Methods development for diffraction and spectroscopy studies of metalloenzymes at X-ray free-electron lasers

Jan Kern\textsuperscript{1,2}, Johan Hattne\textsuperscript{1}, Rosalie Tran\textsuperscript{1}, Roberto Alonso-Mori\textsuperscript{2}, Hartawan Laksmono\textsuperscript{3}, Sheraz Gul\textsuperscript{1}, Raymond G. Sierra\textsuperscript{3}, Jens Rehanek\textsuperscript{4}, Alexei Erko\textsuperscript{4}, Rolf Mitzner\textsuperscript{5}, Phillip Wernet\textsuperscript{5}, Uwe Bergmann\textsuperscript{2}, Nicholas K. Sauter\textsuperscript{1}, Vittal Yachandra\textsuperscript{1} and Junko Yano\textsuperscript{1}

1Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
2CLS, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
3Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
4Institute for Nanometre Optics and Technology, and 5Institute for Methods and Instrumentation for Synchrotron Radiation Research, Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Berlin 12489, Germany

X-ray free-electron lasers (XFELs) open up new possibilities for X-ray crystallographic and spectroscopic studies of radiation-sensitive biological samples under close to physiological conditions. To facilitate these new X-ray sources, tailored experimental methods and data-processing protocols have to be developed. The highly radiation-sensitive photosystem II (PSII) protein complex is a prime target for XFEL experiments aiming to study the mechanism of light-induced water oxidation taking place at a Mn cluster in this complex. We developed a set of tools for the study of PSII at XFELs, including a new liquid jet based on electrofocusing, an energy dispersive von Hamos X-ray emission spectrometer for the hard X-ray range and a high-throughput soft X-ray spectrometer based on a reflection zone plate. While our immediate focus is on PSII, the methods we describe here are applicable to a wide range of metalloenzymes. These experimental developments were complemented by a new software suite, \texttt{cctbx.xfel}. This software suite allows for near-real-time monitoring of the experimental parameters and detector signals and the detailed analysis of the diffraction and spectroscopy data collected by us at the Linac Coherent Light Source, taking into account the specific characteristics of data measured at an XFEL.

1. Introduction

Since the start of operation at Linac Coherent Light Source (LCLS) [1], the application of X-ray free-electron lasers (XFELs) to important biological problems has evolved rapidly. Shortening the exposure to below the time required for a sample to be damaged by the deposition of radiation energy by the X-ray pulse was proposed years before, and has spurred development in these directions [2]. In addition to the potential for sample damage, concerns regarding the feasibility of XFEL studies arose from the following possible X-ray-induced changes: (i) sequential photon absorption within the timescale of the XFEL pulses and (ii) Coulomb explosion resulting from the accumulation of excessive charge in the molecule. The concept of ‘probe before destroy’ entails using a femtosecond X-ray pulse of very high intensity and measuring the response of the system before the manifestation of radiation-induced changes. Over the past three years, several pioneering ‘proof-of-principle’ experiments have been conducted, applying this concept to biological samples, both at LCLS [3–6] and more recently at SACLA [7], the XFEL at SPring-8 in Japan. To allow the rapid evolution of this emerging field, it was necessary to develop new techniques for sample preparation, sample delivery, signal recording and data processing. In this contribution, we review...
the current status of methods developed in these various areas and describe their application for our ongoing experiments to further understand the reaction mechanism of photosystem II (PSII). We start with a short introduction to the structure and function of PSII (§2), followed by a description of sample delivery techniques (§3), summarizing the methods used so far with emphasis on the electrospinning jet developed for our experiments. In addition, we discuss the development of specific tools for performing X-ray spectroscopic measurements on metals contained in dilute biological samples (§4). We finish with a discussion of the software packages developed by us to process both spectroscopic and diffraction data obtained at LCLS (§5), using thermolysin to illustrate the principles for diffraction processing (see [8] for review on PSII).

2. Structure and function of PSII: background and questions

Photosystem II, which catalyses the light-driven oxidation of water into oxygen [9,10], is a membrane protein complex found in green plants, algae and cyanobacteria. The catalytic centre for water oxidation is the oxygen evolving complex (OEC), consisting of an Mn$_4$Ca$_4$O$_5$ cluster and its ligand environment [11]. At the OEC, two molecules of water are oxidized to form dioxygen under the release of four protons and four electrons. To catalyse this reaction, the OEC successively accumulates oxidation equivalents, transitioning through different states in the so-called S-state cycle [12]. The precise mechanism of this process is still not well understood. Detailed mechanistic studies are hampered by the intrinsic radiation sensitivity of the metal cluster to X-rays [13–15], a problem also encountered in other redox-active metalloenzymes [16–19]. In PSII, the higher oxidation states of Mn (Mn$^{III}$ and Mn$^{IV}$) that are present in the intact OEC are rapidly reduced to Mn$^{II}$ upon radiation damage, resulting in structural modifications (e.g. changes in Mn–O–Mn bridging structure, changes in metal–metal and metal–ligand bond lengths). Here, the advantages of XFELs can be exploited, especially the possibility of collecting data at room temperature under near-physiological conditions.

3. Sample preparation and delivery for X-ray free-electron laser experiments

The ‘probe before destroy’ approach requires the sample to be replenished in the interaction region between individual XFEL pulses. For the 120 Hz operation frequency of LCLS, this means that the sample has to be fully replaced within 8.3 ms, and the displacement has to be on the order of hundreds of micrometres to ensure that no previously illuminated sample remains in the interaction region. Furthermore, the sample position has to overlap reproducibly with the X-ray focus that is a few μm$^2$ in size. In addition, the sample is often delivered into a vacuum environment to minimize background signal from air scattering. An approach based on a liquid jet or a droplet dispenser can fulfil these requirements. Typically, a crystal suspension of about 10$^9$ crystals ml$^{-1}$ is continuously delivered from a reservoir through a silica capillary to near the X-ray interaction region where the suspension is then focused to a thin liquid microjet in vacuum before interacting with the X-rays. The most common liquid microjet for serial femtosecond crystallography (SFX) experiments, the gas dynamic virtual nozzle (GDVN) [20], uses a gas sheath to focus the liquid stream. It has a typical flow rate of 10–30 μl min$^{-1}$, and thus consumes 7–20 ml volume of sample for a single 12 h shift of an SFX experiment [21]. This large sample consumption is not sustainable to study reaction cycles of samples that are difficult to crystallize or only available in limited amounts, such as PSII.

The physics of electric field focusing leads to microjet formation similar to that from gas focusing, suggesting that electrospinning has the potential for SFX sample delivery. We therefore designed an injector based on the principle of electrospinning in vacuum [22], where an electric field is applied to focus the liquid jet emitted from a Taylor cone [23]. This cone results from the exposure of an electric field to a small volume of electrically conductive liquid, which deforms the liquid shape beyond that from surface tension alone, and ejects a jet of liquid upon reaching a certain threshold voltage. By adjusting the field strength together with the liquid viscosity, it is possible to control the length of the microjet at the end of the Taylor cone before it breaks up into droplets. In addition to the liquid viscosity, the capillary diameter and backing pressure applied have a significant influence on the flow rate. In our design (figure 1a), the sample is loaded in a pressure cell at 15–20 psi to maintain the desired flow rate. Owing to the low backing pressure, when compared with the GDVN system, samples can be injected directly from standard autosampler vials. The vials are sealed by a screw cap with a septum that is pierced by the sample capillary, the counter electrode and a gas line to pressurize the vial. The use of only one continuous capillary with tapered outer walls at both ends minimizes the possibility of particle aggregation and subsequent injector clogging and reduces the surface area at the exit of the capillary, leading to an easier jet formation and less solution creeping up along the capillary. A positive high voltage of up to 5 kV can be applied in the liquid reservoir and a negative voltage of up to −5 kV can be applied to the counter electrode to focus the microjet at the capillary exit in vacuum. The open geometry design around the interaction region also allows collecting X-ray diffraction and emission data simultaneously (figure 1a,b).

Currently, the electrospinning jet flow rate can be tuned from 0.15 to 3.1 μl min$^{-1}$ [22]. In our PSII experiments (I24,26), see also [8] for review), we used crystals of about 5–20 μm in the longest dimension at concentrations of about 10$^7$–10$^8$ crystals per 1 ml buffer. For capillaries of 100 μm inner diameter, crystals up to 30 μm in the longest dimension were successfully injected. Typical flow rates used for suspensions of PSII crystals were in the range of 0.3 μl min$^{-1}$ (in 1.2 M sucrose buffer using a 75 μm ID capillary) to 3.1 μl min$^{-1}$ (in 30% glycerol buffer using a 100 μm ID capillary). The flow rate can be varied to optimize for minimal sample consumption, while allowing reasonable hit rate and providing compatibility with an illumination scheme to study the various states of PSII in its reaction cycle.

4. X-ray spectroscopic measurements on dilute biological samples

Many important enzymatic reactions rely on metal ions. Understanding the detailed reaction mechanism of these
metalloenzymes requires knowledge about the changes in the electronic structure of the metal in its binding environment during the reaction. X-ray spectroscopy provides an element-specific tool to monitor changes in the oxidation state, electron configuration, ligand–environment and bonding state of the metal. Combining specific X-ray spectroscopic techniques with the special features of XFELs allows for conducting these experiments with subpicosecond time resolution, under close to physiological conditions and without the manifestation of X-ray-induced changes in the measured data. The interaction of transition metal ions with X-rays can lead to absorption of an X-ray photon if the energy of that photon is sufficient to promote one of the bound electrons into a higher shell or into the continuum. This interaction can be measured by X-ray absorption spectroscopy (XAS), either in a transmission set-up as an absorption change or in a fluorescence set-up where the energy and intensity of the X-ray photons emitted from the sample is recorded. These processes can be observed both for the metal K-edge (absorption by the K shell) as well as for the metal L-edge (absorption by the L shells) depending on the incident X-ray energy. As the metal concentration in biological samples is often very low (submillimolar), measurement in transmission mode is rarely possible. In these cases, fluorescence detection is used instead as it allows for a higher sensitivity.

As shown in figure 1c, the hole created in one of the inner electron shells after the absorption process is subsequently filled by the decay of a higher shell electron, which takes place in the subfemtosecond timescale. This process leads to the emission of the respective energy difference between these electron levels in the form of X-rays, which can be collected by X-ray emission spectroscopy (XES) with secondary optics.

(a) K-edge X-ray emission spectroscopy
Contrary to XAS, where the lowest unoccupied orbitals are probed (figure 1c), XES probes the energy levels of the highest occupied orbitals [27]. The highest occupied orbitals are of special interest as they are involved in the actual chemistry during a reaction. XES has also some advantages over XAS for experiments at XFELs: (i) there is no requirement to have a monochromatic excitation pulse, as long as the...
photon energy of the pulse is sufficiently far above the edge of interest. This allows for the use of the full XFEL spectrum for the excitation, resulting in a flux increase of two orders of magnitude compared with monochromatized XFEL pulses. (ii) Energy-dispersive detection schemes for emission can be used, which allows a full spectrum to be collected for each X-ray pulse and therefore avoids errors owing to normalization (especially with fluctuations in pulse intensity, sample concentration, etc.).

In order to record Kβ XES (3p to 1s transitions) from transition metal elements, we developed an energy-dispersive spectrometer [28] (figure 1a,b). It comprises 16 analyser crystals arranged in von Hamos geometry. Each of the analyser crystals (110 x 25 mm in size with a bending radius of 500 mm) is equipped with three actuators to allow fine adjustment of yaw, pitch and distance between the analyser crystals and the interaction point, and the entire array covers a solid angle of 1.3% of the sphere for 6490 eV (Mn Kβ). Different analyser crystals (e.g. Si(440), Ge(620), etc.) can be used to cover the energy range of Kβ transitions from different elements such as Mn, Fe, Ni and Co. The spectrometer was designed to fit into the large focus vacuum chamber of the CXI instrument, sitting at 87° with respect to the incoming beam and at a distance of 500 mm from the interaction region. The X-ray emission signal is focused in the horizontal direction and dispersed in energy along the vertical direction. To achieve good focusing, the entire spectrometer is mounted on a translation stage to allow fine adjustment of the distance between spectrometer and interaction region. For detection of the energy-dispersed signal, a small version of the CSPAD detector (CSPAD140k) [29] is mounted underneath the X-ray interaction point facing the spectrometer. The signal, spread out over about 10 columns of pixels, is recorded for each individual shot, and the single spectra for thousands of shots are added to obtain sufficient signal-to-noise ratio.

The spectrometer was first tested at the LCLS measuring XES of Mn model compounds—MnIIICl2 and MnIIIUPy—terpyridine in aqueous solutions at room temperature [30]. The recorded spectra agreed with spectra measured at cryogenic temperature [31].

(b) Fluorescence-detected L-edge X-ray absorption spectroscopy

While Kβ XES can probe the electronic structure of transition metals, it cannot directly probe the transitions from the valence electrons that are mostly located in d-orbitals, as the d–s transition is not dipole-allowed and therefore very weak. Therefore, it provides only an indirect indication of the occupancy of the d-orbitals owing to the p–d spin-exchange interaction. By contrast, p–d transitions (figure 1c) are dipole-allowed and can directly provide information about the energy levels of the valence electrons. Despite its chemical sensitivity, the transition metal L-edge spectroscopy is not widely used for biological samples for two main reasons. First, the radiation damage is orders of magnitude faster than that of transition metal K-edge spectroscopy owing to the higher absorption cross section. Second, binding energies of elements are close to each other. For example, the L-edge of Mn (640 eV) is located very close to the oxygen K-edge (560 eV). As a result, the Mn L-edge fluorescence signal of dilute biological systems (concentration on the order of mM) is concealed by a huge background of oxygen fluorescence signal from the protein itself (approx. 1 M) and the aqueous buffer (55 M).

In order to discriminate these two signals and collect intact data by taking advantage of the short XFEL pulse, we developed a high-transmission spectrometer that relies on a reflection zone plate (RZP) as the sole X-ray optical element [25]. This zone plate was designed to focus photons of 640 eV energy into a small spot, spatially well separated from the region where photons originating from the oxygen K-edge hit the detector (figure 1d). In contrast to conventional soft X-ray spectrometers [31], the RZP-spectrometer offers less energy resolution ($\Delta E / E$ of about 100 compared with 1000) but orders of magnitude higher efficiency. For detection of the signal, an X-ray CCD camera was mounted in the focal plane of the RZP. In a first proof-of-principle experiment, we used this set-up at the SXR instrument of LCLS and collected Mn L-edge spectra from aqueous solutions of MnCl2 at 500 mM Mn concentration (figure 1d [25]), which confirmed the feasibility of collecting intact Mn L-edge data at room temperature. Based on this result, further experiments on PSII are planned.

5. Software development to process X-ray free-electron laser data

To process XFEL data, several specific features of the experiment have to be taken into account. Owing to the stochastic nature of the lasing process, the intensity and spectral distribution of the incident light is different for each single XFEL pulse. In addition, each probed sample volume is destroyed during the exposure and therefore all the information necessary to scale and integrate the measurements has to be obtained from a single X-ray shot. One other aspect is the shear volume of data acquired owing to the high frequency of the pulses; at 120 Hz, the data rate when using two full CSPAD and one CSPAD140k at the CXI instrument is about 1.3 Gb s⁻¹. In order to use the very limited beam time as efficiently as possible, it is necessary to have a fast feedback regarding the data quality during the measurement.

To address all these challenges, we developed a new software package cctbx.xfel [32,33] based on the computational crystallography toolbox cctbx. One important feature is the online monitoring. The tool can be customized to display hit rates, numbers of spots per image and monitors various other signals and beam line parameters (e.g. XES detector signal, reported beam energy, status of laser shutters, motor positions, etc.; figure 2a) in near-real-time [33]. The principles of this software package will be described here using the
protein thermolysin. A review of the work done with PSII is described elsewhere [8].

(a) Crystal diffraction data processing

For processing of crystallographic data, the software package has to address some specific additional challenges: (i) individual Bragg spots often cover only a few pixels owing to the small point-spread function of the CSPAD detector (figure 2b); (ii) the CSPAD detectors are multi-tiled detectors, and the relative position and orientation of all detector elements has to be known with subpixel accuracy for successful indexing and integration; (iii) there can be more than one crystal in the probed volume leading to diffraction patterns of several lattices on the same image (figure 2c); (iv) the orientation for each crystal has to be determined from a single still image; (v) each diffraction image is different with respect to beam energy, signal intensity and diffracting power of the crystal (figure 2c); and (vi) the signal especially at high q is often dominated by background noise [34].

The first step in the data processing is triaging via a spotfinder [35], as implemented in cctbx, with threshold settings optimized for each individual dataset. Indexing is then performed with LABELIT [36], using algorithms from Rossmann’s data-processing suite. To improve the success rate of the indexing algorithm, we implemented the possibility of providing a target unit cell. This is especially helpful for SFX experiments, where images often contain few spots, and there is no possibility of observing the same crystal under two different orientations, in contrast to the procedure used for indexing in normal oscillation crystallography. This approach can even be applied to data from unknown unit cells: indexing can initially be attempted without a target unit cell and the resulting mean cell parameters can then be used as the target unit cell for a second round of indexing, thereby increasing the number of indexable lattices. In SFX, a single pulse occasionally exposes two or more crystals, leading to several lattices on the same detector image. cctbx.xfel handles this situation by first indexing the dominant lattice and then performing a second indexing run using the spots not accounted for to index the weaker lattice (figure 2c).

In order to correctly predict the position of Bragg spots on the detector, it is necessary to know the arrangement of the detector tiles of the CSPAD with high precision, especially as many Bragg spots are only one or two pixels in area. The orientation and position of the detector elements are refined starting from a rough starting model based on optical measurements of the tile positions. Next, tile positions are iteratively refined by nonlinear least-squares minimization of the distance between observed Bragg reflections and reflections predicted by indexing the spots on an entire dataset of a model protein. The final improved detector model then results in a subpixel RMS deviation of the observed versus the predicted spot positions. The optimization of the detector position has a significant impact on the integration of high-resolution reflections. For example, in the thermolysin dataset...
[32], a perturbation of detector positions by just one pixel leads to loss of about 30% of the high-resolution signal in the dataset when compared with the indexing using the optimized detector metrology.

In order to minimize the contribution of noise to the signal, it is important to model the shape of the Bragg spots precisely. ccxtbx.xfel implements two different spot models, an empirical and a parametric model [32]. The empirical model derives the spot shape for each individual spot from the union of the 10 strongest Bragg spots located in the vicinity of the spot. The mask constructed this way defines the pixels used for integration. The second, more sophisticated model aims to account for the detailed difference in individual spot shapes using a physical model. As a first approximation, the finely textured spectrum of the stochastic SASE pulse is modelled by a top hat function bounded by a low- and a high-energy limit. The other main factor influencing the spread of a spot is the mosaicity of the individual crystal, which is introduced as a third parameter. By refining the energy limits and the mosaicity for each observed lattice, the extent of each spot on that lattice can be predicted, yielding a more accurate spot model (figure 2B). This way the number of non-signal pixels included in the integration mask for each individual spot can be further reduced and the noise contribution to the signal be minimized.

One other important aspect to reduce the contribution of noise to the data is to estimate the resolution cut-off for each individual lattice indexed. As each observed lattice derives from a different crystal measured under slightly different conditions (orientation, beam intensity and sampled crystal volume), it is expected that each lattice exhibits a different diffractive power. A single, global resolution cut-off would either discard valuable information for well-diffracting crystals, or dilute the information content by integrating noise for poorly diffracting crystals. Instead, ccxtbx.xfel computes an individual resolution cut-off for each lattice. Paired refinement can subsequently be used to find the overall high-resolution limit of the merged dataset as the resolution at which the data fail to contribute useful information to atomic model refinement [37].

A detailed description of ccxtbx.xfel and the results of processing thermolysin and lysozyme XFEL diffraction data with our software package can be found in Hattne et al. [32].

(b) X-ray emission spectroscopy data processing

The processing of the XES data for PSII is challenging owing to signal levels well below one photon per pixel. To reliably extract the single-photon signals for each shot, care had to be taken to determine the gain of each individual pixel of the CSPAD140k [24,30]. After initial dark-current (pedestal) subtraction and frame-noise correction for each image, we constructed histograms of the number of analogue–digital counts read out for each pixel on each image frame. By fitting Gaussians to the zero- and one-photon peaks of these histograms, dark and gain corrections to the histograms directly from the data are possible. The correction was applied, so that the zero-photon peak is centred at zero analogue-to-digital units (ADUs), and the separation between the zero- and one-photon peaks is identical for all pixels, thereby correcting for the gain variations between the different pixels. From these gain-corrected ADU values, it is then possible to extract the real photon counts for each pixel and reconstruct the spectrum from these values. All these signal processing procedures were implemented in the ccxtbx framework. We found that we can measure good-quality XES data at LCLS from a few picomoles of Mn [24]. In practice, however, the total sample consumption is higher owing to the use of a continuously running jet and a limited hit rate.

6. Summary

With the methods briefly described in this contribution, we conducted XRD and XES experiments on PSII [24,26], also reviewed in [8]. Using microcrystals and our electrospinning jet, we demonstrated that we can measure and process diffraction data from a large radiation-sensitive metalloprotein complex, and especially that we can obtain electron density for the Mn cluster in PSII at room temperature by this method [26]. Further refinement of our crystallization protocol, injection set-up and inclusion of the X-ray emission spectrometer allowed us to simultaneously measure XRD and XES from PSII crystals at RT [24]. In this experiment, we confirmed that the ‘collect before destroy’ approach also works for the highly radiation-sensitive Mn cluster and we observed diffraction from PSII microcrystals to around 4 Å. We also collected data not only from the dark resting state, but also from the first light-excited state in the reaction cycle of PSII [24]. The results for PSII obtained by us so far are described in more detail in the contribution by Tran et al. [8]. Further improvements of both the methods and sample quality are underway, and future XFEL experiments are planned to follow PSII through its reaction cycle with time-resolved simultaneous XRD and XES studies using in situ illumination at RT. We are also planning to complement these studies with time-resolved spectroscopic studies of the Mn L-edge and K-edge of PSII in solution at LCLS. Altogether, these studies will allow a better understanding of the changes in the geometrical and electronic structure of the catalytic Mn-centre in PSII during the light-driven water oxidation reaction.

Acknowledgements. We thank all our colleagues who have been involved in various parts of this work over the past years. In particular, we want to thank Johannes Messinger and his group, Paul Adams and his group, Mike Bogan, Athina Zouni and her group, and Alexander Foehlisch and his group. We also want to thank the staff at the CXI and SRX instruments at LCLS.

Funding statement. This work was supported by NIH grant no. GM05302 (V.K.Y.) for PSI1 biochemistry, structure and mechanisms, the Director, Office of Science, Office of Basic Energy Sciences (OBES), Division of Chemical Sciences, Geosciences and Biosciences (CSGB) of the Department of Energy (DOE) under Contract DE-AC02-05CH11231 (J.Y., V.K.Y.) for X-ray methodology and instrumentation, by NIH grant no. P41GM103393 for part of the XES instrumentation and support of U.B.; an LBNL Laboratory Directed Research and Development award (DOE contract DE-AC02-05CH11231) to N.K.S. and NIH grants GM098887 and GM102520 (N.K.S.) for data-processing methods. U.B., P.W. and J.Y. also acknowledge support through a Human Frontier Research grant (no. RGP0063/2013) for spectroscopy on photosystem II. We also acknowledge support through the Alexander von Humboldt Foundation (J.K.) and the Ruth L. Kirschstein National Research Service Award (F32GM100595, R.T.). The injector work was supported by DOE Office of Basic Energy Sciences, Chemical Sciences Division, under Contract DE-AC02-05CH11231 (J.Y., V.K.Y.) for X-ray methodology and instrumentation, by NIH grant no. P41GM103393 for part of the XES instrumentation and support of U.B.; an LBNL Laboratory Directed Research and Development award (DOE contract DE-AC02-05CH11231) to N.K.S. and NIH grants GM098887 and GM102520 (N.K.S.) for data-processing methods. U.B., P.W. and J.Y. also acknowledge support through a US Department of Energy Office of Science by Stanford University.
References