm) and snap-cooled in liquid nitrogen.

2.2. BL-1A experimental setup

The long-wavelength native-SAD experiments were carried out at beamline BL-1A at the Photon Factory, KEK, Japan, at X-ray wavelengths of 1.9, 2.7 and 3.3 \AA using one or two EIGER 4M detectors enclosed in a helium chamber to overcome the X-ray absorption caused by air. When two EIGER 4M detectors were used, they were configured with V-shape

geometry at an adjacent tilt angle of 25° . The detector threshold energy was set to half of the X-ray energy for the 1.9 and 2.7 Å experiments, and 2.3 keV for the 3.3 Å experiment, which corresponds to a threshold of 50, 50 and $\sim 61\%$, respectively. BL-1A is equipped with a mini-kappa goniometer with an arm offset of 20° . We used a $40 \times 40 \mu\text{m}$ beam size for all experiments. The flux values are approximately 1.5×10^{11} , 1.2×10^{11} and 1.1×10^{11} photons s^{-1} at 1.9, 2.7 and 3.3 Å, respectively.

2.3. Native-SAD data collection on T₂R-TTL and Sen1 crystals at BL-1A

2.3.1. T₂R-TTL. We collected native-SAD data from a T₂R-TTL crystal of size $500 \times 70 \times 50 \mu\text{m}$ [Fig. 2(a) and Table 1] at wavelengths of 1.9 and 2.7 Å. We collected $14 \times 360^\circ$ datasets at different κ angles (Table 1) at 2.7 Å with 3.4% beam transmission and $21 \times 360^\circ$ datasets at 1.9 Å with 16.5% beam transmission on an EIGER 4M detector placed 60 mm away from the crystal. During data collection, we travelled the longest dimension of the crystal to reduce damage and minimize the systematic errors by introducing fresh crystalline material. We collected all native-SAD datasets from one T₂R-

TTL crystal; the right-hand part of the crystal was collected at 2.7 Å and the left-hand part at 1.9 Å [Fig. 2(a)]. The total doses were estimated as 3.9 MGy and 9.1 MGy for 14 of the 2.7 Å datasets and 21 of the 1.9 Å datasets, respectively.

2.3.2. Sen1. Native-SAD datasets from the helicase protein Sen1 crystal of size $200 \times 100 \times 50 \mu\text{m}$ [Fig. 3(a)] were collected at a wavelength of 2.7 Å on an EIGER 4M detector placed 60 mm away from the crystal. We collected $4 \times 360^\circ$ datasets at different orientations using the mini-kappa goniometer (Table 1). A beam size of $40 \times 40 \mu\text{m}$ and a beam transmission of 3.4% were used. The total accumulated dose was 1.8 MGy.

2.4. Laser-shaping machine and shaping of lysozyme crystals

A compact, fast and user-friendly laser shaping system (Murakami *et al.*, 2004; Kitano *et al.*, 2004, 2005) was developed at RIKEN, SPring-8, Japan, to trim crystals in to various shapes [Fig. 4(a)]. The deep-UV laser, which uses an NSL-193L laser source (Nikon) of wavelength 193 nm and a pulse duration of ~ 1 ns (Kitano *et al.*, 2005), operates at an energy of 8.0 μJ . High-speed scanning galvanometer mirrors focus the beam to $4.6 \times 3.9 \mu\text{m}$ ($H \times V$, FWHM) [Fig. S2(b)]. Crystals

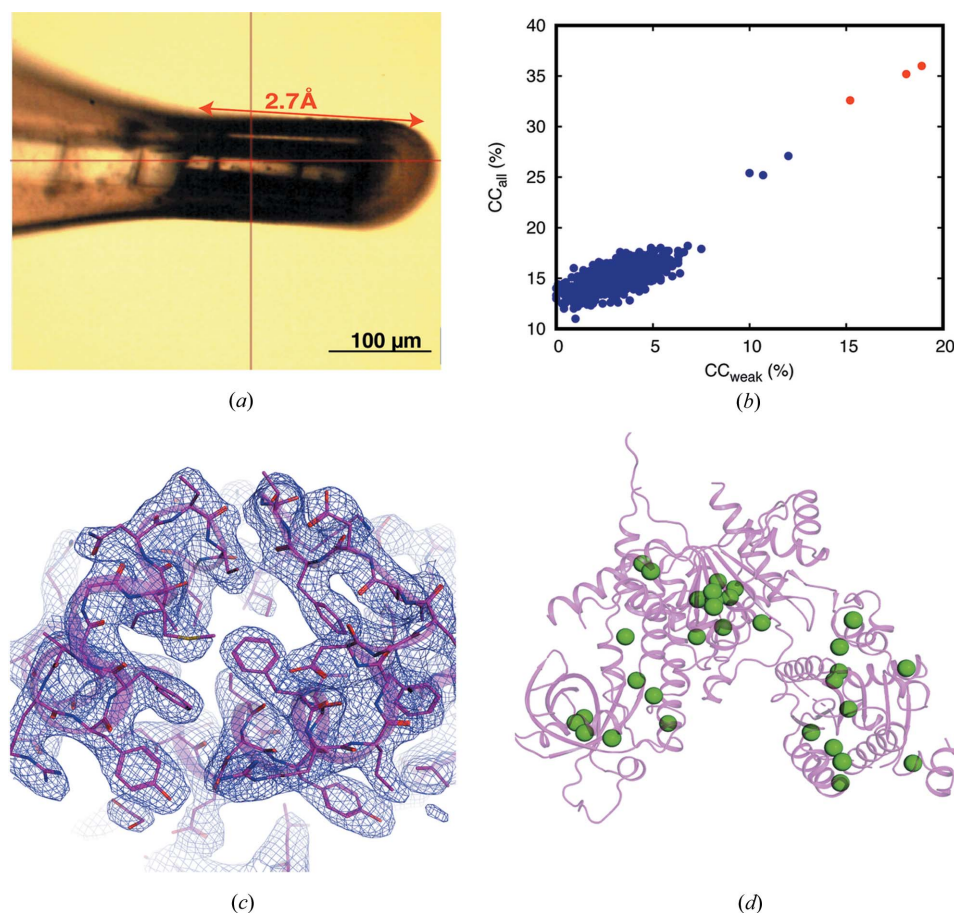


Figure 3

Measurement and native-SAD phasing for Sen1 protein using 2.7 Å. (a) The crystal mounted on an elliptical ActiLoop with the data-collection region marked with a red double-headed arrow. (b) The CC_{all} versus CC_{weak} plot shows the successful substructure determination by *SHELXD*. (c) Experimental phasing map of a selected region of Sen1 after density modification (shown in blue), contoured at 1.0σ along with the C^α trace, produced by *CRANK2* (shown as a light-pink colored cartoon representation). (d) A cartoon representation of Sen1 protein, with anomalous scatterers (*i.e.* S atoms) highlighted as green spheres.

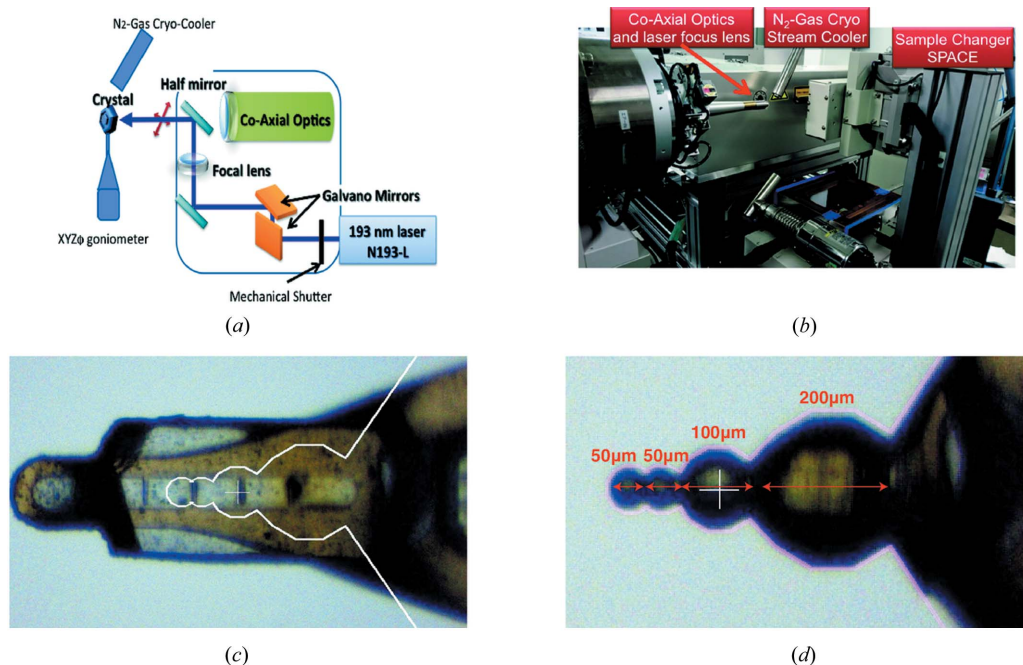


Figure 4 Deep-UV-laser machine setup and laser shaping of lysozyme crystal into spheres of different diameters. (a) Schematic diagram of the deep-UV-laser machine along with focusing optics, goniometer and cryojet. (b) Top-view of the real-life setup of the deep-UV-laser system at SPring-8, Japan. (c) Original large lysozyme crystal of $800 \times 500 \times 400 \mu\text{m}$ (before laser cutting) and the white contour was a template made for a spherical shape with precise diameters of each of the spheres, including the error margin for the laser-cutting process. (d) The same crystal after laser cutting with a deep-UV laser of wavelength 193 nm. The size of each sphere is written in red and the diameters are marked with corresponding red lines. There are four spheres – two of 50 μm , one of 100 μm and one of 200 μm diameter. The part of the crystal on the extreme right is the ‘unshaped’ region, along with the base of the original loop. A supplementary movie of the laser-cutting process is also available in the Supporting information.

were mounted on a high-precision single-axis goniometer with $X/Y/Z$ linear stages using the SPACE sample changer (Murakami *et al.*, 2012) and kept at 100 K under a cryostream [Fig. 4(b)]. We shaped one lysozyme crystal ($800 \times 500 \times 400 \mu\text{m}$) mounted on a MicroLoops ETM (MiTeGen) into four connected spheres with diameters of 50, 50, 100 and 200 μm [Figs. 4(c) and 4(d), and Supplementary movie S1] and another lysozyme crystal into a cylindrical shape of $500 \times 50 \times 50 \mu\text{m}$. The procedure took about 20 min per crystal.

Irradiation damage of the deep-UV laser was evaluated with a micro-focused X-ray beam at beamline BL32XU at SPring-8, Japan (Hirata *et al.*, 2013) using a cytochrome *c* oxidase crystal (Tsukihara *et al.*, 1995). The crystal was shaped using the deep-UV system with lines [Figs. S2(a) and S2(b)] and then rastered using a beam of $1.0 \times 5.0 \mu\text{m}$ ($H \times V$, FWHM) at a wavelength of 1.0 \AA and a flux of 6.0×10^9 photons s^{-1} . The diffraction images were processed using SHIKA (Hirata *et al.*, 2014). We observed a loss of diffraction over an $\sim 10 \mu\text{m}$ thick area [Fig. S2(c)], which implies that the radiation damage extends by $\sim 3 \mu\text{m}$ on each side of the deep-UV laser-beam footprint.

2.5. Dose-normalization measurement

Dose-normalization analysis was carried out using a cylindrical lysozyme crystal that had been shaped by the laser [Fig. S3(a)] of size $550 \times 50 \times 50 \mu\text{m}$, to compare flux and beam transmission between the 2.7 and 3.3 \AA wavelengths.

The X-ray dose was estimated based on intensity decays as measured by the relative B factors, followed by linear curve fitting. At 2.7 \AA , $8 \times 360^\circ$ datasets were collected with 12.4% beam transmission per crystal position. Each dataset ($1 \times 360^\circ$) was collected with 0.2° and 0.1 s per step, corresponding to an accumulated dose of 3.87 MGy [Fig. S3(b)]. We repeated the same experiment at 3.3 \AA with 2.27% beam transmission at a different crystal position [Fig. S3(a)]. Here, the accumulated dose was 0.97 MGy per 360° dataset [Fig. S3(c)]. The dose ratio between the two wavelengths (2.7 versus 3.3 \AA) was therefore ~ 4 . Thus, $12.4/4 = 3.1\%$ beam transmission at 2.7 \AA and 2.27% beam transmission at 3.3 \AA , which should deposit similar doses on the sample. These beam transmissions were used in the subsequent experiments where anomalous diffraction efficiencies at 2.7 and 3.3 \AA were compared.

2.6. Data collection on laser-shaped lysozyme crystals

A lysozyme crystal was mounted on the mini-kappa goniometer at beamline BL-1A, KEK Photon Factory, Japan (Fig. 4). The shaped crystal consisted of four connected spheres: two 50, one 100 and one 200 μm in diameter. Datasets with comparable dose were collected with the bottom EIGER 4M detector in a V-shape configuration at wavelengths of 2.7 and 3.3 \AA with beam transmission of 3.10 and 2.27%, respectively. Two 360° datasets were collected from each of the 50 μm and each of the 100 μm diameter spheres, while only one 360° dataset was collected from the 200 μm diameter sphere at each

wavelength. The accumulated doses per dataset at both wavelengths were about 0.9, 0.45 and 0.225 MGy per dataset for 50, 100 and 200 μm spheres, respectively. All the 2.7 \AA datasets were collected after the 3.3 \AA datasets. All data were collected using oscillation steps of 0.2° and an exposure time of 0.1 s per step.

2.7. Data processing, phasing and refinement

All diffraction data were processed with *XDS* (Kabsch, 2010a) and scaled with *XSCALE* (Kabsch, 2010b). Anomalous data were analyzed with *SHELXC/D/E* (Sheldrick, 2010) using the *HKL2MAP* interface (Pape & Schneider, 2004). The substructure determination was performed using *SHELXD* (Schneider & Sheldrick, 2002), followed by phasing, density modification and automatic model building using *CRANK2* (Skubák & Pannu, 2013). The final refinements were carried out using *phenix.refine* (Afonine *et al.*, 2012). The anomalous peak heights were calculated by *ANODE* (Thorn & Sheldrick, 2011).

2.7.1. T₂R-TTL. For T₂R-TTL, both 1.9 and 2.7 \AA datasets were processed with *XDS*. The multiple datasets at each wavelength were scaled with *XSCALE*. To have common reflections for a direct comparison between datasets at two wavelengths, all data were truncated to 2.95 \AA resolution, which resulted in nearly the same unique reflections (only 0.4% difference) (Tables S1 and S2). The structure was solved from $14 \times 360^\circ$ datasets collected at 2.7 \AA . The substructure was determined using *SHELXD* (Schneider & Sheldrick, 2002) by searching for 100 sites at a resolution cutoff of 3.5 \AA with an E_{min} value of 1.3 and 10 000 trials. This yielded a CFOM of 52.3% ($\text{CC}_{\text{all}} = 38.3\%$ and $\text{CC}_{\text{weak}} = 13.9\%$). The substructure sites were parsed to the *CRANK2* pipeline (Skubák & Pannu, 2013), which completed the sites and carried out phasing, density modification and automatic model building. The final structure was refined at 2.95 \AA resolution in *phenix.refine* (Afonine *et al.*, 2012) with resulting R_{work} of 17.0% and R_{free} of 20.8% (Table 1).

2.7.2. Sen1. For Sen1, the crystal diffracted to 2.8 \AA at a wavelength of 2.7 \AA . The structure was determined from $4 \times 360^\circ$ datasets. *SHELXD* successfully produced a substructure of 22 sites at a resolution cutoff of 3.3 \AA with 1000 tries, resulting in a CFOM of 54.9% ($\text{CC}_{\text{all}} = 36.0\%$ and $\text{CC}_{\text{weak}} = 18.9\%$). *CRANK2* (Skubák & Pannu, 2013) automatically built 692 out of 720 residues. The final refinement of Sen1 structure was performed at 2.95 \AA resolution with *phenix.refine* (Afonine *et al.*, 2012), resulting in final R_{work} of 16.9% and R_{free} of 21.3% (Table 1).

2.7.3. Laser-shaped lysozyme. For laser-shaped lysozyme, a V-shaped detector configuration allowed diffraction resolutions of 2.3 and 2.8 \AA at wavelengths of 2.7 and 3.3 \AA , respectively. During data processing in *XDS* (Kabsch, 2010b), `STRICT_ABSORPTION_CORRECTION` was set to `TRUE`, `AIR` was set to `ZERO` and we manually defined a mask to eliminate shadowed regions caused by overlap between two adjacent detectors in a V configuration. Only the data from the bottom detector were used in data analysis. In the study of the effect

of sample thickness at each wavelength, data were used to the full resolution (Tables S3 and S4). In the direct comparison between 2.7 and 3.3 \AA datasets, only common reflections to 2.8 \AA were used. These reflections were selected using a custom script from the unmerged data in *INTEGRATE.HKL* before scaling by the *CORRECT* routine in *XDS* (Table S5).

3. Results

3.1. Dose-normalized intensity across wavelength

To study the optimal wavelength for native-SAD phasing, the measured anomalous signal per absorbed X-ray dose needs to be compared at different wavelengths. This dose-normalized anomalous efficiency [Appendix A and equation (4)] can be approximated by a dose-normalized diffracted intensity [Appendix A and equation (3)] multiplied by the anomalous scattering factor (f''), assuming the X-ray dose is proportional to the absorbed photon energy. Equation (4) suggests that 1.9, 2.7 and 3.3 \AA wavelengths are optimal for a crystal size of >200 , 125 and 75 μm , respectively, under ideal experimental conditions (*i.e.* no surrounding solvent, loop or air, perfect X-ray beam and detector) [Fig. 1(a)]. The intensity in equation (3) can be used to calculate the theoretical intensity ratio at different wavelengths with an equivalent dose. For example, the expected intensity ratio is 1.16 for a 70 μm thick crystal between 1.9 and 2.7 \AA , and is 1.15 for a 50 μm thick crystal between 2.7 and 3.3 \AA . These intensity ratios were compared with experimentally observed intensity ratios between two wavelengths – 1.9 versus 2.7 \AA (*i.e.* T₂R-TTL) and 2.7 versus 3.3 \AA (*i.e.* laser-shaped lysozyme spheres) – to validate the dose normalization.

To study the crystal size dependence of native SAD at a given wavelength using spherically shaped lysozyme crystals, the theoretical diffracted intensity was estimated by calculating both diffraction volume and absorption correction numerically (Appendix B). Then the theoretical diffracted-intensity ratios across various thicknesses of the crystal were calculated (Table S6) and compared with experimentally observed intensity ratios across different diameters of the laser-shaped lysozyme spheres in Section 3.3.

3.2. Advantage of a wavelength of 2.7 \AA over 1.9 \AA for 100 μm or smaller crystals

3.2.1. T₂R-TTL. We used the tubulin complex T₂R-TTL (Prota *et al.*, 2013) to assess the advantages of native SAD at 2.7 \AA . A T₂R-TTL needle-shaped crystal ($500 \times 70 \times 50 \mu\text{m}$) was mounted on an elliptical ActiLoop (Molecular Dimensions) with minimum surrounding solvent [Fig. 2(a)]. Using different crystal orientations (Weinert *et al.*, 2015), $21 \times 360^\circ$ and $14 \times 360^\circ$ datasets were collected on two crystal positions using 1.9 and 2.7 \AA , respectively [Fig. 2(a)]. We used the ratio of observed diffraction intensities between the two wavelengths to achieve a dose-normalized comparison. The mean intensity ratio of the two 360° datasets at the two wavelengths was 1.8 (Fig. S4), while the theoretical dose-normalized intensity ratio was estimated to be 1.16 (Appendix A).

Therefore, $9 \times 360^\circ$ datasets at 1.9 Å and $14 \times 360^\circ$ datasets at 2.7 Å had a comparable dose [Fig. 2(b)]. We observed the expected Bijvoet ratio (1.5 and 2.8% at 1.9 and 2.7 Å, respectively) in the measured anomalous differences ($\langle |\Delta F| \rangle / \langle F \rangle$) [Fig. 2(c)] and an abrupt rise of $\langle |\Delta F| \rangle / \langle F \rangle$ above 3.5 Å resolution in the 2.7 Å dataset, which does not represent the true anomalous signal but instead indicates that the anomalous signal is buried in the exceeding errors in the weak data at high resolution (Dauter *et al.*, 2002). The corresponding increase of $\langle |\Delta F| \rangle / \sigma(\Delta F)$ is also in a good agreement with the $\sim 86\%$ gain in anomalous scattering factor f'' of sulfur at the two wavelengths [Fig. 2(d)]. In addition, the higher anomalous signal in the 2.7 Å dataset is very visible in the half-dataset anomalous correlation [Fig. 2(e)] and the average anomalous peak heights ($\langle \text{APHs} \rangle$) [Fig. 2(f)]. The merged data at 2.7 and 1.9 Å gave 67 and 40 anomalous peaks above 10σ , respectively [Fig. 2(g)]. Here, the 2.7 Å wavelength data produced successful substructures using *SHELXD* [Fig. 2(h)], and 2152 out of 2363 residues could be built correctly using density modification and automatic model building in *CRANK2* (Skubák & Pannu, 2013). The final structure was refined to 2.95 Å resolution with an R_{work} and an R_{free} of 17.0% and

20.8%, respectively, using *phenix.refine* (Afonine *et al.*, 2012). Native-SAD phasing was also possible at 1.9 Å but only by merging 21 datasets, which constituted more than double the dose used at 2.7 Å. This example clearly illustrates the benefits of performing native SAD at a wavelength of 2.7 Å over 1.9 Å for crystals with 100 μm diameter or less and with minimum extra surrounding materials in the absence of air absorption.

3.2.2. Sen1. We applied native-SAD measurement using a wavelength of 2.7 Å on Sen1: an 85.7 kDa helicase protein with 32 S atoms, involved in the termination of non-coding transcription processes. The Sen1 crystal measured as $220 \times 100 \times 50 \mu\text{m}$ was carefully mounted on an elliptical loop with minimum surrounding solvent [Fig. 3(a)]. Using $4 \times 360^\circ$ datasets collected at multiple crystal orientations and merged together, the substructure was readily solved by *SHELXD* (Schneider & Sheldrick, 2002) using a 3.3 Å resolution cutoff [Fig. 3(b)]. The subsequent density improvement and phasing in *CRANK2* pipeline (Skubák & Pannu, 2013) produced an interpretable map of excellent quality [Figs. 3(c) and S5]. We traced 692 residues successfully and the structure [Fig. 3(d)] was refined to 2.95 Å resolution, resulting in an $R_{\text{work}}/R_{\text{free}}$ of 16.9%/21.3% using *phenix.refine* (Afonine *et al.*, 2012).

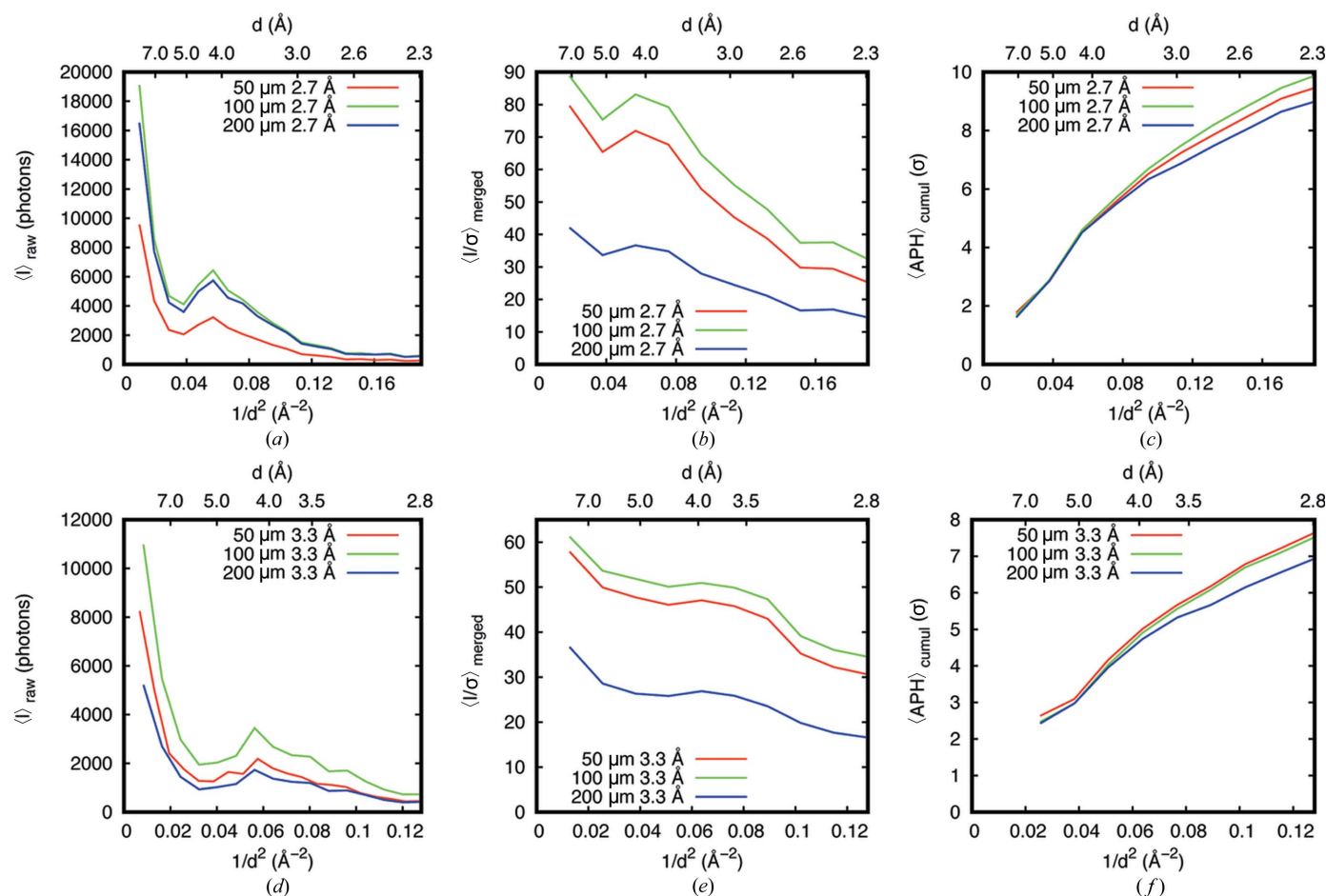


Figure 5 Comparison of data statistics among different spheres of different diameters at wavelengths of 2.7 and 3.3 Å. (a) and (b) Observed diffracted intensities and $\langle I \rangle / \sigma(I)$ over resolution shells at 2.7 Å. (c) Cumulative average anomalous peak height $\langle \text{APH} \rangle$ as a function of resolution at 2.7 Å. (d) and (e) Observed diffracted intensities and $\langle I \rangle / \sigma(I)$ over resolution shells at 3.3 Å. (f) Cumulative $\langle \text{APH} \rangle$ as a function of resolution at 3.3 Å.

3.3. Native SAD with spherical laser-shaped crystals at 2.7 and 3.3 Å

Upon establishing the benefits of using 2.7 Å for native SAD, we next explored the potential of an even longer wavelength of 3.3 Å. Theoretically (Appendix A and Fig. 1), the sample absorption can be detrimental in abstracting accurate anomalous signals at such a wavelength. We there-

fore carried out a systematic study to compare the quality of datasets collected at both 2.7 and 3.3 Å using a lysozyme crystal with various thicknesses. Using a deep-UV laser (*Materials and methods*), we shaped a large lysozyme crystal into connected spheres of 50, 100 and 200 µm diameter (Fig. 4 and Supplementary movie S1). An added benefit of a spherical shaped crystal is that it minimizes the angular dependence of

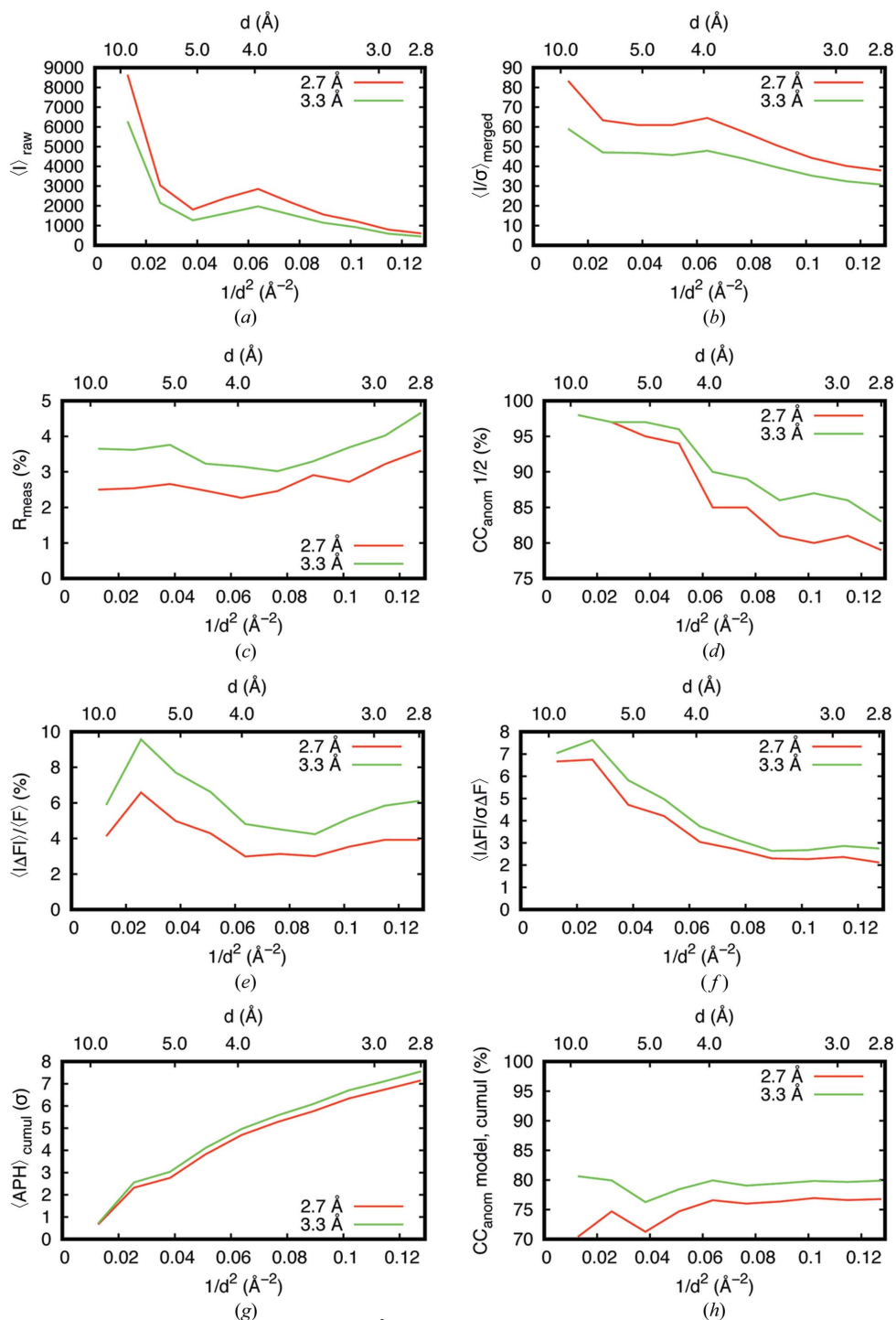


Figure 6 Comparison of data statistics between wavelengths of 2.7 and 3.3 Å on a 50 µm diameter lysozyme sphere. (a)–(f) Observed diffracted intensities (I), $I/\sigma(I)$, R_{meas} , $CC_{\text{anom}}(1/2)$, $\langle |\Delta F| \rangle / \langle F \rangle$ and $\langle |\Delta F| \rangle / \sigma(\Delta F)$ over resolution shells. (g) and (h) Cumulative average anomalous peak height (APH) and correlation coefficient between the observed anomalous difference and the calculated anomalous difference from a refined model as a function of resolution.

absorption (Appendix B). Indeed, the data-processing statistics with and without absorption correction are very similar for the spherical lysozyme crystals except for the 200 μm sphere, where the data quality improved slightly with absorption correction (Fig. S6).

To understand the sample absorption effect, we first compared datasets from 50, 100 and 200 μm spheres at 2.7 \AA . Diffraction intensities are plotted in Fig. 5(a). Their ratios were about 2.0 (100:50 μm) and 1.9 (200:50 μm), in good agreement with the theoretical values of 1.9 and 2.1 (Table S6). However, $\langle I/\sigma(I) \rangle$ values showed very different behavior [Fig. 5(b)], e.g. ratios of $\langle I/\sigma(I) \rangle$ between 100 and 50 μm datasets were less than the expected 1.41 (square root of 2) from Poisson statistics, particularly at low to medium resolutions. The 200 μm dataset had much lower $\langle I/\sigma(I) \rangle$ compared with the other two datasets. A similar trend is also observed in the $\langle \text{APHs} \rangle$ [Fig. 5(c)]. We conclude that the sample absorption for crystals of 100 μm or larger (44% and 70% for 100 and 200 μm crystals, respectively) and the inaccuracy in their corrections in data processing inflate the $\sigma(I)$ estimation, which further diminishes the benefits of the increased diffraction volume. Therefore, in order to profit from the improved f'' at a wavelength of 2.7 \AA for native SAD, crystal size of 100 μm or smaller should be used, as demonstrated here with T₂R-TTL (70–50 μm diameter) and Sen1 (50–100 μm diameter).

Similar analyses were carried out for 3.3 \AA data. The intensity ratios (100:50 μm and 200:50 μm) again followed the theoretical values (observed 1.5 and 0.7 versus theoretical 1.4 and 0.9, respectively) [Fig. 5(d)]. We suspected that the small differences between theoretical and experimentally measured values were caused by a slight miscentering of the X-ray beam with respect to the crystal. As expected, the gain in diffraction volume is further reduced by the excessive absorption at this wavelength (65 and 89% for 100 and 200 μm crystals, respectively). Therefore, the $\langle I/\sigma(I) \rangle$ values only increased slightly between the 50 μm dataset and the 100 μm dataset and both datasets have comparable APHs [Figs. 5(e) and 5(f)]. The $\langle I \rangle$, the $\langle I/\sigma(I) \rangle$ and the anomalous peak height were lowest for the 200 μm dataset. As clearly shown here, small crystals (<50 μm) are a prerequisite to take full benefit of native-SAD phasing at 3.3 \AA and longer wavelength.

We then attempted a comparison between 2.7 and 3.3 \AA using two datasets from the same 50 μm lysozyme sphere collected with a comparable accumulated X-ray dose (see Section 2.5 for experimental dose measurement). The 2.7 \AA dataset had higher observed intensities by about 40% compared with the 3.3 \AA dataset [Fig. 6(a)], which is much higher than the theoretical dose-normalized intensity ratio of 1.15 (Appendix A). The corresponding $\langle I/\sigma(I) \rangle$ was also much higher at 2.7 \AA [Fig. 6(b)]. We attributed these differences to the lower detector efficiency at 3.3 \AA , caused by both energy threshold (61%) and absorption from the non-sensitive surface layers. Indeed, the thin aluminium and silicon layers on the sensor, together with a Mylar window in front of the detector absorb as much as 20–30% more photons at 3.3 \AA than at 2.7 \AA (Donath *et al.*, 2013). In terms of accuracy, the

R_{meas} values of the 3.3 \AA dataset were slightly higher than that of the 2.7 \AA dataset [Fig. 6(c)]. This is likely caused by the so-called ‘corner effect’ from hybrid pixel-array photon-counting detectors (HPCs), which could inflate the R_{meas} by introducing systematic intensity-measurement errors, particularly when the detector energy threshold is above 50% as was the case at 3.3 \AA (Leonarski *et al.*, 2018). In addition, the inaccuracy of absorption correction in data processing could also reduce the accuracy of long-wavelength data.

Nevertheless, the 3.3 \AA dataset featured higher anomalous signal as measured by the half-dataset anomalous correlation [Fig. 6(d)]. The observed $\langle |\Delta F| \rangle / \langle F \rangle$ values were as expected from the Bijvoet ratio estimations (~ 3.5 and $\sim 5\%$ at 2.7 and 3.3 \AA , respectively) [Fig. 6(e)]. The higher anomalous difference at low resolution ($\sim 10\text{--}5 \text{\AA}$) was caused by the enhanced contribution from the four disulfides in lysozyme, unresolved at that resolution. The increase in $\langle |\Delta F| \rangle / \langle F \rangle$ at 3.3 \AA compared with 2.7 \AA was in accordance with the 40% gain in the anomalous scattering factor f'' of sulfur [Fig. 6(e)]. The corresponding $\langle |\Delta F| / \sigma(\Delta F) \rangle$ also increased at 3.3 \AA but to a lower extent [Fig. 6(f)], indicating the higher noise in the 3.3 \AA data as explained earlier. Overall, the $\langle \text{APH} \rangle$ was improved by 0.5 σ to 2.8 \AA resolution [Fig. 6(g)] and correlations between the observed anomalous differences (ΔF_{obs}) and the calculated one (ΔF_{calc}) from the refined model were improved by a few percent [Fig. 6(h)]. We also noticed that the APH for disulfide was clearly lower at 2.7 \AA than at 3.3 \AA (7.1 σ versus 8.1 σ) while the APH was higher for Met at 2.7 \AA (8.7 σ versus 8.1 σ) and comparable for Cl (6.7 σ) at both wavelengths. With a total accumulated dose of $\sim 2\text{--}3 \text{ MGy}$, we attribute this difference to radiation damage on the sensitive disulfide bridges (Murray & Garman, 2002) because the 2.7 \AA dataset was collected after the 3.3 \AA dataset formed the same 50 μm crystal. Therefore, when taking the radiation-damage effect into consideration, the obtained anomalous signal improvement at 3.3 \AA was found to be marginal in this particular experiment. Further improvement in absorption correction and detector performance at such wavelength is needed in order to harness the gain in f'' for native-SAD phasing fully.

4. Discussion

Optimization of native-SAD phasing experiments at wavelengths $> 2 \text{\AA}$ is being addressed at dedicated MX beamlines with reduced air absorption and scattering effects, as well as special detector geometry. However, so far there has been little research into the adverse effect of sample absorption, as well as detector efficiency at such long wavelengths.

In this work, performed at beamline BL-1A at the Photon Factory with two real-life targets T₂R-TTL and Sen1, we have demonstrated that the increased anomalous signal at 2.7 \AA (sulfur $f'' = 1.5 e^-$) over shorter wavelengths can be harnessed effectively as long as the crystal dimension in the beam path is smaller than 100 μm , and extra material around the crystal and air scattering are minimized. Both Sen1 and T₂R-TTL can diffract to about 2.4 \AA resolution, but only data up to 3 \AA were collected and successfully used for phasing. Therefore, native

SAD at a wavelength of 2.7 Å has the potential to reach targets with lower S-atom content and/or lower diffraction resolution where enhanced anomalous signals are needed. While the specialized sample environment at BL-1A is essential, the results reported here also clearly highlight that both the crystal thickness and surrounding materials – loop and solvent – should to be carefully considered prior to data collection at long wavelengths.

Despite an ~40% increase in anomalous signal compared with 2.7 Å ($f'' = 2.1 e^-$ versus $1.5 e^-$), the potential of native SAD at wavelengths of 3.3 Å and beyond is limited in practice to crystals of 50 µm or smaller in size. We have shown in this study that spherical laser-shaping offers an appealing solution to tackle absorption effects by both realizing a fine control of the sample thickness and simplifying absorption-correction procedures. Indeed, the absorption effect is then identical for reflections at a given scattering angle when a spherically homogeneous crystal is illuminated by an X-ray beam with a symmetric profile at the center of the sphere. Here, an added benefit is that the angular dependency of the absorption can be numerically calculated and applied (Appendix B). While X-ray tomography has been attempted to reconstruct the shape and volume of macromolecular crystals (Brockhauser *et al.*, 2008), including the loop and surrounding materials, the correction of the absorption has remained empirical (Blessing, 1995) in most data-processing software suites and relies largely on the collection of data with high multiplicity and in multiple crystal orientations (Liu *et al.*, 2012; Weinert *et al.*, 2015). The possibility to use UV-laser ablation to both remove non-diffracting materials and to shape crystals as spheres is therefore an interesting tool to better deal with absorption effects at long wavelength.

While it may be possible to shape large crystals using UV-laser ablation, microcrystals still remain a challenge for long-wavelength native-SAD phasing. Thanks to recent developments in serial crystallography (SX) methods at both X-ray free-electron lasers and synchrotrons, sample delivery of microcrystals benefits from low scattering background solid-supports, which have been designed to facilitate crystal loading with minimum solvent (Meents *et al.*, 2017; Owen *et al.*, 2017; Huang *et al.*, 2015; Warren *et al.*, 2015; Huang *et al.*, 2016; Wierman *et al.*, 2013; Baxter *et al.*, 2016; Sui *et al.*, 2016). The subsequent serial data collection and data-merging methods have been adapted as well (Zander *et al.*, 2015; Hirata *et al.*, 2013; Wojdyla *et al.*, 2018; Yamashita *et al.*, 2018; Huang *et al.*, 2018; Basu *et al.*, 2019). These developments are particularly relevant to small membrane-protein crystals, for which *de novo* phasing is in demand. The recent native-SAD phasing of the membrane protein PepT_{st} with an SX approach required data from about 2 000 microcrystals (10–20 µm) collected at 6 keV. In contrast, only about 100 Se-Met PepT_{st} crystals of similar size were needed for Se-SAD (Huang *et al.*, 2018). Given the f'' values for S at 4 keV compared with Se at 12.67 keV, ($1.8 e^-$ and $3.8 e^-$ respectively) we estimate an order of magnitude less PepT_{st} microcrystals to be required for solving PepT_{st} by native-SAD phasing at 4 keV or lower energies.

Another obstacle of native-SAD phasing at long wavelength is the detector inefficiency. Detection of low-energy photons is challenging for HPCs – the current standard in MX beamlines. The performance of the EIGER 4M detector used in this study was indeed affected by the lowest reachable energy threshold, inaccuracy in threshold calibration (Leonarski *et al.*, 2018) and absorption from an ~1 µm Al/Si layer on the surface of the silicon sensor, as well as from a protective 20 µm-thick Mylar foil. Unfortunately, the latter effect becomes more pronounced for high-angle reflections because of parallax effects but can be minimized using a curved detector (Wagner *et al.*, 2016) or flat detectors in a V-shape configuration. Note that the faster intensity decay towards a high diffraction angle caused by this parallax can also induce artifacts in data processing, *e.g.* inflating the Wilson B factor. As an alternative to HPC technology, new hybrid charge integrating technology is being developed and is expected to perform better at low energy (Leonarski *et al.*, 2018).

Overall, this work highlights the 2.7 Å wavelength as a very suitable energy with current instrumentation for sample thicknesses ≤ 100 µm when mounted appropriately. In addition, it emphasises that minimization of X-ray absorption by careful sample preparation or accurate control of sample thickness and shape by laser ablation, together with improved detector technology, will be instrumental in realising the full potential of long wavelengths (>3 Å) for solving challenging novel structures using native-SAD phasing.

APPENDIX A

Theory of the optimal wavelength for native SAD

There are theoretical foundations to the optimal X-ray wavelengths for native-SAD phasing (Arndt, 1984; Polikarpov *et al.*, 1997; Hendrickson, 2013; Wagner *et al.*, 2016). Here, we briefly revisit the theory and add considerations on the absorption by both non-crystalline materials around the crystal and the air in the diffraction beam path. Optimal native-SAD data collection primarily relies on two interdependent variables – the materials in the X-ray beam path and the X-ray wavelength. The diffraction efficiency (I_E) defined as the integrated intensity per absorbed energy in a crystal bathed in an X-ray beam (Arndt, 1984; Polikarpov *et al.*, 1997) is expressed as

$$I_E \propto t_{\text{xtal}}^3 \lambda^3 \frac{\exp(-\mu_{\text{xtal}} t_{\text{xtal}})}{1 - \exp(-\mu_{\text{xtal}} t_{\text{xtal}})}, \quad (1)$$

where t_{xtal} is the crystal dimension, λ is the X-ray wavelength, and μ_{xtal} is the linear absorption coefficient and is estimated to be $0.32 \lambda^3 \text{ mm}^{-1}$ for a lysozyme crystal. The $\exp(-\mu_{\text{xtal}} t_{\text{xtal}})$ term is the X-ray transmittance of the path through the crystal. The $1 - \exp(-\mu_{\text{xtal}} t_{\text{xtal}})$ term is the absorption, which is approximately related to the X-ray induced radiation damage. For the scenario where the X-ray beam is smaller than the crystal, I_E becomes

$$I_E \propto t_{\text{beam}}^2 t_{\text{xtal}} \lambda^3 \frac{\exp(-\mu_{\text{xtal}} t_{\text{xtal}})}{1 - \exp(-\mu_{\text{xtal}} t_{\text{xtal}})}, \quad (2)$$

where t_{beam} is the X-ray beam dimension. This case is close to experiments described in this study and its detailed treatment for spherical crystals is given in Appendix B.

Next, we include the solvent around the crystal and the air in the X-ray path from the crystal to the detector,

$$I_E \propto t_{\text{xtal}} \lambda^3 \frac{\exp[-\mu_{\text{xtal}}(t_{\text{xtal}} + t_{\text{sol}})]}{1 - \exp(-\mu_{\text{xtal}} t_{\text{xtal}})} \exp(-\mu_{\text{air}} t_{\text{air}}), \quad (3)$$

where t_{sol} is thickness caused by surrounding solvent/mother liquor around the crystal (we assume that crystal and solvent have similar linear absorption coefficient), μ_{air} is the linear absorption coefficient of air ($3.3 \times 10^{-4} \text{ mm}^{-1} \text{ \AA}^{-3}$), and t_{air} is the path length of air between the crystal and the detector. The t_{beam}^2 term is constant for this experiment and was removed from equation (3).

Based on equation (3), the expected intensity ratio is 1.16 for a given dose between wavelengths of 1.9 and 2.7 Å using an ~70 µm thick crystal of T₂R-TTL. Similarly, for a 50 µm diameter spherical lysozyme crystal that had been shaped by the laser, the expected dose-normalized intensity ratio between wavelengths of 2.7 and 3.3 Å is 1.15.

When searching for the optimal wavelength for native-SAD phasing, anomalous diffraction efficiency (ΔI) is used as a metric and is defined as the diffraction efficiency multiplied by the anomalous scattering factor f'' (Hendrickson, 2013; Wagner *et al.*, 2016; Liebschner *et al.*, 2016),

$$\Delta I = I_E f'' \propto t_{\text{xtal}} \lambda^3 \frac{\exp[-\mu_{\text{xtal}}(t_{\text{xtal}} + t_{\text{sol}})]}{1 - \exp(-\mu_{\text{xtal}} t_{\text{xtal}})} \times \exp(-\mu_{\text{air}} t_{\text{air}}) f''. \quad (4)$$

Based on equation (4), anomalous diffraction efficiencies were visualized in 2D contour plots as a function of crystal size and X-ray wavelength with or without air and solvent around the crystal [Figs. 1(a)–1(d) and S1(a)–S1(d)].

APPENDIX B

Theoretical calculation of diffraction volume and absorption correction of spherical crystals

The mean intensity is related to diffraction volume and dose absorbed by a spherical crystal by (Holton & Frankel, 2010)

$$\langle I(\theta) \rangle \simeq I_0 V_{\text{xtal}} \lambda^3 \langle A(\theta) \rangle, \quad (5)$$

$$\langle A(\theta) \rangle = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{V_{\text{xtal}}} \iiint_{V_{\text{xtal}}} \exp[-\mu T(x, y, z, \theta, \phi)] dx dy dz d\phi \quad (6)$$

and

$$V_{\text{xtal}} = \iiint_{V_{\text{xtal}}} 1 dx dy dz. \quad (7)$$

I_0 = incident X-ray flux density.

V_{xtal} = diffraction volume or illuminated volume, defined by equation (5).

λ = X-ray wavelength.

μ = absorption coefficient (6.31 mm^{-1} for 2.7 Å and 11.5 mm^{-1} for 3.3 Å).

$T(x, y, z, \theta, \phi)$ = X-ray path length in crystal, *i.e.* thickness along incident beam (T1) and diffracted beam directions (T2) combined at a given scattering angle of θ . Here, ϕ stands for the rotation of the diffracted beam.

$\langle A(\theta) \rangle$ = mean transmittance term for spherical crystal at a given scattering angle of θ .

r = radius of the spherical crystal (µm).

b = X-ray beam size (µm).

We calculated the linear absorption coefficient (*i.e.* μ) of lysozyme using *RADDose-3D* (Zeldin *et al.*, 2013). It implied, $\mu (\lambda \rightarrow 3.3 \text{ \AA}) = 1.15 \times 10^{-2} \text{ \AA}^{-3}$ or 11.5 mm^{-1} and $\mu (\lambda \rightarrow 2.7 \text{ \AA}) = 6.31 \times 10^{-3} \text{ \AA}^{-3}$ or 6.31 mm^{-1} . Based on equations (3) and (4) and analytical μ values, a numerical integration method was adopted to precisely account for absorption effect in a spherical crystal that had been shaped by the laser and absorption-corrected mean intensity at a given scattering angle was calculated.

As shown in Fig. S7, we modelled the crystal as a homogeneous sphere of radius r . The center of the sphere (O) defines the origin of the coordinate system and the X-ray beam from the synchrotron is placed on the Z axis. The square X-ray beam with height and width equal to b and of top-hat profile was assumed to be centered with the center of the sphere.

Equation (5) refers to two competing terms – V_{xtal} and $\langle A(\theta) \rangle$. Since the illuminated volume is the intersection of sphere (crystal) and cuboid (top-hat beam), as presented in Fig. S7, it is difficult to analytically calculate volume integrals from equations (6) and (7). Instead integrals were calculated numerically with *Mathematica* (Wolfram Research Inc.). Assuming that diffraction happened at point (x, y, z) inside the illuminated volume, we can calculate the optical path of the X-ray beam inside the crystal $T(x, y, z, \theta, \phi)$ as the sum of two terms – the path along the incident beam to the point where the X-ray beam intersects inside the crystal [T1 in Fig. S7(b)] and the path along the diffracted beam from the point of interaction to the point where it exits from the spherical crystal [T2 in Fig. S7(b)] (Becker & Coppens, 1974). The length of T2 depends upon the scattering angle (θ) of the diffracted beam. In addition, because of rotation of the spherical crystal, each reflection or diffracted vector will rotate 360°, which is accounted for in equation (4) as the ϕ term. In order to calculate the $T(x, y, z, \theta, \phi)$ term, we used a geometric formula for intersection between a line and a sphere. In practice, the volume integral over V_{xtal} is calculated numerically as a triple integral with $(-b/2, b/2)$, $(-b/2, b/2)$ and $(-r, r)$ as limits for x , y and z , respectively, but with the integral content multiplied by an extra term that equals 1, if point (x, y, z) is inside V_{xtal} and 0 if point (x, y, z) is outside V_{xtal} .

The theoretically calculated diffracted intensities and the corresponding ratios between different sizes of spherical crystals at different scattering angles are provided in Table S6 for wavelengths of 2.7 and 3.3 Å. The example *Mathematica*

script for absorption-corrected intensity for 50 μm spherical crystal at 0° scattering angle, $I(0^0)$, is also provided in the Supporting information.

Acknowledgements

This work was performed at the KEK BL-1A beamline, Japan, as well as at the X06DA beamline at the Swiss Light Source, Switzerland. Laser ablation was performed at SPring-8, Japan.

Funding information

NM and YY were supported by the 'Platform Project for Supporting Drug Discovery and Life Science Research [Platform for Drug Discovery, Informatics and Structural Life Science (PDIS)]' from the Japan Agency for Medical Research and Development (AMED). YK and MY were partially supported by 'Target Proteins Research Program (TPRP)' grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) and the 'Platform Project for Supporting Drug Discovery and Life Science Research [PDIS]' and 'Platform Project for Supporting Drug Discovery and Life Science Research [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)]' from AMED under Grant Number JP18am0101070. C-YH is partially supported by the European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska-Curie grant agreement No. 701647.

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst. D* **68**, 352–367.
- Alag, R., Bharatham, N., Dong, A., Hills, T., Harikishore, A., Widjaja, A. A., Shochat, S. G., Hui, R. & Yoon, H. S. (2009). *Protein Sci.* **18**, 2115–2124.
- Arndt, U. W. (1984). *J. Appl. Cryst.* **17**, 118–119.
- Aurelius, O., Duman, R., El Omari, K., Mykhaylyk, V. & Wagner, A. (2017). *Nucl. Instrum. Methods Phys. Res. B*, **411**, 12–16.
- Austin, H. P., Allen, M. D., Donohoe, B. S., Rorrer, N. A., Kearns, F. L., Silveira, R. L., Pollard, B. C., Dominick, G., Duman, R., El Omari, K., Mykhaylyk, V., Wagner, A., Michener, W. E., Amore, A., Skaf, M. S., Crowley, M. F., Thorne, A. W., Johnson, C. W., Woodcock, H. L., McGeehan, J. E. & Beckham, G. T. (2018). *Proc. Natl Acad. Sci.* **115**, E4350–E4357.
- Banerjee, S., Coussens, N. P., Gallat, F.-X., Sathyanarayanan, N., Srikanth, J., Yagi, K. J., Gray, J. S. S., Tobe, S. S., Stay, B., Chavas, L. M. G. & Ramaswamy, S. (2016). *IUCrJ*, **3**, 282–293.
- Basu, S., Kaminski, J. W., Panepucci, E., Huang, C.-Y., Warshamanage, R., Wang, M. & Wojdyla, J. A. (2019). *J. Synchrotron Rad.* **26**, 244–252.
- Baxter, E. L., Aguila, L., Alonso-Mori, R., Barnes, C. O., Bonagura, C. A., Brehmer, W., Brunger, A. T., Calero, G., Caradoc-Davies, T. T., Chatterjee, R., Degrado, W. F., Fraser, J. S., Ibrahim, M., Kern, J., Kobilka, B. K., Kruse, A. C., Larsson, K. M., Lemke, H. T., Lyubimov, A. Y., Manglik, A., McPhillips, S. E., Norgren, E., Pang, S. S., Soltis, S. M., Song, J., Thomaston, J., Tsai, Y., Weis, W. I., Woldeyes, R. A., Yachandra, V., Yano, J., Zouni, A. & Cohen, A. E. (2016). *Acta Cryst. D* **72**, 2–11.
- Becker, P. J. & Coppens, P. (1974). *Acta Cryst. A* **30**, 129–147.
- Bent, A. F., Mann, G., Housen, W. E., Mykhaylyk, V., Duman, R., Thomas, L., Jaspars, M., Wagner, A. & Naismith, J. H. (2016). *Acta Cryst. D* **72**, 1174–1180.
- Blessing, R. H. (1995). *Acta Cryst. A* **51**, 33–38.
- Brockhauser, S., Di Michiel, M., McGeehan, J. E., McCarthy, A. A. & Ravelli, R. B. G. (2008). *J. Appl. Cryst.* **41**, 1057–1066.
- Chen, L., Chen, L.-R., Zhou, X. E., Wang, Y., Kahsai, M. A., Clark, A. T., Edmondson, S. P., Liu, Z.-J., Rose, J. P., Wang, B.-C., Meehan, E. J. & Shriver, J. W. (2004). *J. Mol. Biol.* **341**, 73–91.
- Cianci, M., Groves, M. R., Barford, D. & Schneider, T. R. (2016). *Acta Cryst. D* **72**, 403–412.
- Cruz, M. J. de la, Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat. Methods*, **14**, 399–402.
- Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst. D* **58**, 494–506.
- Deng, L., Vysotski, E. S., Markova, S. V., Liu, Z.-J., Lee, J., Rose, J. & Wang, B.-C. (2005). *Protein Sci.* **14**, 663–675.
- Djinovic Carugo, K., Helliwell, J. R., Stuhmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
- Donath, T., Brandstetter, S., Cibik, L., Commichau, S., Hofer, P., Krumrey, M., Lüthi, B., Marggraf, S., Müller, P., Schneebeli, M., Schulze-Briese, C. & Wernecke, J. (2013). *J. Phys. Conf. Ser.* **425**, 062001.
- Esposito, D., Günster, R. A., Martino, L., El Omari, K., Wagner, A., Thurston, T. L. M. & Rittinger, K. (2018). *J. Biol. Chem.* **293**, 5064–5078.
- Gentry, H. R., Singer, A. U., Betts, L., Yang, C., Ferrara, J. D., Sondek, J. & Parise, L. V. (2005). *J. Biol. Chem.* **280**, 8407–8415.
- Hendrickson, W. A. (2013). *Acta Cryst. A* **69**, 51–59.
- Hendrickson, W. A. (2014). *Q. Rev. Biophys.* **47**, 49–93.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature*, **290**, 107–113.
- Hirata, K., Kawano, Y., Ueno, G., Hashimoto, K., Murakami, H., Hasegawa, K., Hikima, T., Kumasaka, T. & Yamamoto, M. (2013). *J. Phys. Conf. Ser.* **425**, 012002.
- Hirata, K., Kawano, Y., Yamashita, K., Ueno, G., Hikima, T., Murakami, H., Hasegawa, K., Baba, S., Kumasaka, T. & Yamamoto, M. (2014). *Acta Cryst. A* **70**, C333–C333.
- Holton, J. M. & Frankel, K. A. (2010). *Acta Cryst. D* **66**, 393–408.
- Huang, C.-Y., Olieric, V., Howe, N., Warshamanage, R., Weinert, T., Panepucci, E., Vogeley, L., Basu, S., Diederichs, K., Caffrey, M. & Wang, M. (2018). *Commun. Biol.* **1**.
- Huang, C.-Y., Olieric, V., Ma, P., Howe, N., Vogeley, L., Liu, X., Warshamanage, R., Weinert, T., Panepucci, E., Kobilka, B., Diederichs, K., Wang, M. & Caffrey, M. (2016). *Acta Cryst. D* **72**, 93–112.
- Huang, C.-Y., Olieric, V., Ma, P., Panepucci, E., Diederichs, K., Wang, M. & Caffrey, M. (2015). *Acta Cryst. D* **71**, 1238–1256.
- Kabsch, W. (2010a). *Acta Cryst. D* **66**, 133–144.
- Kabsch, W. (2010b). *Acta Cryst. D* **66**, 125–132.
- Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I. & Yao, M. (2008). *J. Biol. Chem.* **283**, 36328–36337.
- Kitano, H., Adachi, H., Sato, A., Murakami, A., Matsumura, H., Takano, K., Inoue, T., Mori, Y., Doi, M. & Sasaki, T. (2004). *Jpn. J. Appl. Phys.* **43**, L1271–L1274.
- Kitano, H., Matsumura, H., Adachi, H., Murakami, S., Takano, K., Inoue, T., Mori, Y., Doi, M. & Sasaki, T. (2005). *Jpn. J. Appl. Phys.* **44**, L54–L56.
- Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O. & Conti, E. (2017). *EMBO J.* **36**, 1590–1604.
- Leonarski, F., Redford, S., Mozzanica, A., Lopez-Cuenca, C., Panepucci, E., Nass, K., Ozerov, D., Vera, L., Olieric, V., Buntschu, D., Schneider, R., Tinti, G., Froejdh, E., Diederichs, K., Bunk, O., Schmitt, B. & Wang, M. (2018). *Nat. Methods*, **15**, 799–804.
- Liebschner, D., Yamada, Y., Matsugaki, N., Senda, M. & Senda, T. (2016). *Acta Cryst. D* **72**, 728–741.
- Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancía, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.

- Liu, Q. & Hendrickson, W. A. (2015). *Curr. Opin. Struct. Biol.* **34**, 99–107.
- Liu, Q. & Hendrickson, W. A. (2017). *Protein Crystallography: Methods and Protocols*, Vol. 1607, ch.16, *Methods in Molecular Biology*, edited by A. Wlodawer, Z. Dauter & M. Jaskolski, pp. 377–399. New York: Springer.
- Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* **D69**, 1314–1332.
- Liu, Z.-J., Vysotski, E. S., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- Meents, A., Wiedorn, M. O., Srajer, V., Henning, R., Sarrou, I., Bergtholdt, J., Barthelmess, M., Reinke, P. Y. A., Dierksmeyer, D., Tolstikova, A., Schaible, S., Messerschmidt, M., Ogata, C. M., Kissick, D. J., Taft, M. H., Manstein, D. J., Lieske, J., Oberthuer, D., Fischetti, R. F. & Chapman, H. N. (2017). *Nat. Commun.* **8**, 1281.
- Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* **D63**, 366–380.
- Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* **D61**, 1263–1272.
- Murakami, A., Kitano, H., Adachi, H., Matsumura, H., Takano, K., Inoue, T., Mori, Y., Doi, M. & Sasaki, T. (2004). *Jpn J. Appl. Phys.* **43**, L873–L876.
- Murakami, H., Ueno, G., Shimizu, N., Kumasaka, T. & Yamamoto, M. (2012). *J. Appl. Cryst.* **45**, 234–238.
- Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
- Olieric, V., Weinert, T., Finke, A. D., Anders, C., Li, D., Olieric, N., Borca, C. N., Steinmetz, M. O., Caffrey, M., Jinek, M. & Wang, M. (2016). *Acta Cryst.* **D72**, 421–429.
- Owen, R. L., Axford, D., Sherrell, D. A., Kuo, A., Ernst, O. P., Schulz, E. C., Miller, R. J. D. & Mueller-Werkmeister, H. M. (2017). *Acta Cryst.* **D73**, 373–378.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Polikarpov, I., Teplyakov, A. & Oliva, G. (1997). *Acta Cryst.* **D53**, 734–737.
- Prota, A. E., Bargsten, K., Zurwerra, D., Field, J. J., Díaz, J. F., Altmann, K.-H. & Steinmetz, M. O. (2013). *Science*, **339**, 587–590.
- Read, J. A., Winter, V. J., Eszes, C. M., Sessions, R. B. & Brady, R. L. (2001). *Proteins*, **43**, 175–185.
- Rose, J. P., Wang, B.-C. & Weiss, M. S. (2015). *IUCrJ*, **2**, 431–440.
- Ru, H., Zhao, L., Ding, W., Jiao, L., Shaw, N., Liang, W., Zhang, L., Hung, L.-W., Matsugaki, N., Wakatsuki, S. & Liu, Z.-J. (2012). *Acta Cryst.* **D68**, 521–530.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Skubák, P. & Pannu, N. S. (2013). *Nat. Commun.* **4**, 2777.
- Sui, S., Wang, Y., Kolewe, K. W., Srajer, V., Henning, R., Schiffman, J. D., Dimitrakopoulos, C. & Perry, S. L. (2016). *Lab Chip*, **16**, 3082–3096.
- Thorn, A. & Sheldrick, G. M. (2011). *J. Appl. Cryst.* **44**, 1285–1287.
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1995). *Science*, **269**, 1069–1074.
- Wagner, A., Duman, R., Henderson, K. & Mykhaylyk, V. (2016). *Acta Cryst.* **D72**, 430–439.
- Warren, A. J., Crawshaw, A. D., Trincao, J., Aller, P., Alcock, S., Nistea, I., Salgado, P. S. & Evans, G. (2015). *Acta Cryst.* **D71**, 2079–2088.
- Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Rose, J., Ebihara, A., Kuramitsu, S., Li, D., Howe, N., Schnapp, G., Pautsch, A., Bargsten, K., Prota, A. E., Surana, P., Kottur, J., Nair, D. T., Basilico, F., Cecatiello, V., Pasqualato, S., Boland, A., Weichenrieder, O., Wang, B.-C., Steinmetz, M. O., Caffrey, M. & Wang, M. (2015). *Nat. Methods*, **12**, 131–133.
- Wierman, J. L., Alden, J. S., Kim, C. U., McEuen, P. L. & Gruner, S. M. (2013). *J. Appl. Cryst.* **46**, 1501–1507.
- Wojdyla, J. A., Kaminski, J. W., Panepucci, E., Ebner, S., Wang, X., Gabadinho, J. & Wang, M. (2018). *J. Synchrotron Rad.* **25**, 293–303.
- Yamashita, K., Hirata, K. & Yamamoto, M. (2018). *Acta Cryst.* **D74**, 441–449.
- Zander, U., Bourenkov, G., Popov, A. N., de Sanctis, D., Svensson, O., McCarthy, A. A., Round, E., Gordeliy, V., Mueller-Dieckmann, C. & Leonard, G. A. (2015). *Acta Cryst.* **D71**, 2328–2343.
- Zeldin, O. B., Brockhauser, S., Bremridge, J., Holton, J. M. & Garman, E. F. (2013). *Proc. Natl Acad. Sci.* **110**, 20551–20556.