This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting.
All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings

Faraday Discussions   Volume 164

Tropospheric Aerosol – Formation, Transformation, Fate and Impacts

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Signal to noise considerations for single crystal femtosecond time resolved crystallography of the Photoactive Yellow Protein

Jasper J. van Thor,[a]* Mark M. Warren,[a] Craig N. Lincoln,[a]

5 **Matthieu Chollet,[b] Henrik Till Lemke,[b] David M. Fritz,[b] Marius Schmidt,[c] Jason Tenboer,[c] Zhong Ren,[e] Vukica Srajer,[e] Keith Moffat[d,e] and Tim Graber[e,f]**

[a] *Imperial College London, Division of Molecular Biosciences, South Kensington Campus, London SW7 2AZ*

10 [b] *LCLS, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*

[c] *Department of Physics, University of Wisconsin-Milwaukee, 1900 E. Kenwood Blvd, Milwaukee, WI 53211, USA*

[d] *Department of Biochemistry and Molecular Biology, and Institute for Biophysical Dynamics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, United States*

15 [e] *Center for Advanced Radiation Sources, The University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, United States*

[f] *Current address: Washington State University c/o Argonne National Laboratory Bld. 438F 9700 S. Cass AveArgonne, IL 60439*

20 Femtosecond time resolved pump-probe protein X-ray crystallography requires highly accurate measurements of the photoinduced structure factor amplitude differences. In the case of femtosecond photolysis of single $P6_3$ crystals of the Photoactive Yellow Protein, it is shown that photochemical dynamics place a considerable restraint on the achievable time resolution
25 due to the requirement to stretch and add second order dispersion in order to generate threshold concentration levels in the interaction region. Here, we report on using a 'quasi-cw' approach to use the rotation method with monochromatic radiation and 2 eV bandwidth at 9.465 keV at the Linac Coherent Light Source operated in SASE mode. A source of significant
30 Bragg reflection intensity noise is identified from the combination of mode structure and jitter with very small mosaic spread of the crystals and very low convergence of the XFEL source. The accuracy with which the three dimensional reflection is approximated by the 'quasi-cw' rotation method with the pulsed source is modelled from the experimentally collected X-ray
35 pulse intensities together with the measured rocking curves. This model is extended to predict merging statistics for recently demonstrated self seeded mode generated pulse train with improved stability, in addition to extrapolating to single crystal experiments with increased mosaic spread. The results show that the noise level can be adequately modelled in this
40 manner, indicating that the large intensity fluctuations dominate the merged signal-to-noise ($I/\sigma I$) value. Furthermore, these results predict that using the self seeded mode together with more mosaic crystals, sufficient accuracy may be obtained in order to resolve typical photoinduced structure factor amplitude differences, as taken from representative synchrotron results.

1 Introduction

Structural reaction dynamics of proteins are fundamental to their biological function. Femtosecond time resolved protein X-ray crystallography would provide a direct structural probe of these events, providing compelling motivation for exploiting
5 novel XFEL sources for such studies.[1] Traditionally, time resolved pump-probe spectroscopic techniques provide access to the early time processes of activation in biological molecules. One particular example of photo-induced biological activation concerns the photoisomerisation of the biological p-coumaric acid chromophore of the Photoactive Yellow protein photoreceptor. Focusing on the sub-picosecond
10 excited state dynamics, photoisomerisation couples the initial absorption event to trigger protein structural changes, which cause a cascading series of events and formation of reaction intermediates ranging from femtosecond to second time scales.[2-20] The primary photoproduct, 'I0', which is formed after excited state decay with a typical time constant of ~0.7 ps and approximately 30% primary quantum
15 yield[2] is characterized by visible absorption at 510 nm, red-shifted from the ground state absorption at 446nm. [21, 22] Ultrafast infrared spectroscopy has shown the trans-cis isomerisation is accompanied by a disruption of the hydrogen bonding between the chromophore C=O group and the Cys69 backbone, and includes the formation of a short lived ground state intermediate with distorted trans-cis configuration.[17]
20 Further rearrangements of the chromophore and the surrounding protein environment occur in the electronic ground state configuration, and launch a photocycle which is thermally reversible on the ms-second time scale. The following intermediates I1, I2' and I2 occur on ns, ms and ms time-scale, followed by reformation of the ground state, and involve global protein structural changes that
25 persist for ms-s, and are considered to form the biological signaling state of the photoreceptor.[7]

Using synchrotron radiation, pump-probe time resolved crystallography experiments of crystals of the Photoactive Yellow Protein use the Laue X-ray diffraction method to capture the photoinduced structure factor amplitude
30 differences.[23-26] The PYP is one of relatively small number of light sensitive protein crystals that have been amenable to time resolved pump-probe studies using this technique, with other examples including heme proteins and photosynthetic reaction center.[27-30] Since their first demonstration, these experiments have continued to generate considerable interest and activity, particularly to exploit the technique to
35 probe in detail the various reaction mechanisms and assignments from extensive time resolved measurements.[24-26, 31]

With the emergence of novel X-FEL femtosecond pulsed hard X-ray sources, new capabilities for pump-probe crystallography now exist.[1] It has been demonstrated that XFELs can be used for protein X-ray crystallography, particularly using the
40 stream of micron-sized crystals injected into the XFEL beam using a microjet.[32-36] Considering the crystallographic data quality, compared to that obtained at synchrotron sources, it is noted that good quality measurements of the structure factor amplitudes can be obtained by employing a 'Monte-Carlo' method of data integration of many observations from a large number of micro-crystals. This
45 approach has now also allowed the observation of anomalous scattering using the serial femtosecond crystallography approach, demonstrating the possibility to use XFELs also for phasing of protein structures.[33, 34] . This result indicates that also time resolved pump-probe serial femtosecond crystallography may be possible,

which requires a very precise measurement of the small photo-induced structure factor amplitude differences, with precision that exceeds the noise levels of the collected and processed data. Here, we consider the alternative possibility of using a single, macro-crystal for pump-probe studies with XFEL radiation, particularly in view of the noise levels which are considerably higher than those for datasets collected at synchrotron sources using either the rotation method or the Laue method. A distinct advantage of the Serial Femtosecond Crystallography approach is that radiation damage should not contribute to the photoinduced difference measurements, whereas single crystal experiments collected using a stroboscopic approach do experience the effects of radiation damage.

Achieving sufficient signal-to-noise of the difference measurements depends on both the concentration of the photointermediate as well as the crystallographic data quality. Femtosecond photolysis is thus a critical parameter, which has been considered in detail on the basis of extensive power-, pulse duration- and wavelength dependence as well as the addition of second order dispersion using passive pulse shaping techniques.[2] In the case of PYP it was found that photochemical dynamics fundamentally limit the achievable time resolution to a few hundred femtoseconds, requiring stretching and addition of second order dispersion to suppress the stimulated emission path. Numerical modelling extracted the non-linear cross sections and projected photoproduct yields which were also confirmed using conventional model-free Z-scan measurements. Perhaps counter-intuitively,[37] optical penetration depth is not critical in this regime, in contrast to nanosecond excitation. Since the non-linear cross sections rapidly dominate the population transfer with high peak power for femtosecond excitation, little or no additional gain will be achieved for reduction of the photolysed depth from large single crystals to micro-crystals. From the ultrafast spectroscopy studies, a compromise could be formulated which results in an instrumental cross correlation of ~300 fs with approximately 10% photolysed yield in the interaction region.[2] In this contribution we consider the impact of the X-ray source parameters on the crystallographic data quality collected on single crystals, firstly to evaluate the ability to scale and merge ground state data sets with sufficient accuracy.

2 Experiment

X-ray crystallography experiments were conducted at the X-ray Pump-Probe station (XPP) at the Linac Coherent Light Source (LCLS) at Stanford, USA during experiments xpp23410 and xpp44112. For the xpp23410 experiments the X-ray slits were set to 80x40 μm or to 200x200 μm, the wavelength was 1.37757 Å, and a monochromator was used with a 1.6 eV band pass. The beam intensity was adjusted by inserting attenuators such that the strongest Bragg reflections did not saturate the detector and was focused to give a convergence of 0.007°, and a spot size of 80x40 μm. The distance to the detector was 68.5 mm, as refined from the data processing, using a MAR 165 detector. The repetition rate was 2 Hz and five exposures were used to fill a single image, with either continuous rotation or of a stationary crystal. For the xpp44112 experiments, the wavelength was 1.3099 Å and a monochromator was used with 1.6 eV bandpass. The pulse duration was ~ 70 fs the X-ray slits were set to 40x40 μm. The beam was attenuated to below detector saturation with 5 X-ray pulses, with the aim to approach typical synchrotron conditions with regard to incident flux and pump-probe cycles. The distance to the detector was 59.1 mm and

was calibrated from the powder pattern from SeO2. The detector was a mar 165 CCD. P63 crystals of PYP were mounted in capillaries at room temperature and were typically ~ 50x50x400 μm dimension. The beam was centered on the crystals and scanned across the needles to collect data from multiple positions. Data were
5 processed using MOSFLM[38] and reduced with Scala[39] and CCP4,[40] and also with HKL-2000.[41] Rigid body refinement was performed with REFMAC[42] using coordinates from 2PHY.[43]
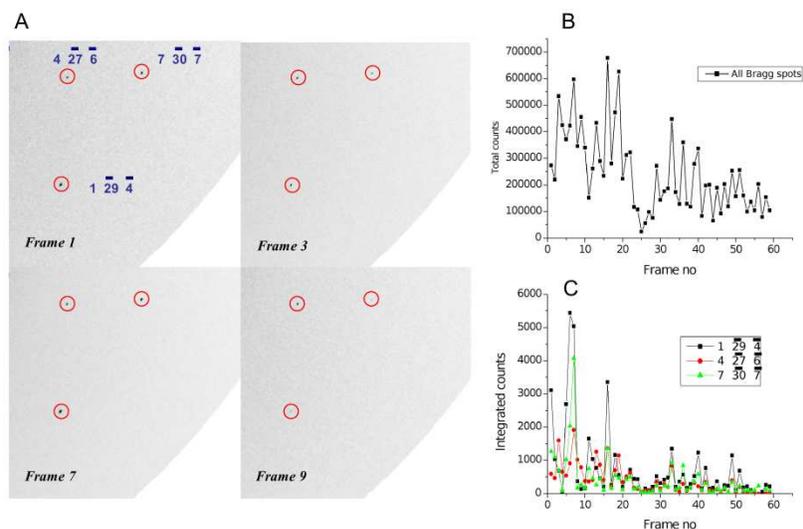

10 **Fig. 1** Bragg diffraction intensity fluctuations, and decay due to radiation damage, with a static crystal orientation. A). At the detector edge, reflections (1,-29,-4), (4,-27,-6) and (7,-30,-7) are shown for representative frame numbers 1,3,7, and 9 in the data collection series of 59 images. B) The total integrated counts for all Bragg spots for frames 1 through to 59. C) Integrated intensities for (1,-29,-4) (squares), (4,-27,-6) (circles) and (7,-30,-7) (triangles) on each frame.

15 ## 3 Results

3.1 – X-ray beam mode structure and jitter in this structure are the origins of Bragg reflection intensity noise with static crystal orientation

With a detector distance of 68.5 mm, a wavelength of 1.3776 Å and a 1.6 eV bandpass, $P6_3$ crystals of PYP shows diffraction spots to the edge of the detector
20 with up to 1.5 Å resolution. A test of the intrinsic noise of the Bragg diffraction intensity, as well as a characterization of the radiation damage under the attenuated conditions, included a data collection of a single crystal volume with static crystal orientation. Setting the X-ray slits to 200x200 μm, a series of 59 images, with 5 exposures each, followed the general decay and reproducibility of the diffraction
25 intensities. Figure 1 shows an example of three high resolution Bragg reflections and their integrated intensities on the 59 subsequent images (panels A and C). Figure 1B shows the total integrated intensities of all Bragg spots on each image.

The total diffraction intensities show large fluctuations which are primarily

caused by the intensity fluctuations of the X-ray source. The intrinsic bandwidth of the SASE generated beam is ~30 eV, with a center wavelength of 1.38 Å. Therefore, the inclusion of a monochromator with a 1.6 eV bandpass results in essentially 100% pulse intensity fluctuations, such that also the average of the sums of five pulses still
5 produces several-fold diffraction intensity fluctuations as seen in Figure 1B. Furthermore, there is a general decay caused by radiation damage of the protein crystal, which limit the number of exposures of each crystal volume. Importantly, it is seen in Figure 1C that there are also large fluctuations in the relative diffraction intensities of individual Bragg reflections, in spite of the stationary orientation and
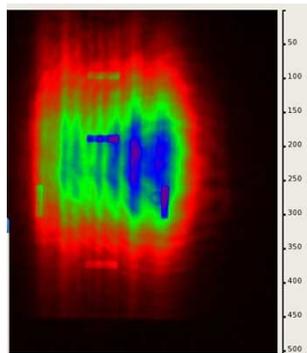10 energy selection by the monochromator. Whereas theses three reflections are

Fig. 2 Measured mode structure of the X-ray beam with open aperture showing ~ 250x250 μm spot size. The structure is caused by diffraction from an upstream mirror, which has smaller active length than the footprint of the beam. The structure shows additional shot-to-shot jitter.

15 presented as examples, this observation was also true for all other reflections measured. The possibility of crystal 'twitching' in response to X-ray exposure can very likely be discounted under these attenuated conditions. In contrast, un-attenuated conditions were seen to ablate protein crystals and generate acoustic shock sufficient to disorder adjacent unexposed crystal volume to the extent that
20 diffraction was lost. Under attenuated condition used here, the absorbed X-ray dose is on the same order of magnitude as that under conditions of pulsed Laue X-ray crystallography at synchrotron sources such as BioCARS 14ID.[44] Since both the 100 ps synchrotron source pulse duration and the 70 fs XFEL pulse duration are within the acoustic propagation time across the crystal, differences in pulse duration should
25 not account for twitching occurring under XFEL radiation exposure. Instead, the differences in relative intensities of reflections are likely due to stimulation of spatially separated mosaic blocks due to strong structure of the X-ray beam mode and jitter of this structure (Figure 2).

In comparison, a study of tetragonal Thaumatin crystals at the 19-BM synchrotron
30 station at the Advanced Photon Source established the repeatability of retrieving the relative Bragg diffraction intensity after correction for radiation damage decay and additionally a small storage ring current drop, to be on the order of 0.7-7% r.m.s.d. for low and high resolution diffraction bins, respectively.[45]

The feasibility of small scale mosaic block microstructure is taken from phase
35 contrast X-ray diffraction imaging of lysozyme crystals, which together with observation of anisotropic rocking curves indicate a micro-structure of crystal defects on length scales which would be expected to affect the measurements made

with a 200x200 μm area shown in Figure 1.[46] Reducing the X-ray slits to smaller size did not appreciably change the noise levels.

3.2 – Rocking curves of PYP crystals measured with non-diverging XFEL beam.

The reported convergence of the LCLS X-ray source is 0.007°, which is considerably smaller than the mosaic spread of PYP crystals. A rocking curve was measured over a 0.33° range, with incrementing the goniometer in 0.01° steps, and each image read after five pulses, as used in Figure 1. Each image contained between 150 and 200 Bragg spots with I/σI of 3.0 or above. Figure 3 shows representative rocking curves collected for spots in low and medium resolution bins. The full width half max ($\theta_{1/2}$) of a Bragg diffraction of a perfect crystal is given by[46]

$$\Theta_{1/2} = 2d\xi^{-1} \tag{1}$$

where d is the interplanar spacing and ξ is the extinction length which in the symmetric Laue case is given by

$$\xi = \frac{\pi \cos \theta_B}{Cr_e\lambda(F_g/V)} \tag{2}$$

where r_e is the classical electron radius ($r_e = 2.82$ Å 10^{-5} Å), F_g is the structure factor, V is the volume of the unit cell, C is the polarisation factor, λ is the wavelength and θ_B is the Bragg angle. With the addition of crystal mosaicity the phenomenological appearance of the angular intensity spread was however better represented by Gaussian functions rather than cosine functions,[47] and the fitted f.w.h.m. values are shown in Figure 3 for four representative reflections at various resolution.
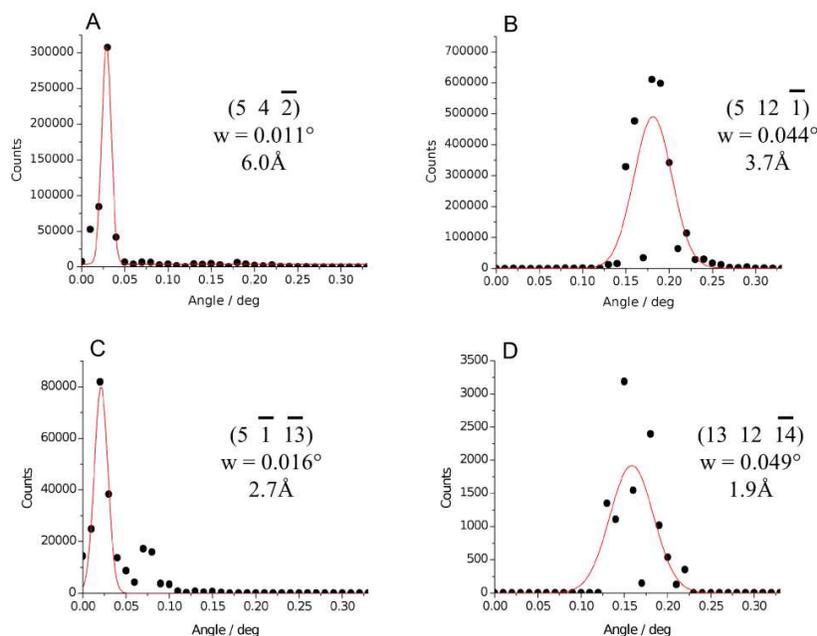
Fig. 3 Representative measured rocking curves for A) (5,4,-2) (f.w.h.m. 0.011°), B) (5,12,-1) (f.w.h.m. 0.044°), C), (5,-1,-13) (f.w.h.m. 0.016°) and D) (13, 12, -14) (f.w.h.m. 0.049°).

The histogram of a set of selected reflections shows a considerable spread in the f.w.h.m. values between 0.002° and 0.075 ° (Figure 4). These values should be taken only as estimates given the fluctuations of the source intensity as well as the additional structure factor amplitude noise demonstrated in Figure 1. However, the results resembles for instance rocking curves determined for lysozyme crystals in their anisotropic distribution but also absolute values as well as line shape defects and splitting (Fig. 3C).[47] The results indicate that the combination of the low mosaic spread of PYP crystals and the very low convergence angle result in very sparse observation when using the rotation method. Below, we show that increased mosaic spread would significantly improve the resulting signal-to-noise of data collected under such conditions.
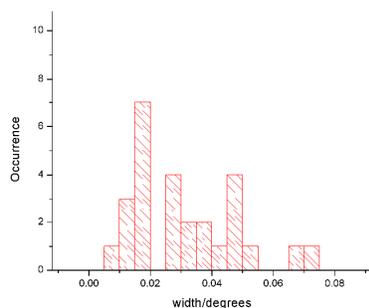

Fig. 4 Histogram of 27 individually fitted rocking curves of a single PYP crystal rotated through a 0.33° range, for reflections selected at various resolution and I/σI exceeding 10 at the peak.
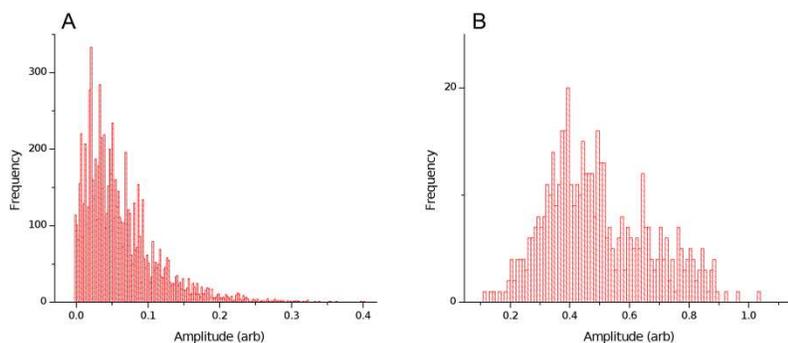

Fig. 5 Histogram of 8,475 pulse intensities from SASE mode at 9.47 keV with a monochromator. The I/σI from this series is 1.21 B) Histogram of 500 pulses from self-seeded mode at 8.4 keV using the entire length of the undulator, as recently demonstrated. The I/σI from this series is 2.84.

3.3 – Source noise and modelled diffraction intensity envelope projections with SASE and self-seeded mode

In the following, we consider the source intensity noise of SASE mode as well as recently demonstrated self-seeded mode, which have energy selection by including the monochromator and seeding, respectively. A photodiode detector recorded the

intensity of each shot during data collection using multiple single crystals of PYP, using the SASE mode with energy selection at 1.3099 Å and 1.6 eV bandpass. Within a single period of data collection, a series of 8,475 pulses were measured (Figure 5A). Recently, LCLS was operated in self-seeded mode, which leads to
5 stabilization of the central wavelength, which a typical single shot f.w.h.m. of 0.5 eV and time averaged 1 eV bandwidth, at 8-9 keV.[48] The reported parameters included a 50% r.m.s. stability fluctuation, in part due to lack of FEL saturation in the seeded half of the undulator. Experimental results presented in Amann et al (2012) Nature Photonics come from the very first self-seeding runs [48] (January
10 2012). At that time, the last four LCLS undulators were not used for the self-seeding. Recently, it was reported that all the undulators can be used for self-seeding operation having an impact both on the intensity and the stability of the self-seeded beam (personal communication, Alberto Lutman and Paul Emma, LCLS, SLAC). Figure 5B shows a recent set of 500 pulses at 8.4 keV, generously provided by
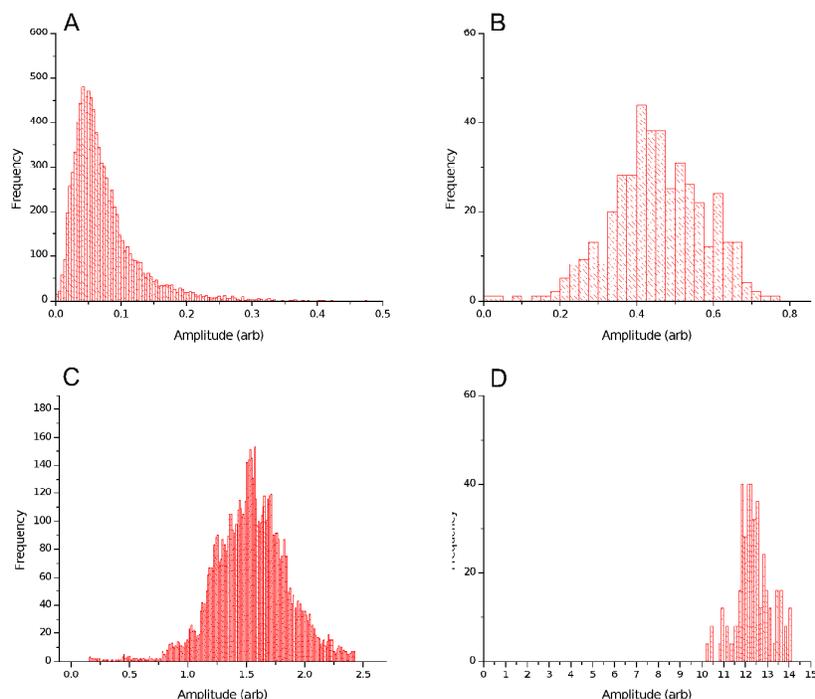
15

Fig. 6 Modelled distributions of the total integrated Bragg reflection intensity for SASE mode (A,C) and self-seeded mode (B,D), using data from Figure 5A,B. (A) Expected distribution for a mosaic spread of f.w.h.m. of 0.04° and a step of 0.02° using SASE mode. The I/σI for this is expected to be 1.47. (B) Expected distribution for a mosaic spread of f.w.h.m. of 0.04° and a step of 0.02° using
20 self-seeded mode. The I/σI for this is expected to be 3.81. (C) Expected distribution for a mosaic spread of f.w.h.m. of 0.4° and a step of 0.02° using SASE mode. The I/σI for this is expected to be 5.05. (D) Expected distribution for a mosaic spread of f.w.h.m. of 0.4° and a step of 0.02° using self-seeded mode. The I/σI for this is expected to be 16.4.

Alberto Lutman and Paul Emma, which may be used to model the resulting
25 crystallographic noise in comparison with the SASE mode.

Together with the rocking curve information it is possible to estimate the

accuracy of the estimate for the integrated structure factor amplitudes, on the basis of these intensity distributions alone, neglecting the additional noise resulting from mode structure and jitter. The rotation method was experimentally implemented by *continuous* rotation over a 0.1° range for each integrated image, with 5 X-ray probes, thus probing at 0.02° intervals. In this 'quasi cw' manner, the XFEL pulse train probes the envelope of the Bragg reflection up to 6 times, assuming an average value for the f.w.h.m. of 0.04° (Figure 4). Figure 6A shows the simulated histogram of the total resulting integrated intensity using the data shown in Figure 5A, which correspond to the actual experimental conditions of data collection (see below section 3.4). This distribution thus corresponds to the integrated 2D projection of a single observation of a Bragg spot, assuming only X-ray intensity noise contributions. Figure 6B shows the modelled distribution expected under conditions of using self-seeded mode, using data from Figure 5B. Figure 6C simulates the resulting distribution if the mosaic spread would increase 10-fold to a value of f.w.h.m. of 0.4°, but maintaining the 0.1° rotation, resulting in an average number of 76 probes for each reflection. Finally, figure 6D simulates a mosaic spread of f.w.h.m. of 0.4° and a 0.1° rotation per image.

Table 1 Data reduction statistics

	Multi-crystal dataset	Single-crystal dataset
# of frames / # of crystals	96/6	23/1
Space group	$P6_3$	$P6_3$
Unit cell (a,b,c(Å);α,β,γ(°))	66.934, 66.934, 40.986	66.867, 66.867, 40.985
	90.000 90.000, 120.000	90.000 90.000, 120.000
Resolution range (Å)	57.97-1.60	100-1.50
Reflections (Total/Unique)	137,459/13,877	20,145/16,947
Rejected outlier (#/%)	9,986/7.3	1,286/6.4
Completeness[a]	84.3(63.6)	51.4(21.6)
I/σI	3.1(1.4)	7.1(1.5)
Average Multiplicity	4.8	1.2
R_{merge}[b]	0.400 (0.534)	0.239 (0.470)
R_{meas}[c]	0.445 (0.614)	N.D.
$R_{p.i.m.}$[d]	0.190 (0.289)	N.D.
PCV[e]	0.635 (0.822)	N.D.
R_{crys}/R_{free} (%)	0.26 (0.23)	N.D.

[a] Numbers appearing in parenthesis are for the high resolution shell.

[b] $$R_{merge}(I) = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$$

[c] $$R_{meas}(I) = \sum_{hkl} \sum_i^n \left(\frac{n}{n-1} \right)^{1/2} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$$

[d] $$R_{p.i.m.}(I) = \sum_{hkl} \sum_i^n \left(\frac{1}{n-1} \right)^{1/2} |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$$

[e] $$PCV(I) = \sum_{hkl} \left(\frac{1}{n-1} \sum_i^n (I_i(hkl) - \langle I(hkl) \rangle)^2 \right)^{1/2} / \sum_{hkl} \langle I(hkl) \rangle$$

3.4 – Applying the 'quasi-cw' rotation method to PYP crystals using the SASE mode

Data were collected using the rotation method at 1.3099 Å and 1.6 eV bandpass, using a rotation of 0.1° and 5 probes at 0.02° steps for each image, corresponding to conditions shown in Figure 5A and Figure 6A. The beam was 80x40 μm and was

translated by 100 μm to collect data on multiple crystal volumes on a single crystal. Six crystals in total were used, with each changeover starting at the equivalent indexed orientation to ensure continuous coverage of reciprocal space between crystals. Data were processed using both MOSFLM[38] and HKL-2000,[41] exploring
5 conditions for outlier rejections. Table 1 presents statistics for a multi-crystal averaged dataset and a single crystal dataset.

The multi-crystal averaged dataset was processed in order to maximize the completeness, with an average multiplicity of 4.8. The relatively low completeness is caused by overlap of missing reflections from similar physical orientation of all
10 mounted crystals, with the crystallographic c-axis corresponding to the long axis. Processing of data from a single crystal with increased rejection criteria appears to improve the R_{merge} and I/σ_I value of the data at the expense of 4.9% rejection, but having only low completeness. The R_{merge} and I/σI values are however meaningless in this regime with an average multiplicity of 1.2, thus lacking most of the intensity
15 noise except for a small portion of reflections which had a multiplicity of 2 and were distributed throughout all resolution bins. For a Gaussian distributed error, it is expected that the PCV(I) value is a factor of sqrt(π/2)~ 1.25-fold higher than the R_{meas} value.[49] This ratio is 1.43 and 1.34 for the average and highest resolution shell,
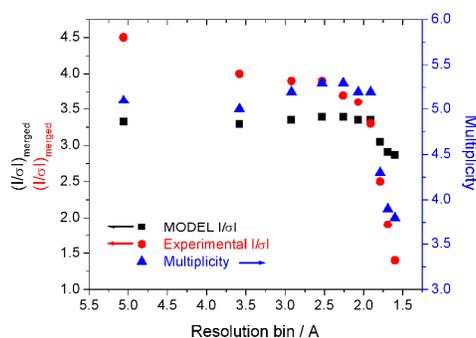

20 **Fig. 7** Modelled (I/σI)merged values (black squares) and experimental (I/σI)merged values (dots) as determined by Scala (Evans 2006), for separate resolution bins. The Multiplicity (triangles), together with a fit of (I/σI)unmerged taken from Figure 6A are used with equation (4) to model the plotted values (squares).

indicating that with the number of observations available some deviation from
25 Gaussian distribution is present.

A simple relationship may be used to consider the dominating experimental error arising from the source fluctuations in terms of an expected R-value[49]

$$R_{merge} = \left(\frac{2}{\pi}\right)^{1/2} \frac{\langle\sigma_I\rangle}{\langle I_o\rangle} \simeq \frac{0.7979}{\langle I_o\rangle / \langle\sigma_I\rangle} \tag{3}$$

Furthermore, by approximating the simulated distributions seen in Figure 6 as purely
30 statistical variation, the I/σI value may be estimated

$$\left(\frac{\langle I_o\rangle}{\langle\sigma_I\rangle}\right)_{merged} = n^{1/2}\left(\frac{\langle I_o\rangle}{\langle\sigma_I\rangle}\right)_{unmerged} \tag{4}$$

where n is the multiplicity

Following this model, based on the modelled I/σI value of 1.47 for a single observation under precisely the conditions of data collection (Fig 5A, 6A, Table 1), multiplication with the square root of the average multiplicity yields a value of 3.22
5 for the expected average value of $(I/\sigma I)_{merged}$ for the multi-crystal dataset (Table 1). The close correspondence suggest that the source intensity fluctuations dominate the experimental error of this dataset, and that the used (average) value of 0.04° for the f.w.h.m. of the mosaic spread in order to simulate the distribution in Figure 6A which provides $(I/\sigma I)_{unmerged}$ as a single observation appears to be appropriate. An
10 analysis of this model for each resolution bin shows deviation between predicted and experimental $(I/\sigma I)_{merged}$ values in opposite directions at low and high resolution respectively (Figure 7).

This deviation could arise from the error-normalisation procedure that is carried out by Scala, which adjusts the errors according to [50]

$$\sigma'(I_{hl}) = Sdfac \ [\sigma^2(I_{hl}) + SdB \cdot \langle I_h \rangle + \left(Sdadd \cdot \left\langle I'_h \right\rangle \right)^2]^{1/2} \tag{5}$$

15

Sdfac, SdB and Sdadd are fitted parameters used in the scaling of the errors for all intensity bins, in order to normalize the deviations δ_{hl}.

$$\delta_{hl} = \left(\frac{n-1}{n} \right)^{1/2} (I_{hl} - \langle I_h \rangle) / \sigma'(I_{hl}) \tag{6}$$

Sdadd is adjusted to model errors that are proportional to the intensity, and thus
20 includes source fluctuations. Scala uses a default value of Sdadd = 0.02, whereas the multi-crystal dataset (Table 1) resulted in a values of 0.091 and 0.084 for full and partials reflections, respectively. The normalized deviation resulted in values of ~0.7 and ~0.5 for the two highest resolution bins, which also correspond with the lowest reported $(I/\sigma I)_{merged}$ values from Scala results (Fig 7; Red dots), indicating that the
25 standard error is overestimated for the weakest reflections.

The crystallographic R-factor at 26%, and R-free of 29%, indicate that the scaling and averaging has led to generally reasonable amplitudes. Figure 8 shows resulting 2Fo-Fc electron density. It is noted however that at this level of accuracy the shape of the electron density is almost entirely determined by the synchrotron derived
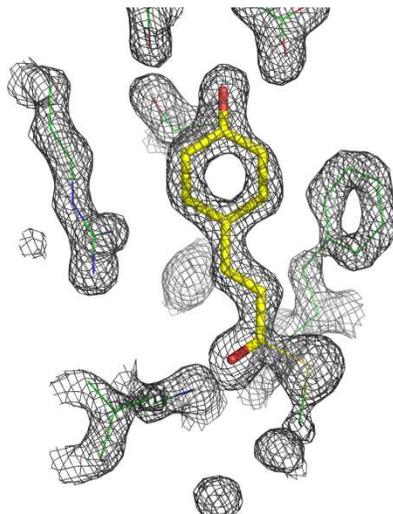30 phase angles used in the Fourier synthesis.

Fig. 8 2Fo-Fc electron density of the p-coumaric acid chromophore and direct environment contoured at 2σ level using the multi-crystal dataset (Table 1, center column).

4 Discussion

This work shows that crystallographic data quality collected under 'quasi-cw' and monochromatic conditions using the rotation method, in terms of its signal to noise parameter $(I/\sigma I)_{merged}$, may be understood from the dominating pulse intensity fluctuations. From data with a multiplicity of 4.8 there is correspondence with the square root law (eq. 4) which is sufficiently to predict the average $(I/\sigma I)_{merged}$ vale, but showing some deviations in opposite directions for high and low resolution bins as seen in Figure 7. We have identified an additional source of Bragg reflection intensity noise which is likely to be caused by the X-ray mode structure (Figure 2) and jitter of this mode structure. Nevertheless, with up to five 'probes' of the XFEL pulse train, assuming an average mosaic spread and using a Gaussian distribution model of the source noise fluctuations, a reasonable estimate of the signal to noise for a single reflection can be made, when fully rotated through its rocking curve.

The analysis that is presented here is in essence a 'Monte Carlo' application as no knowledge or measurement of the source is used in the data processing. In principle, post-refinement of partiality from a fine-slicing method might take advantage of the pulse intensity recorded for each image. In this work that would still require the refinement of a single integral with five pulses and rotation through a known portion of the rocking curve. This could still result in increased signal to noise of the merged and scaled observations. However, the presence of the noise caused by the mode and mode jitter strongly suggests that this will not improve the data in the case of rocking curves with small f.w.h.m. as for the PYP crystals. Such an approach may be reconsidered if beamline optics could be improved to remove the diffraction and mode structure.

Taking these results together with recent progress at LCLS which shows considerably improved stability, it may be predicted that in particular for single crystals with larger mosaic spread, signal to noise may be attained that could be comparable with that collected at synchrotron sources. For example, assuming a

mosaic spread of f.w.h.m.=0.4º, a $(I/\sigma I)$ value of 16.4 was predicted for the self-seeded mode at LCLS (Figure 6D). Assuming that equation 3 remains valid in this regime, the $(I/\sigma I)_{merged}$ would become 32.8 with a multiplicity of four. Depending on the space group symmetry, this could be achieved with collecting ~ 100 images,
5 which is a similar amount as for the dataset presented here (Table 1). The $(I/\sigma I)_{merged}$ value is presumably the most appropriate parameter in order to make a comparison with published synchrotron derived datasets. For example, time resolved pump-probe Laue X-ray crystallography data sets of the PYP which are sufficiently sensitive to detect the small photoinduced differences reproducibly, typically show
10 an average value of $(I/\sigma I)_{merged} \sim 35$ for data with a characteristic resolution of 1.6 Å.[23, 24, 31, 51, 52]

Interestingly, these results also suggest that XFELs are a particularly well placed complementary sources to synchrotron stations, also for picosecond and longer time studies. This argument considers mainly the mosaic spread of protein crystals. Fast
15 time-resolved synchrotron measurements rely on the Laue method, and a 'quasi-cw' scheme as used here is usually not feasible because of the lower intrinsic pulse power of spontaneous radiation. Even with a very bright source such as 14-ID beamline at BioCARS which uses two in vacuum undulators,[44] a monochromatic pulsed application would require too many pump-probe cycles in order to collect
20 complete datasets. Instead, Laue crystallography is used. The Laue technique however requires very high quality crystals particularly those that have a very small mosaic spread, typically below f.w.h.m = 0.1º. Since the multiple frequencies stimulate mosaic blocks at different Bragg angles, Laue diffraction of crystals with higher mosaic spread show elongation of the Bragg spots which presents difficulties
25 for the data integration due to close overlap and additionally the energy gradient that is present along the stretched reflection. This work suggests that for crystals which are not well suitable for Laue crystallography, pump-probe time resolved monochromatic experiments at XFELs are especially promising as shown in this contribution.

30 5 Acknowledgements

40 References

[*] *Imperial College London, Division of Molecular Biosciences, South Kensington Campus, London SW7 2AZ, United Kingdom. Tel:+44(0)207 594 5071; E-mail: j.vanthor@imperial.ac.uk*
[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/
45 [‡] Footnotes should appear here. These might include comments relevant to but not central to the matter under discussion, limited experimental and spectral data, and crystallographic data.

Faraday Discussions Accepted Manuscript

1.   J. C. J. M. Glownia, J. Andreasson, A. Belkacem, N. Berrah, C. I. Blaga, C. Bostedt, J. Bozek, L. F. DiMauro, L. Fang, J. Frisch, O. Gessner, M. Gühr, J. Hajdu, M. P. Hertlein, M. Hoener, G. Huang, O. Kornilov, J. P. Marangos, A. M. March, B. K. McFarland, H. Merdji, V. S. Petrovic, C. Raman, D. Ray, D. A. Reis, M. Trigo, J. L. White, W. White, R. Wilcox, L. Young, R. N. Coffee, and P. H. Bucksbaum, *Opt. Express*, 2010, **18**, 17620-17630.

2.   C. N. Lincoln, A. E. Fitzpatrick and J. J. van Thor, *Phys Chem Chem Phys*, 2012, **14**, 15752-15764.

3.   Y. Imamoto, M. Kataoka, F. Tokunaga, T. Asahi and H. Masuhara, *Biochemistry*, 2001, **40**, 6047-6052.

4.   Y. Imamoto, K. Mihara, F. Tokunaga and M. Kataoka, *Biochemistry*, 2001, **40**, 14336-14343.

5.   Y. Imamoto, H. Kamikubo, M. Harigai, N. Shimizu and M. Kataoka, *Biochemistry*, 2002, **41**, 13595-13601.

6.   S. Devanathan, A. Pacheco, L. Ujj, M. Cusanovich, G. Tollin, S. Lin and N. Woodbury, *Biophys J*, 1999, **77**, 1017-1023.

7.   K. J. Hellingwerf, *Antonie Van Leeuwenhoek*, 2002, **81**, 51-59.

8.   J. Hendriks, T. Gensch, L. Hviid, M. A. van Der Horst, K. J. Hellingwerf and J. J. van Thor, *Biophys J*, 2002, **82**, 1632-1643.

9.   W. D. Hoff, A. Xie, I. H. Van Stokkum, X. J. Tang, J. Gural, A. R. Kroon and K. J. Hellingwerf, *Biochemistry*, 1999, **38**, 1009-1017.

10.   T. E. Meyer, G. Tollin, T. P. Causgrove, P. Cheng and R. E. Blankenship, *Biophys J*, 1991, **59**, 988-991.

11.   M. Unno, M. Kumauchi, N. Hamada, F. Tokunaga and S. Yamauchi, *J Biol Chem*, 2004, **279**, 23855-23858.

12.   M. A. van der Horst, W. Laan, S. Yeremenko, A. Wende, P. Palm, D. Oesterhelt and K. J. Hellingwerf, *Photochem Photobiol Sci*, 2005, **4**, 688-693.

13.   B. Borucki, C. P. Joshi, H. Otto, M. A. Cusanovich and M. P. Heyn, *Biophys J*, 2006, **91**, 2991-3001.

14.   J. T. Kennis and M. L. Groot, *Curr Opin Struct Biol*, 2007, **17**, 623-630.

15.   D. S. Larsen and R. van Grondelle, *Chemphyschem*, 2005, **6**, 828-837.

16.   I. H. van Stokkum, B. Gobets, T. Gensch, F. Mourik, K. J. Hellingwerf, R. Grondelle and J. T. Kennis, *Photochem Photobiol*, 2006, **82**, 380-388.

17.   L. J. van Wilderen, M. A. van der Horst, I. H. van Stokkum, K. J. Hellingwerf, R. van Grondelle and M. L. Groot, *Proc Natl Acad Sci U S A*, 2006, **103**, 15050-15055.

18.   M. L. Groot, L. J. van Wilderen, D. S. Larsen, M. A. van der Horst, I. H. van Stokkum, K. J. Hellingwerf and R. van Grondelle, *Biochemistry*, 2003, **42**, 10054-10059.

19.   M. L. Groot, L. J. van Wilderen and M. Di Donato, *Photochem Photobiol Sci*, 2007, **6**, 501-507.

20.   C. P. Joshi, B. Borucki, H. Otto, T. E. Meyer, M. A. Cusanovich and M. P. Heyn, *Biochemistry*, 2006, **45**, 7057-7068.

21.   D. S. Larsen, M. Vengris, I. H. van Stokkum, M. A. van der Horst, F. L. de Weerd, K. J. Hellingwerf and R. van Grondelle, *Biophys J*, 2004, **86**, 2538-2550.

22.   A. D. Stahl, M. Hospes, K. Singhal, I. van Stokkum, R. van Grondelle, M. L. Groot and K. J. Hellingwerf, *Biophys J*, 2011, **101**, 1184-1192.

23.   B. Perman, V. Srajer, Z. Ren, T. Teng, C. Pradervand, T. Ursby, D. Bourgeois, F. Schotte, M. Wulff, R. Kort, K. Hellingwerf and K. Moffat, *Science*, 1998, **279**, 1946-1950.

24.   H. Ihee, S. Rajagopal, V. Srajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff and K. Moffat, *Proc Natl Acad Sci U S A*, 2005, **102**, 7145-7150.

25.   Y. O. Jung, J. H. Lee, J. Kim, M. Schmidt, K. Moffat, V. Srajer and H. Ihee, *Nat Chem*, 2013, **5**, 212-220.

26.   M. Schmidt, V. Srajer, R. Henning, H. Ihee, N. Purwar, J. Tenboer and S. Tripathi, *Acta Cryst. D.*, 2013, **69**, 2534–2542.

27.   V. Srajer, T. Teng, T. Ursby, C. Pradervand, Z. Ren, S. Adachi, W. Schildkamp, D. Bourgeois, M. Wulff and K. Moffat, *Science*, 1996, **274**, 1726-1729.

28.   F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, Jr., M. Wulff and P. A. Anfinrud, *Science*, 2003, **300**, 1944-1947.

29.   J. E. Knapp, R. Pahl, V. Srajer and W. E. Royer, Jr., *Proc Natl Acad Sci U S A*, 2006, **103**, 7649-7654.

30. A. B. Wohri, G. Katona, L. C. Johansson, E. Fritz, E. Malmerberg, M. Andersson, J. Vincent, M. Eklund, M. Cammarata, M. Wulff, J. Davidsson, G. Groenhof and R. Neutze, *Science*, 2010, **328**, 630-633.

31. F. Schotte, H. S. Cho, V. R. Kaila, H. Kamikubo, N. Dashdorj, E. R. Henry, T. J. Graber, R. Henning, M. Wulff, G. Hummer, M. Kataoka and P. A. Anfinrud, *Proc Natl Acad Sci U S A*, 2012, **109**, 19256-19261.

32. 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. 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. N. 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. Homke, C. Reich, D. Pietschner, L. Struder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. U. Kuhnel, 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. Jonsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. D. Schroter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Y. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. H. Spence, *Nature*, 2011, **470**, 73-U81.

33. T. R. Barends, L. Foucar, S. Botha, R. B. Doak, R. L. Shoeman, K. Nass, J. E. Koglin, G. J. Williams, S. Boutet, M. Messerschmidt and I. Schlichting, *Nature*, 2013.

34. T. R. Barends, L. Foucar, R. L. Shoeman, S. Bari, S. W. Epp, R. Hartmann, G. Hauser, M. Huth, C. Kieser, L. Lomb, K. Motomura, K. Nagaya, C. Schmidt, R. Strecker, D. Anielski, R. Boll, B. Erk, H. Fukuzawa, E. Hartmann, T. Hatsui, P. Holl, Y. Inubushi, T. Ishikawa, S. Kassemeyer, C. Kaiser, F. Koeck, N. Kunishima, M. Kurka, D. Rolles, B. Rudek, A. Rudenko, T. Sato, C. D. Schroeter, H. Soltau, L. Strueder, T. Tanaka, T. Togashi, K. Tono, J. Ullrich, S. Yase, S. I. Wada, M. Yao, M. Yabashi, K. Ueda and I. Schlichting, *Acta Crystallogr D Biol Crystallogr*, 2013, **69**, 838-842.

35. L. C. Johansson, D. Arnlund, G. Katona, T. A. White, A. Barty, D. P. Deponte, R. L. Shoeman, C. Wickstrand, A. Sharma, G. J. Williams, A. Aquila, M. J. Bogan, C. Caleman, J. Davidsson, R. B. Doak, M. Frank, R. Fromme, L. Galli, I. Grotjohann, M. S. Hunter, S. Kassemeyer, R. A. Kirian, C. Kupitz, M. Liang, L. Lomb, E. Malmerberg, A. V. Martin, M. Messerschmidt, K. Nass, L. Redecke, M. M. Seibert, J. Sjohamn, J. Steinbrener, F. Stellato, D. Wang, W. Y. Wahlgren, U. Weierstall, S. Westenhoff, N. A. Zatsepin, S. Boutet, J. C. Spence, I. Schlichting, H. N. Chapman, P. Fromme and R. Neutze, *Nat Commun*, 2013, **4**, 2911.

36. S. Boutet, L. Lomb, G. J. Williams, T. R. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, M. Messerschmidt, A. Barty, T. A. White, S. Kassemeyer, 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. Spence, H. N. Chapman and I. Schlichting, *Science*, 2013, **337**, 362-364.

37. R. Neutze and K. Moffat, *Curr Opin Struct Biol*, 2012, **22**, 651-659.

38. A. G. Leslie, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 48-57.

39. W. Kabsch, *J. Appl. Crystallogr.*, 1988, **21**, 916-924.

40. 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. 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, *Acta Crystallogr D Biol Crystallogr*, 2011, **67**, 235-242.

41. Z. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307-326.

42. G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr D Biol Crystallogr*, 2011, **67**, 355-367.

43. G. E. Borgstahl, D. R. Williams and E. D. Getzoff, *Biochemistry*, 1995, **34**, 6278-6287.

44. T. Graber, S. Anderson, H. Brewer, Y. S. Chen, H. S. Cho, N. Dashdorj, R. W. Henning, I. Kosheleva, G. Macha, M. Meron, R. Pahl, Z. Ren, S. Ruan, F. Schotte, V. Srajer, P. J.

Viccaro, F. Westferro, P. Anfinrud and K. Moffat, *J Synchrotron Radiat*, 2011, **18**, 658-670.

45.   D. Liebschner, M. Dauter, G. Rosenbaum and Z. Dauter, *Acta Crystallogr D Biol Crystallogr*, 2012, **68**, 1430-1436.

5 46.   Z. W. Hu, Y. S. Chu, B. Lai, B. R. Thomas and A. A. Chernov, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 621-629.

47.   D. Lubbert, A. Meents and E. Weckert, *Acta Crystallogr D Biol Crystallogr*, 2004, **60**, 987-998.

48.   J. Amann, W. Berg, V. Blank, F. J. Decker, Y. Ding, P. Emma, Y. Feng, J. Frisch, D.
10   Fritz, J. Hastings, Z. Huang, J. Krzywinski, R. Lindberg, H. Loos, A. Lutman, H. D. Nuhn, D. Ratner, J. Rzepiela, D. Shu, Y. Shvyd'ko, S. Spampinati, S. Stoupin, S. Terentyev, E. Trakhtenberg, D. Walz, J. Welch, J. Wu, A. Zholents and D. Zhu, *Nat. Photonics*, 2012, **6**, 693-698.

49.   K. Diederichs and P. A. Karplus, *Nat Struct Biol*, 1997, **4**, 269-275.

15 50.   P. Evans, *Acta Crystallogr D Biol Crystallogr*, 2006, **62**, 72-82.

51.   K. Moffat, *Nat Struct Biol*, 1998, **5 Suppl**, 641-643.

52.   S. Rajagopal, K. S. Kostov and K. Moffat, *J Struct Biol*, 2004, **147**, 211-222.
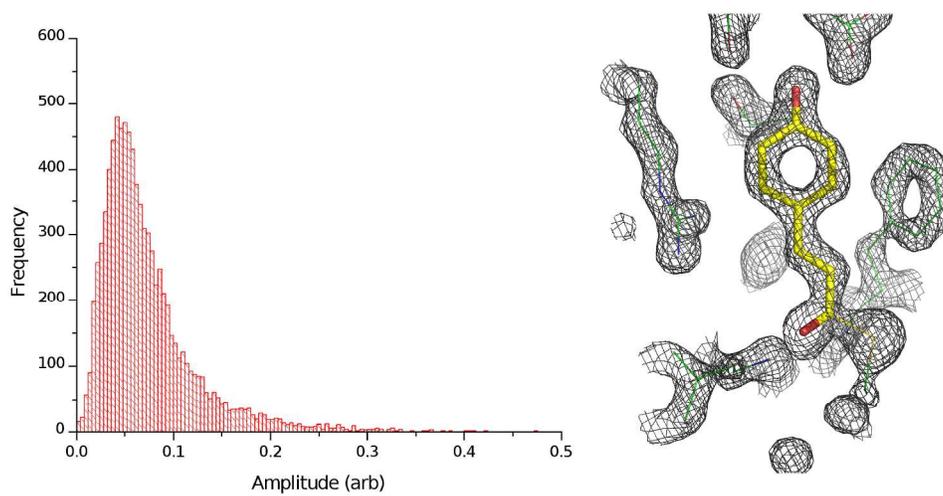
A model for the signal-to-noise value for applying the rotation method to single crystals at LCLS