UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Studying Photosystem II at X-ray Free Electron Lasers

Permalink

https://escholarship.org/uc/item/35h3z2j3

Author

Doyle, Margaret Donovan

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Studying Photosystem II at X-ray Free Electron Lasers

by

Margaret Donovan Doyle

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Applied Science and Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Roger Falcone, Chair
Dr. Junko Yano
Professor Graham Fleming
Professor Dave Attwood

Summer 2024

Studying Photosystem II at X-ray Free Electron Lasers

Copyright 2024
by
Margaret Donovan Doyle

Abstract

Studying Photosystem II at X-ray Free Electron Lasers

by

Margaret Donovan Doyle

Doctor of Philosophy in Applied Science and Technology

University of California, Berkeley

Professor Roger Falcone, Chair

X-ray free-electron lasers (XFELs) enable the study of biological systems under functional conditions. This has sparked a scientific revolution, beginning around the time when the world's first hard X-ray FEL, the Linac Coherent Light Source (LCLS) at the Stanford Linear Accelerator Center (SLAC) became operational in 2009. By focusing a large number of photons into a small area, and having them all arrive within a few femtoseconds, the world of ultra-small and ultra-fast phenomena of biochemical processes opened up significantly.

These machines operate through a process called Self Amplified Spontaneous Emission (SASE), which begins from the shot noise of an electron beam. These electrons are accelerated to relativistic speeds before they are fed through a periodic arrangement of magnets, called undulators, which forces them to weave and emit bright bursts of X-ray light. The power of the light gets amplified along the length of the undulators as the electrons self-organize through the SASE process. The electrons are eventually discarded, while the ultra-short and ultra-bright pulses of X-ray light are directed towards experimental halls to serve as powerful probes in a variety of scientific investigations. Further details surrounding XFEL operation, as well as the data handling considerations which stem from high repetition rate data acquisition paired with customized detectors, are discussed thoroughly in this dissertation.

An optimal candidate for XFEL studies is Photosystem II (PS II), the enzyme responsible for catalyzing the light-induced oxidation of water to molecular oxygen. Nature has perfected this harvesting of light energy and conversion to chemical energy over billions of years. This makes it an excellent target of study that could yield insights useful for the development of clean energy. PS II contains a metal catalytic site crucial for its function, called the oxygen evolving complex (OEC; Mn_4CaO_5), which undergoes a variety of structural and chemical changes during its light-induced reaction cycle. The surrounding water and hydrogen-bonding network also play an important role in enzyme function via the transport of protons and substrate water. These processes can be tracked through a combination

of time-resolved X-ray emission spectroscopy (XES) and serial femtosecond crystallography (SFX) at XFELs.

This dissertation covers three studies at the intersection of XFELs and PS II. The methods described may be applied to other metal containing enzymes as well, and need not be specific to PS II. Collecting time resolved SFX simultaneously with XES allows for the structural changes of the protein and the chemical state of the metal catalytic site to be bridged and reaction kinetics to be studied. Interpreting the XES data collected at XFELs from low concentration metalloenzymes can present its own challenges due to relatively low signal levels and chemical sensitivity. Therefore, this dissertation will also discuss an innovative technique at XFELs which uses the stimulated XES method. This approach, analogous to the stimulated Raman method in the optical regime, offers the potential to collect XES data with greater efficiency. It may also be used for uniquely probing the less intense, but more chemically rich, spectral transitions.

Interpreting the rich information content contained within the electron density maps collected at XFELs during SFX experiments can also bring about challenges, especially in the context of PS II's water networks; however, crystalline molecular dynamics (MD) simulations can serve as a promising computational tool to aid in the interpretation of the electron density maps. The simulation work described in this dissertation has the potential to extract more information from room temperature XFEL crystallography data, especially in the context of water dynamics and hydrogen bonding networks, and also holds the promise of improving our crystallographic refinement.

To the loving memory of William Curtis Doyle.
Engineer, innovator, and exceptionally supportive grandfather.
January 19, 1929 - November 27, 2015

Contents

Contents **ii**

List of Figures **vi**

List of Tables **xviii**

1 Introduction **1**
1.1 Background and Purpose 1
1.1.1 Timescales and Challenges for Monitoring Dynamic Processes in Biological Systems 1
1.1.2 Harnessing the Power of the X-ray Free Electron Laser 3
1.2 Metalloenzymes 4
1.2.1 Importance in Biological Processes and Significance of Manganese . . 4
1.2.2 Photosystem II 5
1.3 Overview of Techniques Used 7
1.3.1 X-ray Emission Spectroscopy 9
1.3.2 Stimulated X-ray Emission Spectroscopy 10
1.3.3 Molecular Dynamics Simulations 12

2 X-ray Free Electron Lasers **15**
2.1 Origin of XFELs 15
2.2 Synchrotron Radiation 17
2.3 From Undulator to XFEL 21
2.3.1 Self-Amplified Spontaneous Emission 23
2.4 Seeding Techniques 24
2.5 Time Structure 25
2.6 Outlook 26

3 Navigating Data-Related Challenges **28**
3.1 Introduction to Data Reduction at XFELs 28
3.1.1 Importance of Data Reduction 28
3.1.2 Data Reduction Pipeline for XES 29

3.1.3 Challenges and Considerations . . . 31
3.2 Exploring Machine Learning Models for Spectral Classification . . . 34
3.3 Image Curvature Correction Algorithm . . . 38
3.3.1 Origins and Impact of Image Curvature . . . 38
3.3.2 Center-of-Mass Algorithm . . . 39
3.3.3 Image Correction Results . . . 41
3.4 Outlook . . . 42

4 X-Ray Emission Spectroscopy Measurements at XFEL 44
4.1 Introduction . . . 44
4.2 Experimental Overview . . . 46
4.2.1 Spectrometer . . . 47
4.2.2 Experimental Conditions . . . 49
4.3 In Situ Diagnostics . . . 49
4.4 Studying Reaction Kinetics Through X-ray Emission Spectroscopy . . . 51
4.5 Results . . . 55
4.5.1 Kinetics of Mn Redox Chemistry During Water Oxidation in Photosystem II Monitored Using an X-ray Free Electron Laser . . . 56
4.5.1.1 Abstract . . . 56
4.5.1.2 Introduction . . . 56
4.5.1.3 Results and Discussion . . . 58
4.5.1.4 Mn K$\beta_{1,3}$ XES During the Reaction Cycle . . . 58
4.5.1.5 Kinetic Models . . . 60
4.5.1.6 Evaluation of the Models . . . 63
4.5.1.7 Reaction Sequence Leading to O_2 Formation . . . 65
4.5.1.8 Conclusions . . . 67
4.5.2 Supplementary Information . . . 68
4.5.2.1 Materials and Methods . . . 68
4.5.2.2 Time Constants from Other Studies . . . 70
4.5.2.3 Kinetics Modeling . . . 70
4.5.2.4 Results from Different Experiments . . . 72
4.5.2.5 Population Trends of Models . . . 72
4.5.2.6 Sensitivity to Parameter Choices . . . 73
4.6 Outlook . . . 77

5 Enhancing Chemical Sensitivity With Seeded Stimulated Emission 79
5.1 Introduction . . . 79
5.2 Seeded Stimulated X-ray Emission at 5.9 keV . . . 83
5.2.1 Abstract . . . 83
5.2.2 Introduction . . . 84
5.2.3 Experimental Setup . . . 85
5.2.4 Establishing the Pump and Seed Pulse Parameters for SSE . . . 86

5.2.5 Spectral Analysis and Seeding Efficiency . . . 91
5.2.6 Conclusion . . . 93
5.3 Two-Color Mode . . . 94
5.4 Observation of Multiplet Lines in Seeded Stimulated Mn Kα X-ray Emission 97
5.4.1 Sodium Permanganate . . . 97
5.4.2 Manganese Chloride . . . 100
5.5 Outlook . . . 103
5.5.1 Seeded Stimulated Emission on Low Concentration Metalloenzymes . 103
5.5.2 Exploring Manganese Complexes with Different Multiplet Structures 103
5.5.3 Phase Stable Pulse Pairs . . . 104
5.5.4 An X-ray Laser Oscillator . . . 104

6 Molecular Dynamics Simulations on Photosystem II 108
6.1 Introduction . . . 108
6.2 Crystalline Molecular Dynamics Simulation Preparation . . . 109
6.2.1 The Case for Crystalline MD . . . 109
6.2.2 Ensemble Selection . . . 110
6.2.3 Restraints in Production Simulations . . . 110
6.2.4 Computing Densities from MD trajectories . . . 111
6.3 Water Networks in Photosystem II Using Crystalline Molecular Dynamics Simulations and Room Temperature XFEL Serial Crystallography . . . 112
6.3.1 Abstract . . . 112
6.3.2 Introduction . . . 113
6.3.3 Methods . . . 116
6.3.4 Molecular Model and Force Field Parameters . . . 116
6.3.5 MD Simulations . . . 117
6.3.6 Mean Structure Factors . . . 118
6.3.7 Experimental XFEL Crystallography Data . . . 119
6.3.8 Results . . . 119
6.3.9 MD Density and Comparison with XFEL Experimental Data . . . 122
6.3.10 O4 Channel . . . 126
6.3.11 O1 Channel . . . 129
6.3.12 Cl1 Channel . . . 131
6.3.13 Discussion . . . 133
6.3.14 Networks for Proton Transfer . . . 134
6.3.15 Water Transport . . . 136
6.3.16 Conclusions . . . 137
6.3.17 Supporting Information . . . 138
6.3.17.1 Methodology . . . 138
6.3.17.2 Mean Structure Factors (Continued) . . . 138
6.3.17.3 Comparing MD Map to Experimental Map . . . 139
6.3.17.4 MADI: Map Based Donor Acceptor Identification . . . 139

6.4 Outlook 164

Bibliography **166**

List of Figures

1.1 The time scale of various biological processes (top) and the temporal resolution of major techniques deployed in biophysical studies (bottom). The tools explicitly discussed in this dissertation are colored red. Adapted from [3, 4, 5, 6]. 2
1.2 Simplified schematic of the thylakoid membrane and the protein complexes involved in the photosynthetic electron transport chain. Electrons follow the indicated path (red arrows) to the stromal side of PS I where they are used to reduce $NADP^+$ to NADPH. Adapted from [26, 27]. 6
1.3 Schematic of possible X-ray interactions with matter. Adapted from Andrew Minor's Materials Characterization course (MATSCI 204) at UC Berkeley. . . . 8
1.4 A simplified version of the respective transitions which result in Kα and Kβ emission (left) and the resulting spectra collected from Mn foil (right), adapted from [41]. 9
1.5 Schematic of relevant processes involved in stimulated X-ray emission, adapted from [1] and [42]. Illustrated differences spontaneous emission and stimulated emission (a). Difference between ground state and excited state (b). In ASE, a spontaneously emitted photon (red) acts as seed to stimulate the emission of a second photon from a core excited atom (c). In SSE, an external seed pulse (red) stimulates the emission of photons from core–hole excited atoms along the seeding direction (d). 11
1.6 Simplified MD simulation workflow, adapted from [6]. 12

2.1 An historical overview of X-ray radiation sources. (a) Evolution of peak brilliance across generations of synchrotron radiation sources. (b) A concise timeline of X-ray Free Electron Laser history. Adapted from [58] and [59]. 17
2.2 Diagram illustrating the three types of magnetic structures generating synchrotron radiation, along with how their respective intensity scales with number of electrons N_e and number of poles N_{poles}, adapted from [1]. 18
2.3 Generation of narrow cone undulator radiation. As electrons traverse the magnetic structure (a) they oscillate and generate a toroidal $sin^2\Phi'$ radiation pattern in the electron's frame of reference (b). In the laboratory frame of reference, the shape is contracted and folded into a narrow forward cone with an angle of $\theta = \frac{1}{2\gamma}$ (c). Adapted from [1]. 19

2.4 Undulator tuning curve schematic, adapted from [1]. 20
2.5 Evolution from undulator (a) to XFEL (b) accompanied by the resultant gain curve (c). As the undulator length extends, the initially randomly distributed electrons evolves, exhibiting an increasing degree of order before reaching the highest degree of electron modulation, otherwise known as "saturation." Once this point is reached, continued passage through the undulator will not be beneficial. The gain curve visually represents the amplification achieved in the XFEL process. Adapted from [1], and originally from [62]. 22

3.1 Flowchart outlining the typical data reduction steps for XES data collected at an XFEL. First, all raw data is written out to XTC format. The resulting files are quite large and must be reduced. They are then translated into HDF5 files. Finally, analysis is taken care of in a Jupyter notebook environment. 29
3.2 (a) ePix camera, reproduced from [84]. (b) Example detector image reconstructed from the recorded data (Mn K$\beta_{1,3}$). The spectral direction is shown along the vertical axis, and the spatial direction is shown along the horizontal axis. . . . 30
3.3 Schematic of the structure of an HDF5 file generated during an XES experiment. The blue node represents the root group within the HDF5 file. Green nodes represent datasets within the group. Arrows between nodes represent relationships between groups and datasets - with the direction of the arrow indicating the parent-child relationship. 31
3.4 When the signal is collected on a detector with small enough pixel size, the charge cloud gets spread across adjacent pixels in what is known as "charge-sharing." A reconstruction algorithm is required to correct this. 32
3.5 ADU histogram with three prominent features below 200 ADU consisting of an artifact of the detector, the X-ray fluorescence signal, and elastic scattering. The black vertical lines indicate the region of interest. There are two weak peaks at higher energies corresponding to multi-photon hits (around 220 and 280 ADU) which can be incorporated into analysis. The peak around 220 ADU comes from two XES photons, and the peak around 280 ADU comes from one XES photon plus one elastically scattered photon. 33
3.6 Example 2D detector image (left) and corresponding spatial projection (right) of a nickel Kα emission signal. The signal ROI (red dashed lines) and background ROI (orange dashed lines) are marked as well. 34
3.7 Example single shot detector images from each class used constructing my model. 35
3.8 Histogram showing distribution of images for each class used. 36
3.9 Confusion matrix demonstrating performance of the ML model with predicting spectral line and probed metal. 37
3.10 Screen shot of the resulting flask-based web application, showing the page where interested users can upload a 2D detector image and obtain a prediction of the metal probed and transition line measured. 37

3.11 Identifying curvature of the raw detector image. In this image, only a seed pulse is present. 38
3.12 2D detector image of monochromatic seed (left) and its y-projection (right) with the corresponding Gaussian fit. 39
3.13 Linear interpolation to optimized constants in quadratic fit which are used to correct the detector images. 40
3.14 Schematic of full image correction procedure. First, each image is divided into a set of horizontal bars of defined slice width. Next, the interpolation functions are applied for that slice's energy, generating shear parameters. The entire image is sheared, and the slice of interest is placed into a brand new image. The process is repeated for every horizontal slice until the entire image is corrected. 41
3.15 Visual representation of a detector image before (left) and after (right) correction. The broad "cloud" is a stimulated XES signal (see **Chapter 5** for more details), and the intense and narrow feature is the monochromatic seed. The two features become more parallel to each other, as they should be, after the correction is implemented. 42

4.1 Schematic of the drop on tape setup for simultaneous collection of XES and XRD data, reproduced from [113] 47
4.2 Experimental setup geometry for XES. 49
4.3 First moment trends from PS II collected for both crystal and solution, collected across nine different experiments using different experimental stations, samples, sample delivery systems, laser set-ups, detectors, and other parameters. 51
4.4 Schematic of "cross-illumination", which typically results from light scattering off a neighboring droplet which "sees" the laser first. It does not correspond to a droplet that gets a double hit. 52
4.5 Comparing flash and pure state first moment trends from the experiments in Fig. 4.3 53
4.6 Procedure for calculating first moment trends. 55
4.7 Schematic representation of the XES measurement setup at an XFEL. (a) Structure of PS II (pdb ID 6DHO). (b) The Mn_4CaO_5-O_x cluster (in S_3 state) and its reaction (Kok) cycle. (c) Schematic depiction of the pump-probe sequences for measuring time-points between S_i states. (d) K$\beta_{1,3}$ transition that gives rise to the XES emission spectrum measured by employing a von Hamos spectrometer and a position sensitive 2D detector (PSD). Data were obtained with *T. vestitus* PS II core complexes in solution at room temperature. See SI for details on the measurement conditions. 58
4.8 (a) Mn K$\beta_{1,3}$ XES of PS II in different flash states: raw spectra and smoothed spectra are shown. Data are area-normalized over the region from 6483 to 6497 eV. Insets show the zoom in view for 6493.5-6494.0 eV. (b) First moments of flash states (with prevalent S-state indicated). Error bars were calculated from bootstrap estimates. The first moment energy is offset by 6491 eV [87]. 60

4.9 Scenarios modeled in this work for the S_3 to S_0 transition. (a) Schematic depiction of the three variables in the eight scenarios considered in the modeling. (b) Tabulated breakdown of models used in this work, including Mn redox steps anticipated in each model. * Indistinguishable from S_3 oxidation state ($Mn^{IV}{}_4$) by XES, only indicated if long lived due to H^+ release ** Two options: Mn oxidation states are $Mn^{IV}{}_4$-$O^{\cdot}$ and thus indistinguishable by XES from S_3 state ($Mn^{IV}{}_4$) and/or state is so short-lived that it does not accumulate 62
4.10 Calculated Mn XES first moment changes for the eight scenarios for the $S_3 \rightarrow S_0$ transition (A through H, solid lines) (see Fig. 4.9 b), compared to experimental data (blue markers), with small off-set in first moments at nF(0 ms) shown in dark purple (see the SI for details). Upper left panel shows the calculated trends of all models (overlapping gray lines) for $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$, and the remaining panels show $S_3 \rightarrow S_0$ for each model (A through H) individually. The $S_3 \rightarrow S_0$ panels show the oxidation states and include insets showing times up to 1 ms, and rates $\pm$ 25 % are shown as a gray area. Colored horizontal arrows illustrate time-resolved populations of intermediates – see Fig. 4.14 for a detailed explanation. Kinetic rates and mean-squared deviation (MSD) are provided for each model in the form of time constants (τ), which are related to the half-life ($\tau_{\frac{1}{2}}$) by $\tau_{\frac{1}{2}} = \ln(2)\tau$. Error bars were estimated for each data point by bootstrapping as described in the reference [87]. 63
4.11 Estimated first moment and population trends using model D. Top: Calculated trends (black) compared to experimental measurements (blue), with small off-set in first moments at nF(0ns) shown in dark purple (see text for details). Bottom: Time-resolved populations, including the proposed reduced intermediate ($Mn^{III}Mn^{IV}{}_3$). The populations of intermediate $Mn^{III}Mn^{IV}{}_3$ for the $S_0 \rightarrow S_1$ and $S_2 \rightarrow S_3$ transitions have been removed, for clarity (see Fig. 4.14). 65
4.12 Proposed S_3 to S_0 chemical process based on the kinetics of the Mn redox changes, in the context of the structural data by Bhowmick et al [131]. The rate constants of the transition into and out of the peroxide-like intermediate, after a 200 μs lag (τ_1), are estimated as 100 (τ_2) and 2100 μs (τ_3), respectively. The proposed structural models of the $Mn4CaO_{5/6}$ cluster at each step and oxidation state of each Mn are shown. The location of the MnIII in the peroxide-like intermediate depends on the model for O-O bond formation. 67
4.13 Measured and modeled first moment trends for experiments R1-R2, using model D with rates from Fig. 4.10. Each nF corresponds to data collected at 200 ms after n flashes, such that the sample is predominantly in the intermediate S-state reached by this time (S_1, S_2, S_3, and S_0, for n=0,1,2,3). 72
4.14 Population trends of the different models, using rates from Fig. 4.10. 73
4.15 Using [S_x] intermediates with different first moments for [S_x] (top), or this intermediate followed by partial reduction (bottom). Varying [S_x] around [S_3], with rates from models B and D. 74

4.16 Varying rates for model D, including varying the ratio between long- and short-lived intermediates (first panel), the ratio between the two short-lived intermediates (second panel), and varying the total short (third panel) and long (fourth panel) rates by ±25%. Increase in (first) rate shown as dotted line, decrease as dashed line. 75
4.17 Modeled first moments for model D, varying rates and sample turnover parameters. Increasing and decreasing the parameters, with results shown as dotted and dashed lines, respectively. 76
4.18 Sensitivity of models A (a), D (b), and H (c). Calculated from 10 000 iterations, varying sample turnover parameters and rates (left), as well as standard deviations for all combinations (right). 77

5.1 Simulations of spontaneous Kα emission from $MnCl_2$ and $NaMnO_4$, including transition lines, which build each spectrum, shown as vertical sticks. 80
5.2 The characteristic chemical shifts observed between ASE spectra collected from $MnCl_2$ and $NaMnO_4$ across two different facilities. 82
5.3 (a) Experimental layout (not to scale): SACLA XFEL radiation containing the seed and pump pulses is focused with a KB mirror system onto a liquid jet of $NaMnO_4$. 2% of the seed pulse is outcoupled and registered in the Seed Spectrometer. The resulting stimulated emission is dispersed onto the $K\alpha$ Spectrometer with a Si (220) crystal analyzer. (b) The level diagram and concept behind seeded stimulated emission (SSE). The pump pulse (yellow), which is tuned above the Mn K edge, creates a population inverted 1s core hole excited state. The seed pulse (blue) is tuned to the $K\alpha$ emission line in order to seed the SSE signal (red). (c) Schematic showing how the pump and seed pulse intensity is varied by moving the gain medium to various distances out of focus along the beam direction. 87
5.4 (a) Selected single-shot ASE spectrum collected at a pump intensity of 1.7×10^{19} W/cm^2; single-shot SSE spectrum (center) and seed pulse (right vertical line) at (b) $\delta E = 7$ eV above $K\alpha_1$ and (d) $\delta E = 0$ eV at a pump intensity of 1.1×10^{18} W/cm^2; (c) detuning scan of $K\alpha_1$ SSE at a pump intensity of 1.1×10^{18} W/cm^2, in which each horizontal row corresponds to the average spectrum for a given seed pulse energy collected across 1924 shots and each seed is indicated by a square box on the diagonal; (e) 1D version of (c), showing $K\alpha_1$ SSE spectra (solid red lines), in which the center of the seed pulse energy is indicated by a blue vertical line. Each 2D spectrum shown was collected at the $K\alpha$ spectrometer, and is peak normalized. Each 1D spectrum in (e) is normalized to the peak pixel value from (c). 90

5.5 Seed pulse spectrum obtained from averaging 1924 single shots, and the corresponding best fit from equation (5.1). In the inset to the left, the shot-by-shot seed intensity, which exhibits 100% fluctuation and has an estimated average value of 7.8×10^6 (black broken line) is shown. The average spectrum and corresponding fit were shifted to $\delta E = 7$ eV from $K\alpha_1$ for demonstration purposes. Next, two regions were selected from the shifted fit: $\delta E = 0$ eV (a) and $\delta E = 7$ eV (b). In region a), 2.0×10^4 photons were observed within $\Delta E_{K\alpha_1} = 2.1$ eV FWHM. In region b), 8.2×10^6 photons were found in a bandwidth of $\Delta E_{K\alpha_1} = 2.1$ eV FWHM. Integrated photon counts were computed from the fit. In the inset to the right, the single shot spectrum corresponding to Fig. 5.4 B is shown to illustrate how, even in the case of a detuned seed, enough photons from the low-energy tail are able to stimulate emission. 91
5.6 Left: The FWHM, f, of the Mn $K\alpha_1$ line as a function of the seed pulse detuning at different pump intensities. Lines are drawn for reference. Right: Example single-shot spectra, in which the seed is $\delta E = 0$ eV from the Mn $K\alpha_1$ line, to further illustrate spectral narrowing. 92
5.7 (a) Seeding efficiency as a function of seed photon counts that fall within the spectral range of $K\alpha_1$ FWHM for two different pump intensities. Shots are displayed for cases where the seed is in the range of $\delta E = [0, +7]$ eV for the case of 150 μm out of focus, and $\delta E = [0, +2]$ eV for the case of 400 μm out of focus. (b) Photon count distribution of $K\alpha_1$ and seed pulse photons from a seeded scan in focus, at a pump pulse intensity of 1.7×10^{19} W/cm^2. 93
5.8 Array of seed pulse scans collected without sample for a given undulator position. Intensity drops off towards lower energies and it appears to take on a Gaussian envelope (see equation (5.2)). Shot-to-shot fluctuations, as one may expect at an XFEL, do not appear to be prominent enough to compensate for going off of the monochromator - thus demonstrating the relative stability available at SACLA. 96
5.9 Left: Break down of how peakfinder algorithm works. First the raw signal is split into three offset components. Middle: Next, indices are retrieved where the reference intensity (blue signal) is greater than the intensity of either of the two offset signals(green and orange). Right: Finally, in the scenario where a cluster of peaks is found, the window of the peak is checked in order to verify the local maximum has been retrieved. 98
5.10 Illustration of the peak finder in action on a seeded stimulated emission spectrum, averaged over 1924 shots. 99
5.11 Top: Theoretical $NaMnO_4$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Results of the peakfinder algorithm deployed on average stimulated emission spectra. 100
5.12 Top: Theoretical $MnCl_2$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Waterfall plot showing averaged $MnCl_2$ spectra from a seeded scan through the Kα_1 envelope. Seed pulse photon energy is shown as the green vertical line. 101

5.13 Top: Theoretical $MnCl_2$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Results of the peak finder algorithm from the top 2% strongest shots. Light blue vertical bars are shown to guide the eye to partially reproducible peak positions. 102
5.14 Experimental setup (including estimated losses) for copper ASE and SSE measurements at CXI at LCLS. 105
5.15 ASE and seeded photon flux measurements for a jet solution and solid copper gain medium. Reproduced from [205]. 106
5.16 Crater size scaling laws resulting from SEM analysis. Reproduced from [205]. . 107

6.1 Overview of PS II and its electron donor site where water is oxidized to molecular oxygen. (a) The protein structure of PS II from Thermosynechococcus vestitus serves as the starting point for our MD model. The lipids (blue) are highlighted along with cofactors (green) and the three water channels (O1, pink; O4, cyan; Cl1, green). (b) OEC (only the $Mn_4CaO_5(H_2O)_4$ cluster) is shown in the context of the water channels, along with the Kok S state cycle, advancing through the S states (S_i, i = 1-4) by illumination with 1 through 4 flashes (1-4F). The S_1 state (solid red box) is the focus of the current work. 114
6.2 Methodological approaches used in this work. (a) The protein structure of PS II, which crystallizes in the P212121 space group, is first collected at an XFEL. The asymmetric unit is then expanded to the full P1 unit cell. Periodic boundary conditions are applied, wrapping the protein coordinates around the simulation box. The simulation is then carried out in a box approximately the size of the unit cell, with waters shown in light gray, protein atoms shown in green, and lipid/cofactor atoms shown in red. After carrying out the simulation, electron density maps are computed from the trajectory. (b) Comparison of experimental density (1σ / 0.173 e/Å^3) to the simulated water density (2.08 e/Å^3) is first required to make sure the XFEL electron density is reproduced with high fidelity. (c) Once the validation step is complete, simulated oxygen (purple, shown at 2.08 e/Å^3) and hydrogen (H1, cyan; H2, green, both shown at 0.26 e/Å^3 see the Supporting Information) densities are computed for the hydrogen bonding analysis. Average simulated oxygen and hydrogen peak positions, extracted from the densities, are marked as crosses. An example schematic, demonstrating our interpretation of the hydrogen and oxygen maps around W48 and W49, is shown on the far right. W48 and the H1 and H2 associated with it are likely fairly rigid over the course of the simulation, while the H1 and H2 associated with W49 are likely more mobile and are swapping into and out of the hydrogen bonding position. The nonbonded hydrogen can sit anywhere on the dotted cone. 121

6.3 Comparing the MD derived electron density for waters with the crystallographic electron density. (a) Crystallographically modeled waters around the OEC, color coded by their relative agreement to the MD density. Blue colored waters are less than 1.0 Å from a peak in the simulated electron density map, while gold colored waters are more than 1.0 Å away from the nearest simulated peak. The three water channels of interest: The O1 channel (red), Cl channel (green), and O4 channel (blue) are also overlaid. (b) Distribution of distances between crystallographic water positions and MD water density peaks within a 25 Å radius from the OEC (centered on O1). (c) Scatterplot showing the correlation between the MD and experimental electron density peak heights for different groups of MD waters in the inner (blue), middle (cyan), or outer (light blue) shell. (d) Example of a region with close MD water peaks and corresponding crystallographic waters. The simulated MD water density (gray contour at 2.08 e/Å^3) is shown in the context of the crystallographic waters (numbered blue spheres). The experimental 2Fo Fc map is overlaid (orange) and contoured at 1σ. (e) Example of a region with distant MD water peaks and corresponding crystallographic waters. The simulated MD water density (gray and contoured at 2.08 e/Å^3) is shown in the context of the crystallographic waters (numbered yellow spheres). Experimental 2Fo−Fc map is overlaid (orange) and contoured at 1σ. The water numbering convention for all waters in the vicinity of the OEC has been described previously. [206] The numbers of waters are based on the 0F (S_1 state) experimental crystal structure used by Hussein et al. [206] (PDB ID Code 7RF2). The channels were mapped by Caver 3.0 Pymol plugin using this room temperature crystal structure [257, 258] 124

6.4 Hydrogen bond and water dynamics in the O4 channel. (a) Simulated MD solvent map in the O4 channel (gray, contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure (numbered, in color). The results from the hydrogen bonding analysis are depicted as arrows, pointing from the donor toward the acceptor. The strength of the bond identified by our algorithm is represented by a color gradient as defined in the legend. (b) Rotated view of the W50−W51−W52−W53 region, showing that crystallographic waters W50 and W53 lie outside of the MD map wireframe. All simulated maps were carved at a 2 Å radius about each atom. Oxygen and hydrogen peaks from the MADI analysis, demonstrating the average orientation of waters showing favorable hydrogen bonding geometries in this channel, can be found in Figure 6.16. . 127

6.5 Drawbridge mechanism of ASN 338. Snapshots of the ASN structure from a single copy at (a) 90.94 ns, (b) 92.42 ns, and (c) 92.88 ns suggest that the ASN side chain movement is coupled with water transfer away from the OEC. The average solvent map from the entire trajectory across all four copies is overlaid and contoured at 2.08 e/Å^3. As the water molecule approaches ASN 338, it sits in the excess density, not found in the experimental map (blue circle). The side chain momentarily flips, potentially providing more room for the water molecule to hop over. After the water molecule crosses over and fills the W72 position, the drawbridge raises. It is observed to continue to raise and lower for the rest of the simulation. 128

6.6 Hydrogen bond and water dynamics in the O1 channel. (a) Simulated MD solvent map in the O1 channel (contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure. The results from the hydrogen bonding analysis are depicted as arrows, pointing from donor toward acceptor. The strength of the bond identified by our algorithm is represented by a color gradient. (b) W76 and W77, where there is significant disagreement between the map and crystal structure. (c) Crystal water and MD map agreement/H bond network in the region of the penta cluster waters (W26−W27−W28−W29−W30). All simulated maps were carved in a 2 Å radius about each atom. Oxygen and hydrogen peaks from the MADI analysis, demonstrating the average orientation of waters showing favorable hydrogen bonding geometries in this channel, can be found in Figure 6.19. 130

6.7 Hydrogen bond and water dynamics in the Cl1 channel. (a) Simulated MD solvent map in the Cl1 channel (contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure. The results from the hydrogen bonding analysis are depicted as arrows, pointing from donor toward acceptor. The strength of the bond identified by our algorithm is represented by a color gradient. (b) Continuous hydrogen bond wire from W3 to ASP 61, identified through our MADI analysis. Only the core participating waters are shown. Peaks found from MADI analysis in this channel can be seen in Figure S15. 132

6.8 Overview schematic of potential proton pathways, with arrows showing the direction of the hydrogen bond networks inferred for the S1 state, along with highly dynamic regions in each of the three channels. We are interpreting the low density peaks (in the 33rd percentile for waters within a 25 Å radius of the active site, centered on O1) as corresponding to more mobile waters, and the high density peaks (above the 66th percentile for waters within a 25 Å radius of the active site, centered on O1) as corresponding to more rigid waters. Two previously discussed bottleneck regions in the Cl1 and O4 channels are shaded. Possible proton pathways from the OEC to the bulk are indicated by distinctly colored dashed lines, labeled I−IV. Selected waters which are thought to participate in each pathway are labeled based on the numbering convention described previously [206]. . . . 134

6.9 Converging to comparable maps between the experimental $2Fo{-}Fc$ crystallographic densities (left, blue, contoured at 0.173 e/Å^3 which is equivalent to 1σ), and the MD densities (right, contoured at 2.08 e/Å^3). Images shown in Coot[272]. 141
6.10 Individually calculated water maps (colored) overlaid with the total average water map across all copies (grey) to demonstrate how mobility is classified. (Left) Example of how the individual water map looks for a rigidly fixed water. W20's individual map, which shows up as a visible and strong sphere of density, is comparable to the total water map when both are at the same 2.08 e/Å^3 level. (Right) On the other hand, the individual maps for more mobile waters like W32 W33 W34 are not visible at 2.08 e/Å^3, since they are not in any location long enough over the course of the simulation to build up substantial density. At 1 e/Å^3, individual maps for W32 and W33 become visible (and are offset from their crystallographic placements, with the density for W33 sitting closer to W32 and the density for W32 sitting in the W39 position). The density for W34, on the other hand, remains still too weak to be seen at this level due to extreme mobility. It only becomes visible at levels $<$ 0.7 e/Å^3 142
6.11 The full PS II dimer (PDB ID **7RF2**) which was used as the basis for our MD model. The ROI referred to throughout the text (centered on the O1 atom in the OEC with a 25 Å radius) is shown (yellow) along with its enclosed waters (dark blue). Note that the orientation of the dimer is flipped relative to how it is displayed in **Figure 6.1**. All outside waters are shown in cyan. The 'unlocked' monomer, which is the focus of this study, is shown in green. The 'locked' monomer [254] is shown as a sand colored surface. The orange region corresponds to contacts within 3.5 Å of the monomer monomer interface extended to the entire residue/ligand. 143
6.12 Histogram comparing the distribution of peak height of simulated waters inside the three channels of interest (O1, Cl and O4) vs. outside the channels. On average, waters inside the channels demonstrated overall stronger peaks. Such a finding suggests that positioning of waters within the channels is more anchored over the course of the simulation. Each bin corresponds to raw count divided by the total number of counts as well as the bin width. As a result, the area under each histogram is normalized to one thus providing a probability density. . . . 144
6.13 Demonstration of how crystallographic refinement can be improved through the combination of MD. (a) MD solvent map (grey and contoured at 2.08 e/Å^3) overlaid with the experimental 2Fo−Fc map (orange and contoured at 1 σ). (b) After an additional round of MD aided refinement, a new water, W202, was registered in this region (updated 2Fo Fc map is colored magenta and contoured at 1σ). (c) An additional MD solvent density, not found in the experimental map, appears near W127. (d) After an additional round of MD aided refinement, yet another new water, W205, was registered in this region. 145
6.14 Individually calculated water maps (colored) overlaid with the total water map (grey) in the O4 channel. All contoured at 2.08 e/Å^3. 146

6.15 Individually calculated water maps (colored) in the region of W50−W51−W52−W53 overlaid with the total water map (grey). (a) All maps contoured at 2.08 e/Å^3, yet only the full water map is visible at this level. (b) Individual maps for W50 (red), W51 (green) and W52 (cyan) contoured at 1 e/Å^3 for better visualization, while the total water map remains contoured at 2.08 e/Å^3 . Dynamic tendencies are observed in this region. Note that no individual map was calculated for W53 as there was no equivalent water near its position after the 45 ns equilibration (see Methods). 147
6.16 Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the O4 channel. 148
6.17 Snapshot at 45ns (third dimer, monomer A, for details see Methods), at which point an H bond chain transiently bridges the gap between the W50−W51−W52−W53 region and the W71−W72−W73 triangle. The ASN sidechain appears also to bend into a conformation which maximizes space for the transiting water. The resting state structure for ASN 338 is overlaid at a higher transparency to demonstrate ASN movement. 149
6.18 Schematic of the ASN drawbridge mechanism observed at 0 ns (a), 0.28 ns (b) and 2.74 ns (c) during the equilibration period (second dimer, monomer A, for details see Methods). In this example, the twisting of the ASN sidechain makes room for a Na+ ion hopping from the W72 position into a region near W51−W52−W53. 150
6.19 Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the O1 channel. 151
6.20 Calculated water maps for individual waters (colored) in the pentacluster (or water wheel), overlaid with the total water map (grey). Maps contoured at 2.08 e/Å^3 .Note that all individual water maps are visible at this level. The grey density next to W4 corresponds to an offset W4 position during the simulation. 152
6.21 Calculated individual water maps for W35−W76−W77 overlaid with the total water map (grey). (a) All maps contoured at 2.08 e/Å^3 , yet only the total water map (grey) is visible at this level. (b) Individual water maps for W35 and W77 contoured at 1 e/Å^3 for better visualization, yet only the map for W35 (red) is visible. Note that no individual map was calculated for W76 as there was no equivalent water near its position after the 45 ns equilibration (see Methods). The density for W77 only becomes visible at levels < 0.55 e/Å^3. 153
6.22 Distribution of the distances between crystallographic water positions and the nearest strong MD peaks (>2.08 e/Å^3) within three channels of interest. Each bin corresponds to raw count divided by the total number of counts in each channel as well as the bin width. As a result, the area under each histogram is normalized to one thus providing a probability density. 154

6.23 Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the Cl1 channel. 155
6.24 Three extra densities observed on both sides of the GLU 65 −GLU 312 'proton gate' in the Cl1 channel. Solvent map contoured at 2.08 e/Å^3 and overlaid with the crystal structure. In the crystal structure, the GLU 65 and GLU 312 distance is 2.7 Å. In the final snapshot from the MD trajectory, the GLU 65 and GLU 312 distance is at least 3 Å across all four copies. We hypothesize that these additional water positions can only be occupied when the bottleneck is wider than in the XRD structure. 156
6.25 Determining optimal O−H bonds. (A) After oxygen and hydrogen peaks are grouped, forming an average water orientation from the simulation, all combinations of O−H peak distances between nearest neighbors were checked in order to converge on the optimal donor acceptor direction for a given pair. From panel A, water Y is clearly in a more optimal configuration for accepting a proton from water X than the other way around. (B) Once acceptor and donor roles were established, a few strict thresholds were set to start filtering the results. OH distances had to be below 3.0 Å, and O_{donor}−$H_{acceptor}$ distance < O_{donor}−$O_{acceptor}$ distance. The latter filter was put in place to ensure that the orientation between donor acceptor pairs was optimized for hydrogen bonding. While both OW_y−$HW1_x$ and OW_y− $HW2_x$ from panel B are feasible bonds, the distance of the former is smaller, so it is selected as the optimal O H bond distance in the pairing. After all optimal bonds are identified between pairs,if a water is registered as participating as a "donor" in three surviving bonds, the weakest of the three is knocked out. (C) The strength of the final bonds were classified based on the distribution of O−H distances retrieved from the MADI analysis (see SI 18). The results are represented by arrows following a color gradient, pointing from donor to acceptor (see Results). 157
6.26 Distribution of O−O and O−H distances retrieved from the MADI analysis. For the O−H distribution the 33rd, 66th and 99th percentile are at 2.4 Å, 2.6 Å, 3.0 Å, respectively. 158
6.27 Flowchart detailing the filtering logic employed for the hydrogen bonding analysis. 159

List of Tables

1.1 Table comparing both SACLA and LCLS facilities, including parameters such as electron energy E_e, Lorentz factor γ, photon energy range $\hbar\omega$, undulator period λ_u, pulse duration $\Delta\tau_{FWHM}$ and available modes. Adapted from [1]. 4

4.1 Time constants reported from other studies in the literature. Note that time constants are used, not half−lifes. Also note that exact sample conditions vary between experiments, which will also have an effect on estimated time constants. 70
4.2 Sample turnover parameters for the included experiments. 71

5.1 Relevant experimental components and their corresponding transmission factors at 5.9 keV - used to correct for the number of photons at the gain medium. The only losses relevant to the Seed Spectrometer come from the X-ray transport losses and KB reflectivity. The remaining elements in the table are used to correct the $K\alpha$ Spectrometer. *Resulting from the ratio of horizontal beam divergence (3.8 mrad) and Si (220) Darwin width (27 μrad). 86
5.2 Distance from nanofocus, horizontal (x) and vertical (y) beamsize at the gain medium, calculated by assuming the 3.8 mrad (horizontal) and 1.9 mrad (vertical) beam divergence downstream of the focus, and the corresponding pump (I_{pump}) and seed (I_{seed}) pulses intensities at the gain medium. 88

6.1 Breakdown of lipids/cofactors/ions used in simulation 160
6.2 Breakdown of atoms in simulation 160
6.3 Documentation of how various lipids/cofactors/ions were prepared for this study. 160
6.4 Net charges for the lipids/cofactors/ions used in this simulation. 161
6.5 All significant water wires (must contain ≥ 3 waters with a consistent acceptor donor orientation) identified by the MADI analysis 161
6.6 Defining O1 channel by amino acids. ROI is 3.5 Å for all waters. 162
6.7 Defining Cl1 channel by amino acids. ROI is 3.5 Å for all waters except for W40, W121, W24 and W61, which have a radius of 5 Å. 163
6.8 Defining O4 channel by amino acids. ROI is 3.5 Å for all waters. 164
6.9 Number of MD peaks and crystallographic waters found in each of the three channels, using the amino acid boundaries tabulated in tables 6.6 - 6.8. 164

Acknowledgments

I feel incredibly lucky to have received so much support leading up to, and throughout, my PhD. If I had any success in graduate school, it is only because I have been surrounded by so many patient, clever and encouraging people. First, I need to thank Drs. Junko Yano, Jan Kern and Vittal Yachandra. They took me on as a graduate student researcher in their group at Lawrence Berkeley National Laboratory back in 2019, and since then, they have provided me with countless scientific opportunities and helped me thrive. Junko - your generosity and willingness to always put others first, your skill for calmly navigating difficult situations, and your ability to dream big while encouraging others to do the same, are just some of the many qualities I admire about you. I am so grateful for all of the new science you exposed me to - and the trust you put in me to handle large data sets. Jan - I hope to one day come close to emulating your ability for quickly and effortlessly switching your brain between various scientific topics. Thank you for introducing me to so many different systems, and for always answering my questions about them. Vittal - thank you for being such a patient teacher over these past years, and for advocating for me as a student. I feel very grateful for all of the concepts you have taken the time to explain to me in terms I can understand, and for all of the resources you provided me with to supplement this learning.

I also would like to thank Professor Roger Falcone, my campus advisor, the chair of my thesis committee and the person who provided me with the opportunity to come to UC Berkeley for graduate school. Roger has always provided me with great advice and support throughout my PhD, and I feel completely indebted to him for the generosity he has shown towards me over the years. I want to also thank the other members of my dissertation committee, Professor Dave Attwood and Professor Graham Fleming. Dave has spent many hours meeting with me over the course of my PhD in order to help me prepare for final projects, prelims, and qualifying exams. His dedication towards encouraging students and helping them succeed is utterly inspiring. I am also incredibly grateful for the support Graham has given me throughout seminars, during my qualifying exam, and also during the preparation of this thesis.

I have been lucky to be involved in the Applied Science and Technology (AS&T) graduate group at UC Berkeley, which is a unique and interdisciplinary PhD program. I would like to thank Professor Ji Yao, who is the chair of this graduate group and was a very supportive member of my prelim committee. Ji also was involved in making my exit talk possible, which I am very grateful for, as it ended up becoming a very special day shared with my family. In addition, I owe a lot to Ariana Castro for her administrative expertise as AS&T student services advisor.

Over the past five years, I have collaborated with many exceptional scientists across several different institutions. I want to first acknowledge Dr. Mike Wall from Los Alamos National Laboratory, who was involved in mentoring me on the molecular dynamics project. Mike has always been sincere, patient, and kind with me. He has allowed me to explore my interests, while also pushing me to think intentionally and with a purpose. I also want to acknowledge Professor Uwe Bergmann at the University of Wisconsin. The first XFEL

beamtime I ever participated in was for Uwe's experiment at SACLA in January 2020. Since then, I have had the opportunity to work closely with Uwe on the stimulated emission project and participate in several more related beamtimes with him. Uwe's excitement towards science, and life in general, is contagious. He has spent countless hours teaching me physics concepts, helping me with writing, and challenging me to think critically. For this, I am very grateful.

During this pre-covid SACLA trip, I met two more people who would also end up having a very significant impact on my PhD experience: Thomas Kroll and Alex Halavanau. Thomas taught me how to vigorously scrutinize data from all sides. Through our weekly meetings, I found myself developing greater confidence with explaining my findings and sharing my interpretations. Thomas has helped me prepare for many important presentations throughout my PhD, selflessly dedicating his time so that I would feel more confident and relaxed when it came to the big day. Alex also spent countless hours helping me parse through data and write code, while also always reminding me of the bigger picture. Throughout my PhD, Alex providing me with a lot of valuable advice, and always made sure to check in on me when schedules got busy. Later on, Alex welcomed me into several of his other experiments related to the X-ray Laser Oscillator at SLAC, and encouraged me when I wanted to obtain more hands-on experimental experience. I would also like to thank Leland Gee, Roberto Alonso Mori, Fred Poitevin and Andy Aquila at SLAC, who I have had the pleasure to work with during beamtimes over the years. In addition, the stimulated emission work shown in this thesis would not have been possible without the scientific team at SACLA which includes Ichiro Inoue, Yuchi Inubishi, Taito Osaka, Jumpei Yamada, Makina Yabashi, Hitoki Yoneda and Yurina Michine.

Outside of SLAC and SACLA, I worked with many other collaborators who also had a profound impact on me and my development during graduate school. Thank you to Johannes Messinger for his involvement in the PS II kinetics work shown in this thesis, David Wych for answering so many of my GROMACS related questions, Rana Hussein and Mohamed Ibrahim for teaching me so much about water networks and PS II, Clemens Weninger for helping me find joy in data reduction, Christopher Ohmer for the laughs and great conversations, and Thomas Fransson for working with me so closely on all things XES and always having the most cheerful and positive attitude in the process.

The fact that I could call Lawrence Berkeley National Laboratory (LBNL) "home" over the past five years is a privilege. Thank you to Aaron Brewster, Nick Sauter and James Holton, who are all scientists at LBNL that I have overlapped with either during beamtimes or when working on the molecular dynamics project. Aaron, Nick and James answered every question I had, created an encouraging and comfortable environment, all while making me feel valued and respected. Thank you to the entire Yano-Yachandra-Kern lab, there are too many to name individually, for teaching me so much about big team science. Thank you to Louise Lassalle for getting me started on the molecular dynamics project, and Asmit Bhowmick for serving as my mentor on this project over the next five years, teaching me about simulations, crystallography, best practices for data handling, scientific thinking, and paper and proposal writing, Hiroki Makita and Phillip Simon for the countless hours

invested teaching me about the tape drive, Kuntal Chatterjee for being so much fun to work together with on the XES and answering all of my spectroscopy related questions, Corey Kaminskey for inspiring me with her vast science knowledge base, Ruchira Chatterjee for her endless support and wisdom, Medhanjali Dasgupta, Isabela Nangca, Miao Zhang, Moritz Kretzschmar and Steve Keable for all of the great memories at beamtimes together. Finally, I want to thank Isabel Bogacz for her friendship, support, and encouragement, and for helping me feel understood and capable during my PhD.

I would not have even made it to graduate school in the first place had it not been for many supportive mentors and teachers along the way. I must thank Francesco Baldini, Stefan Baessler, Ago Marinelli, James Cryan, Dan Slaughter and Mike McDonald for their important roles as research advisors throughout undergrad. I also would like to acknowledge many of my teachers back in Arlington,Virginia. This includes Paul Hessler, Aaron Scheutz, Deb Waldren, Stefanie Dreizen, Curt Dorman, Janet Fenstermacher, Michelle Breedy and Melville Turner. All of these teachers are responsible for my love of school and passion for learning. Being excited to go to class each day is a privilege that I am incredibly fortunate to have had. I owe it to all of them for creating that environment in their respective classrooms.

Finally, I would like to acknowledge my family, friends and loved ones for their unwavering support and encouragement through the years. Specifically, I would like to thank Ian Hardman for his steadfast friendship as we journeyed through high school, college and graduate school together. I'd also like to thank Chris Li, Cole Miles and Andrew Jiao for all of the UVA physics reunions that we embarked on together in the Bay Area. Thank you to Claire Flynt for exchanging nearly-daily voice-notes with me where we have talked about everything under the sun, Courtney Roark for maintaining our cross-country pen-pal streak over the past five years, as well as Simran Batra and Casey Bowler for teaching me the true meaning of friendship.

Finally, I would like to thank Kirk Larsen for his endless support, for believing in me when I doubted myself, for always putting things into perspective, for sharing my excitement over plots, and for bringing so much color into my life. And to my family, thank you so much for all you have done for me. Matthew, thank you for always giving me the best advice. Brendan, thank you for organizing regular sibling races to keep my motivated throughout graduate school. Mom and Dad, thank you for answering every one of my phone calls, constantly affirming me (while also being frank when I need it), and for giving me the confidence to aim high and follow my heart.

Chapter 1

Introduction

1.1 Background and Purpose

1.1.1 Timescales and Challenges for Monitoring Dynamic Processes in Biological Systems

Thoroughly understanding a biological system requires study of structural dynamics across multiple timescales. The temporal scale of dynamics in biological systems is vast, spanning over 15 orders of magnitude (Fig. 1.1). Absorption and stimulated emission occur at the shortest timescales (approximately femtoseconds), while protein synthesis and cellular processes take place over timescales extending beyond one second. Therefore, it is important to identify a tool or suite of techniques that provide analytical access to this broad range of processes.

Hard X-ray based techniques can be quite powerful, as photons with such high energy (from just below 10 keV up to nearly 100 keV [1]) have Å wavelengths that are comparable to atomic level spacing. When measuring a sample under crystalline conditions, X-ray light can be diffracted, and the captured coherent scattering provides atomic-level structural information. Additionally, many atomic resonances and absorption edges correspond to energies in this regime. Thus, hard X-ray spectroscopy allows scientists to probe the electronic structure of the system under study [2].

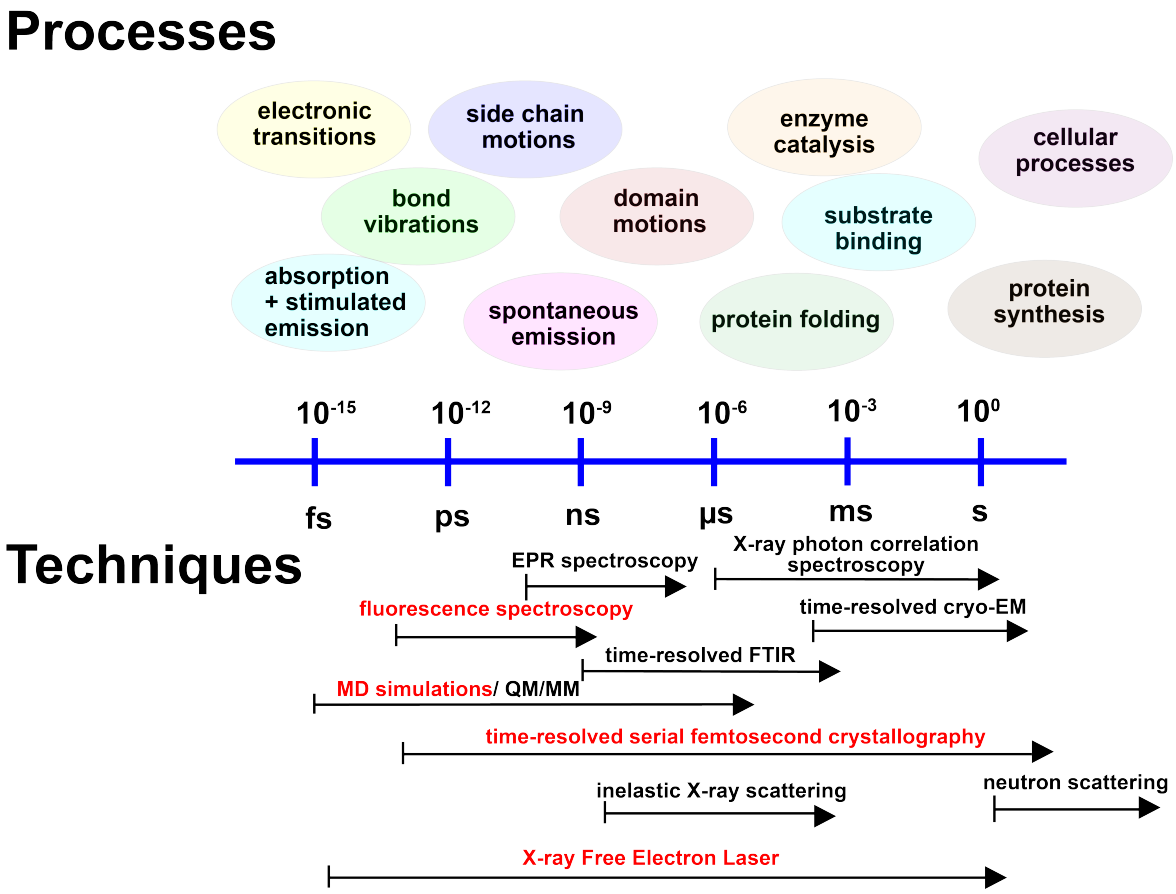


Figure 1.1: The time scale of various biological processes (top) and the temporal resolution of major techniques deployed in biophysical studies (bottom). The tools explicitly discussed in this dissertation are colored red. Adapted from [3, 4, 5, 6].

While X-rays are a powerful tool for imaging and spectroscopic studies of complex biological systems, they have significant drawbacks. As a form of ionizing radiation, X-rays can damage samples, which is a major bottleneck for biological studies. This damage is initiated through absorption of the radiation (see **Section 1.3**), which leads to the generation of highly reactive species called free radicals which are characterized by an unpaired electron [7]. As these free radicals propagate through the sample, their strong oxidizing power removes electrons from other molecules, resulting in further damage in the process. This is one of the main sources of damage at a synchrotron [8].

The damage can be slowed down by bringing the sample to a very low temperature ($\approx$ 100 K) [9]. While this technique is effective at prolonging the lifetime of the sample, it also comes with complications. The handling and mounting required for cryo-cooled samples can lead to additional damage, and some samples even exhibit high levels of cryo-resistance [10]. If one is attempting to carry out time-resolved studies, knowledge of the freezing time

following some excitation is not always going to be precise. Finally, extensive studies have demonstrated how cryocooling can modify the conformation of the sample. Proteins are known to exhibit a large variety of conformations at room temperature that are important for their function. This raises questions about how accurately the collected data reflects the sample's natural state [11, 12, 13].

1.1.2 Harnessing the Power of the X-ray Free Electron Laser

X-ray Free Electron Lasers (XFELs) are an extraordinarily powerful and relatively recent technology that has come online in the past 15 years. They can circumvent many of the challenges discussed in the previous section, partly due to the ultrashort nature of the X-ray pulses generated by XFELs, which have durations on the order of femtoseconds. This allows researchers to bypass cryocooling entirely and effectively outrun radiation damage while collecting data at room temperature. The ultra−bright nature of the pulses enables the collection of high−resolution images, allowing even weakly scattering systems to be studied and subtle changes in pump−probe studies to be detected. Since the Linac Coherent Light Source (LCLS) began operating near Stanford in 2009 (the world's first hard X-ray Free Electron Laser), atomic resolution imaging of molecular processes at room temperature has been achieved, and pump-probe studies have been conducted across timescales ranging from femtoseconds to seconds [14].

The work in this dissertation leverages modern XFELs (see **Chapter 2** for an introduction to XFELs), including their multicolor and time delay capabilities, to enhance our understanding of Photosystem II (PS II). This technology yields massive amounts of data, necessitating the development of specialized analytical tools (see **Chapter 3** for more on data-related challenges). Specifically, the work described here was conducted at two XFEL facilities. The first is LCLS in California, where the X-ray emission data presented in **Chapter 4** was collected, along with the serial femtosecond crystallography data in **Chapter 6**, which is compared to Molecular Dynamics (MD) simulations. The second XFEL used in this dissertation is at SACLA in Japan, where the stimulated emission work shown in **Chapter 5** was performed.

Each of these facilities offers unique benefits. For example, both LCLS and SACLA have two-color modes; however, SACLA offers better beam stability and allows for more intense focusing conditions, which is crucial for the work shown in **Chapter 5**. See Table 1.1 for more quantitative details comparing the two facilities.

Parameters	LCLS	SACLA
E_e	13.6 GeV	8 GeV
γ	26,600	15,700
$\hbar\omega$	250 eV - 10 keV	4.5 - 15 keV
λ_u	30 mm	18 mm
$\Delta\tau_{FWHM}$	$<$ 10 - 70 fs	$<$ 10 fs
Mode	SASE and self-seeded	SASE and self-seeded

Table 1.1: Table comparing both SACLA and LCLS facilities, including parameters such as electron energy E_e, Lorentz factor γ, photon energy range $\hbar\omega$, undulator period λ_u, pulse duration $\Delta\tau_{FWHM}$ and available modes. Adapted from [1].

1.2 Metalloenzymes

1.2.1 Importance in Biological Processes and Significance of Manganese

The chemical processes carried out in biological systems is enabled by catalysts called enzymes, typically proteins that speed up the rate of reactions inside cells by lowering activation barriers [15]. Enzymes that require a carefully selected set of firmly bound metal ions for their activity, metalloenzymes, are involved in some of nature's most complex reactions. The flexibility of the oxidation states of transition metals makes them ideal participants in these processes, which favor chemical diversity through evolution, allowing for the active site to be fine-tuned [16]. To improve the design of our own artificial enzymes, it is crucial to study nature and observe these processes in real time. Time-resolved studies, especially those that provide insight into structural and electronic dynamics under functional conditions at XFELs, are integral to this understanding.

Manganese is a key transition metal of interest, as it is involved in hundreds of metalloenzymes in plants and is highly valued for its oxidation state flexibility [17]. Manganese exists in oxidation states ranging from +2 to +7, with the +2 state being the most prevalent in biological systems [17]. Manganese was first documented in plants over 200 years ago from the ash of several vegetable species [18]. In the 1900s, manganese was recognized as playing an important role in both plant growth and development [19]. Sixty years later, a catalytically active Mn-containing protein complex was isolated from spinach [20].

One of the most extensively researched manganese-containing metalloenzymes is PS II (see **Section 1.2.2**), which is integral to photosynthesis—the process by which plants use the sun's energy to produce simple carbohydrates from carbon dioxide and water, releasing oxygen. PS II contains an active metal site composed of four manganese called the oxygen evolving complex (OEC).

1.2.2 Photosystem II

Photosynthesis, occurring in plants, algae, and cyanobacteria, produces the majority of the molecular oxygen that sustains life on earth. The overall chemical reaction is written as:

$$6CO_2 + 12H_2O + \text{light} \rightarrow C_6H_2O_6 + 6O_2 + 6H_2O \tag{1.1}$$

During this process, carbon dioxide and water are taken in by a plant and used to make glucose and oxygen gas. In reality, this can be separated into two primary components: the light reactions and the dark reactions (the latter is also called the Calvin cycle). Plant cells contain chloroplasts that are responsible for photosynthesis. Inside the chloroplast lies a structured membrane (the thylakoid membrane) which is where the light reactions take place.

The thylakoid membrane contains photosynthetic complexes PS II, cytochrome b_6f (Cyt b_6f) and Photosystem I (PS I) [21] (see Fig. 1.2). For the reaction to kick off, light first must be absorbed by one of the many chlorophyll molecules contained within the thylakoid membrane. This initiates an energy transfer process between molecules that eventually culminates at a special pair of chlorophylls in PS II's reaction center, P680. After absorbing this energy, P680 transitions to an excited state, releasing an electron which is passed through the electron transfer chain [22] (red arrows, Fig.1.2). The electron initially transfers between PS II to Cyt b_6f through a charge carrier, plastoquinone (PQ) [23], which gets reduced to plastoquinol (PQH_2). Next, electron transfer between Cyt b_6f and the reaction center of PS I, P700, is facilitated by another charge carrier, Plastocyanin [24].

When light energy is absorbed by P700, the associated chlorophyll becomes excited. The energy from this excitation facilitates the conversion of $NADP^+$ and a hydrogen ion into NADPH which is a crucial molecule in the second phase of reactions (Calvin cycle). In addition to this, ATP, which is synthesized from ADP and P_i (inorganic phosphate) by ATP Synthase (not shown in Fig. 1.2) is also required for facilitating the assembly of sugar molecules in the second set of reactions [25].

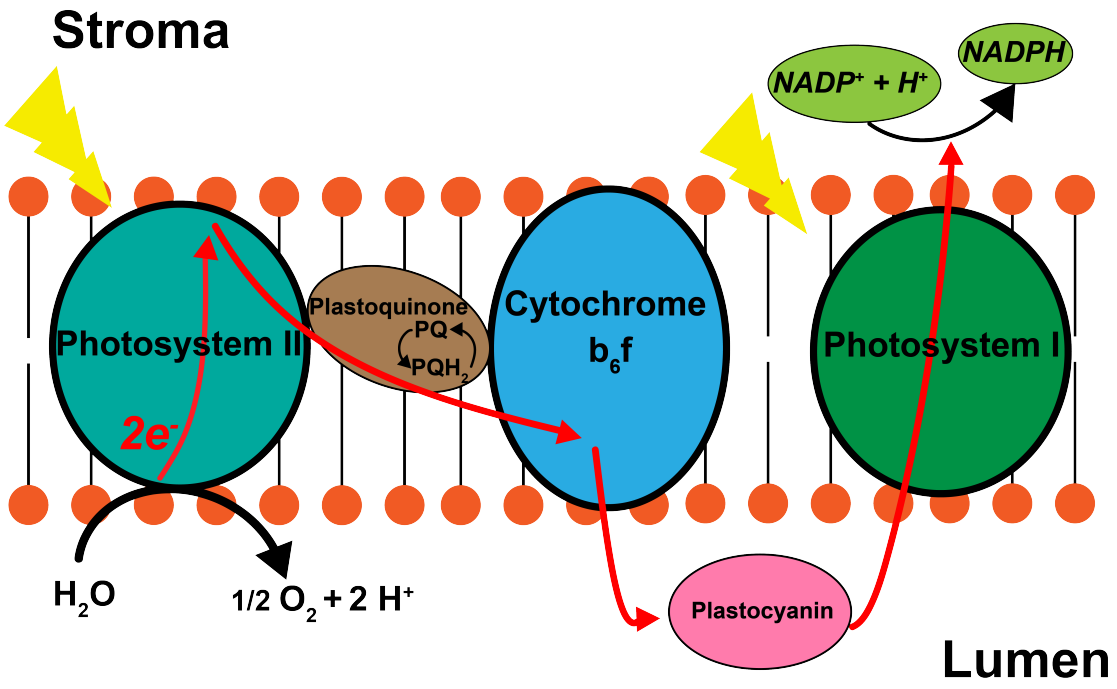


Figure 1.2: Simplified schematic of the thylakoid membrane and the protein complexes involved in the photosynthetic electron transport chain. Electrons follow the indicated path (red arrows) to the stromal side of PS I where they are used to reduce $NADP^+$ to NADPH. Adapted from [26, 27].

This dissertation, however, is primarily concerned with the first stage of photosynthesis involving the light dependent reactions. Let us now turn our attention back to PS II, specifically the OEC, a metal cluster (Mn_4CaO_5) situated within the protein towards the lumen side of the membrane. As previously mentioned, $P680^+$ is in its excited state, and is lurking nearby with an outstanding redox-potential estimated to be $+1.21$ V [28]. It is then able to extract an electron from a Tyrosine residue, Tyr161 [29]. In turn, Tyr161 obtains electrons from the cluster.

Sequential excitation of P680 through subsequent light flashes drives the redox state changes of the this cluster, cycling through a series of five different "S states" as they are commonly called [30]. In this series, S_0 is the most reduced state and S_4 is the most oxidized state. Starting from the dark S_1 state, each photon absorbed leads to oxidation of the cluster and further advancement along the cycle until the transient S_4 state is reached. At that point, molecular oxygen is formed from the oxidation or "splitting" of water, and the cluster is reset for the process to begin again.

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \tag{1.2}$$

In addition to the redox changes which occur throughout this cycle, there are also a number of structural changes which occur at and around the cluster as well. Better understanding the details of this reaction may provide a blueprint for practical applications, such as the splitting of water for solar energy production.

1.3 Overview of Techniques Used

The work in this dissertation employs various techniques centered around the use of the XFEL for the purpose of studying PS II. I will first describe how incident X-rays can interact with a sample, and then I will place this in the context of XFEL-related work. See Fig. 1.3 for a schematic demonstrating the various possibilities in which X-rays can interact with a sample.

In general, X-ray beams induce sample heating. If the temperature of a crystalline sample rises above the glass transition point, damage can get accelerated which will shorten the crystal's lifetime [31]. When collecting data at room temperature, beam−induced heating also can increase reaction rates [31]. This makes it imperative to outrun these heat-induced effects when collecting at high pulse intensities.

This heating is initiated by photoabsorption. When an atom absorbs an X-ray photon, a photoelectron from the inner shell is ejected, leaving behind a highly excited state. From there, the excited state can decay either through the Auger-Meitner effect (non-radiative), in which an inner shell vacancy is filled while an electron from the outer shell is emitted [32], or an X-ray fluorescence photon can be emitted (radiative). Typically, fluorescence competes with Auger-Meitner emission, and the selected process is related to the atomic number (Z) of the element. For high Z elements, such as heavy metals, X-ray fluorescence is favored. Conversely, for low Z elements, Auger-Meitner emission is favored.

The vacancy in an ion which just emitted an X-ray fluorescence photon is susceptible to a cascade of Auger-Meitner decay, generating more electrons in the process which have the potential to further ionize and damage their surroundings.

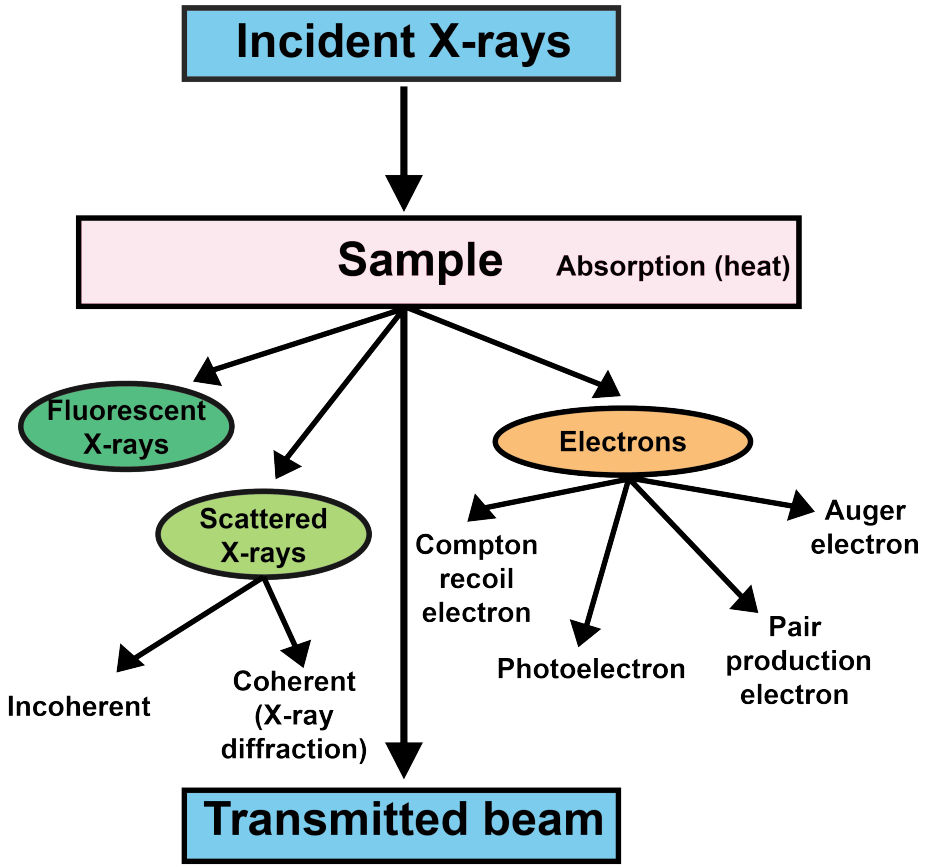


Figure 1.3: Schematic of possible X-ray interactions with matter. Adapted from Andrew Minor's Materials Characterization course (MATSCI 204) at UC Berkeley.

However, the ultra-short X-ray pulses provided by XFELs (femtoseconds) can effectively outrun most damage processes[33]. Furthermore, since the pulses are widely spaced in time (8 ms when running at 120 Hz at LCLS), there is adequate time for sample to replenished when running in a "probe before destroy" mode. This strategy has been an effective measurement approach at XFELs in research areas spanning serial femtosecond crystallography (SFX) to ultrafast X-ray spectroscopy [34, 35, 36, 37, 38].

When considering other ways that X-rays can interact with matter, pair production is a possibility, but it only occurs at energies much higher than those relevant to this dissertation ($\geq$ 1 MeV) [39]. Incoherently scattered X-rays, such as those from Compton scattering, can indirectly contribute to the background in our experiments. They affect the energy of incident X-rays, further decreasing the intensity of the desired signal while slightly increasing noise [39]. Coherently scattered X-rays provide us with X-ray diffraction data, enabling the capture of structures of large proteins such as PS II. While not described in great detail in this section, it is highly relevant for the MD simulations covered in **Chapter 6**.

1.3.1 X-ray Emission Spectroscopy

In X-ray emission spectroscopy (XES), an inner shell electron is first ionized from an atom, leaving behind a photo-ionized core-hole intermediate state. The ion can undergo relaxation through the emission of either a fluorescence photon or an electron. This section focuses on the former possibility.

The energy and labeling of the fluorescence photon depends on two factors: (1) the shell from which the electron was removed (in this dissertation, the focus is on processes involving the removal of an electron from the 1s orbital, which is called K fluorescence) and (2) the orbital from which the filling electron originated [40].

Fig. 1.4a illustrates the two transitions that are particularly pertinent to this dissertation. Transitions between the 2p and 1s orbitals lead to Kα emission, while transitions between the 3p and 1s orbitals lead to Kβ emission. Fig. 1.4b features example spectra for each collected from manganese foil at LCLS. For further insights into XES data collection at XFELs, refer to **Chapter 4**.

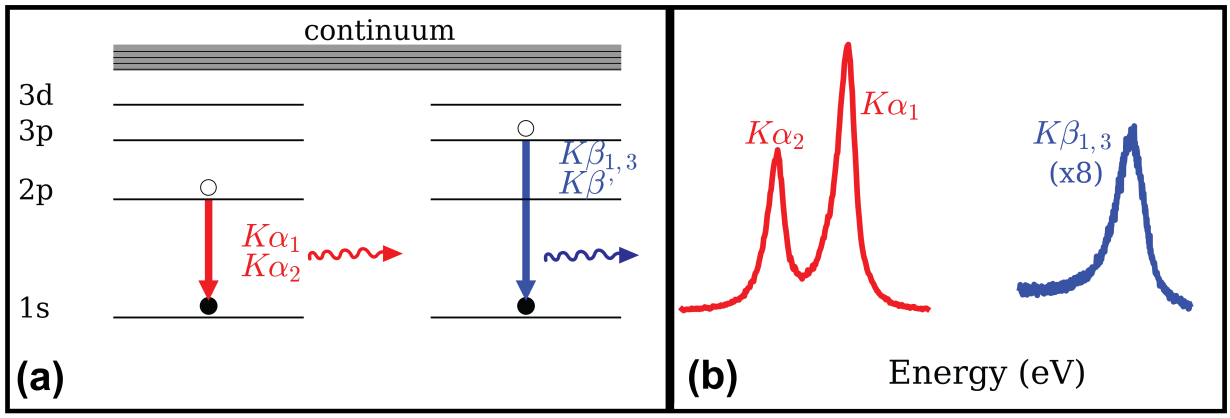


Figure 1.4: A simplified version of the respective transitions which result in Kα and Kβ emission (left) and the resulting spectra collected from Mn foil (right), adapted from [41].

The spectra will change depending on the chemical environment, making it a useful technique for probing chemical properties of 3d transition metals such as manganese [40]. The Kα lines, which come into play primarily in **Chapter 5**, are affected by the spin-orbit influence, a result of the electron's spin interacting with its own momentum [40]. The Kβ lines, which come into play primarily in **Chapter 4**, are mostly influenced by electron-electron interactions between the valence shell and the 3p orbitals, leading to two separate features (K$\beta_{1,3}$ and Kβ').

There is a higher energy region not shown in Fig. 1.4, arising from transitions between the valence orbitals and the 1s shell. These so-called "valence-to-core" or VtC features are the most sensitive to the element's bonding environment. However, VtC spectra are several hundred times weaker than Kβ emission, posing significant data collection challenges. On

the other hand, Kα spectra are approximately eight times stronger than K$\beta_{1,3}$, but are not as sensitive to the valence shell configuration. When studying PS II, K$\beta_{1,3}$ spectra are typically collected as this is a reasonable compromise between sensitivity and signal strength.

1.3.2 Stimulated X-ray Emission Spectroscopy

While the previous section primarily concerned the linear photon-matter interactions involved with spontaneous emission (Fig. 1.5 a, middle), I will now discuss the development of a spectroscopic technique that considers nonlinear photon-matter interactions (Fig. 1.5 a, bottom). This technique involves inner-shell X-ray lasing, which can occur either through Amplified Spontaneous Emission (ASE) or Seeded Stimulated X-ray Emission (SSE).

First, a short ($\sim$10 fs) and highly focused ($\sim$100-200nm FWHM) pulse tuned above the absorption edge of the element of interest creates a population inversion of 1s core-hole excited states (Fig. 1.5 b). In the X-ray regime, this can be achieved using the high peak power of an XFEL as a pump. After a short time, some photons are spontaneously emitted in the forward direction where they stimulate the emission of additional photons at the same wavelength and in the same direction. Thus, this technique allows for exponential gain of the signal in a single pass (Fig. 1.5 c).

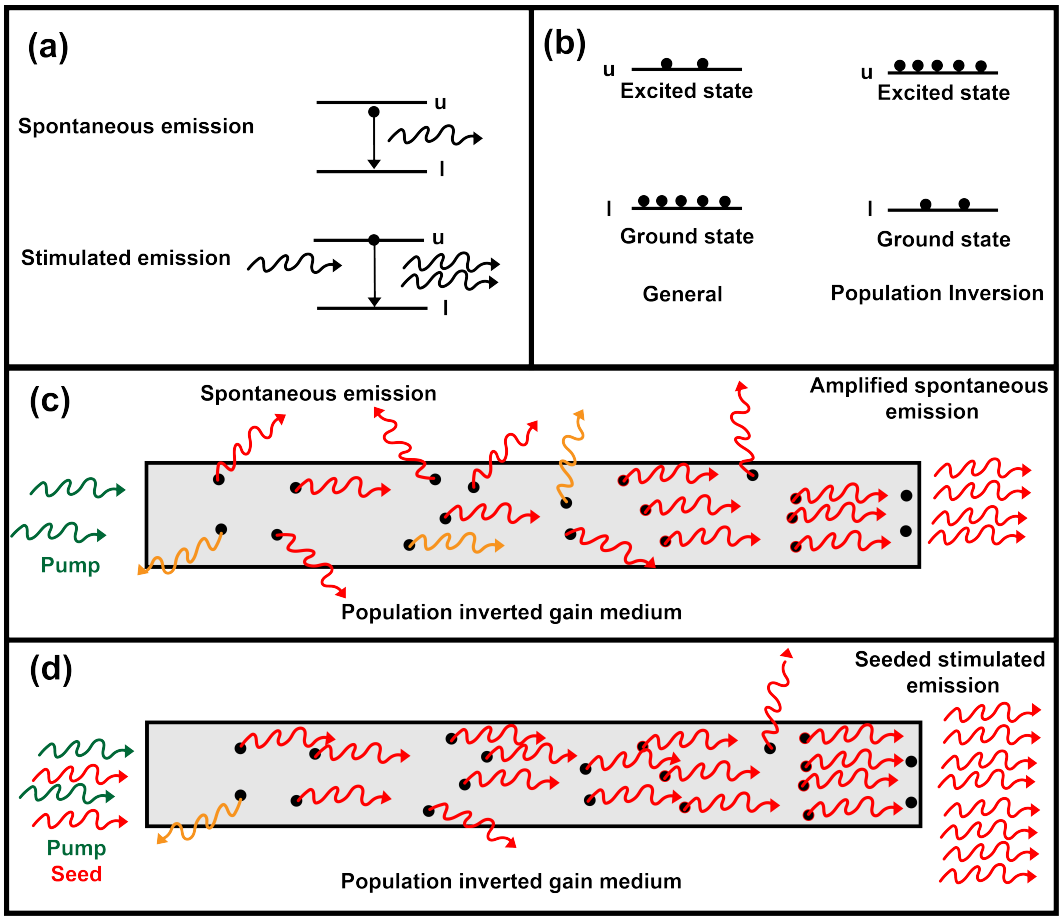


Figure 1.5: Schematic of relevant processes involved in stimulated X-ray emission, adapted from [1] and [42]. Illustrated differences spontaneous emission and stimulated emission (a). Difference between ground state and excited state (b). In ASE, a spontaneously emitted photon (red) acts as seed to stimulate the emission of a second photon from a core excited atom (c). In SSE, an external seed pulse (red) stimulates the emission of photons from core–hole excited atoms along the seeding direction (d).

One can also introduce a second color pulse, tuned precisely to the energy of the transition of interest, for selection and amplification (Fig. 1.5 d). These two colors can be generated at an XFEL using two sets of undulators (see **Chapter 2** and **Chapter 5** for more details). The seed pulse is then timed to arrive within the core-hole lifetime of the excited state, at which point it can stimulate the excited state to decay through a specific path, resulting in enhanced chemical sensitivity.

1.3.3 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations are a powerful tool that, when combined with experimental data, can provide valuable insights into the structure and function of large biomolecules and reveal details about the dynamic processes inherent to protein function [43, 6]. In this dissertation, MD simulations on PS II are carried out (**Chapter 6**) with the specific intention of investigating the water and hydrogen bonding networks surrounding the active site with a level of spatial and temporal detail that surpasses what is currently achievable experimentally.

Most MD simulations follow a similar procedure. A simplified MD workflow, which involves preparing, running, and analyzing the simulation, is shown in Fig.1.6. To begin a simulation, an initial structure is required. These structures are typically collected experimentally at a synchrotron but can also be obtained computationally, or from a combination of the two. For this dissertation, however, the starting structure used was collected under room temperature conditions at an XFEL.

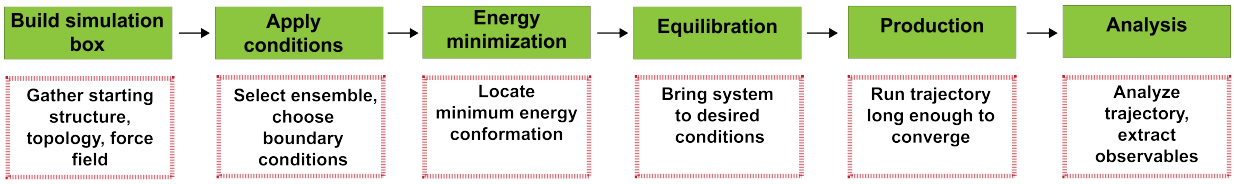


Figure 1.6: Simplified MD simulation workflow, adapted from [6].

A topology file is also necessary. This type of file identifies the atoms used and their interactions (both bonded and non-bonded) with each other. A force field also needs to be chosen. This is a mathematical expression that describes the dependence of the system's potential energy V on the coordinates of the particles [44]. The force F on a given atom can be related to the potential by the following:

$$F = -\nabla V \tag{1.3}$$

Where the total potential of the system V_{tot} includes all bonded potentials as well as non-bonded terms:

$$V_{tot} = V_{bonded} + V_{non-bonded} \tag{1.4}$$

V_{bonded} is summed over the potential energy associated with chemical bonds, bond angles, and angles between planes [6].

$$V_{bonded} = V_{bond} + V_{angle} + V_{dihedral} + V_{improper\,dihedral} \tag{1.5}$$

where

$$V_{bond}(d_{ij}) = \frac{1}{2}K_b(d_{ij} - d_b)^2 \tag{1.6}$$

$$V_{angle}(\nu_{ijk}) = \frac{1}{2}K_a(cos(\nu_{ijk}) - cos\nu_a)^2 \tag{1.7}$$

$$V_{dihedral}(\phi_{ijkl}) = \frac{1}{2}K_d(1 + cos(n\phi_{ijkl} - \phi_d))^2 \tag{1.8}$$

$$V_{improper\ dihedral}(\varphi_{ijkl}) = \frac{1}{2}K_{id}(\varphi_{ijkl} - \varphi_{id})^2 \tag{1.9}$$

The force constants K_b, K_a, K_d and K_{id} in Eqs. 1.6 - 1.9 represent the stiffness of the molecule. They, along with the equilibrium constants d_b, ν_a, ϕ_d and φ_{id}, are derived, validated, and undergo further rounds of refinement in order to reproduce experimental observables [6].

Meanwhile, $V_{non-bonded}$ describes the interaction between pairs of particles and is composed of both the van der Waals force which is modeled as the Lennard-Jones potential,

$$V(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] \tag{1.10}$$

and the Coulomb interactions, which are modeled as such:

$$V_{coulomb} = \frac{q_i q_j}{4pi\epsilon_0\epsilon_r r} \tag{1.11}$$

The choice of force field depends on several factors, including, but not limited to, the system being studied, the force field's compatibility with the simulation software and the chosen water model, and the goals of the simulation. A number of choices must then be made to apply the correct conditions for the simulation. This includes proper selection of the ensemble, which will depend on the experimental conditions that one wishes to simulate (see **Chapter 6** for more details). Selecting periodic boundary conditions (PBC) is a common method used to simulate the bulk behavior of a system of interest [45]. While PBC can simplify the complications associated with treating interactions at the boundary of the simulation box, it can also introduce unwanted artifacts, which are beyond the current scope of this dissertation [45].

The starting configuration is first run through a round of energy minimization to check for inappropriate geometry or steric clashes resulting from the overlap of non-bonding atoms, thus helping optimize the molecular structure [46]. An equilibration phase follows. For the simulations presented in **Chapter 6**, the first 45 ns of the trajectory were used to ensure proper equilibration of the protein, lipids, cofactors, and solvent, and thus were not used for further downstream processing. The trajectory from the final 55 ns of the simulation was extracted for statistical analysis and is considered the production part of the simulation. These trajectories can then be analyzed to extract useful observables for

comparison with experimental data. The aforementioned atomic trajectories are generated by numerical integrating of Newton's equations of motion:

$$F = ma = m\frac{dv}{dt} = m\frac{d^2x}{dt^2} \tag{1.12}$$

This can be achieved through an integrator, the choice of which must consider both accuracy as well as computational cost. The MD integrator that is utilized in **Chapter 6** is the so-called "leap-frog algorithm." Through this integrator, the positions and velocities can be approximated by a Taylor series expansion. The velocities at time $t + \frac{\Delta t}{2}$ are computed from the force:

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{\Delta t}{m}F(t) \tag{1.13}$$

Which is then used to compute the positions x at time t + Δt:

$$x(t + \frac{\Delta t}{2}) = x(t) + v(t + \frac{\Delta t}{2})\Delta t \tag{1.14}$$

So that the velocities and positions leap over each other, meaning that the computed positions and velocities are always shifted one half step from each other. Leapfrog methods are time reversible and are guaranteed to conserve total energy. Further details surrounding the choice of integrator is beyond the current scope of this dissertation. However, the interested reader can learn more from a range of classic MD textbooks, such as those by Daan Frenkel and Berend Smit [47].

Chapter 2

X-ray Free Electron Lasers

2.1 Origin of XFELs

On April 24, 1947, a technician at General Electric's 70-MeV electron synchrotron in Schenectady, New York observed a bright arc of light emanating from a partially transparent vacuum chamber where electrons were being accelerated. This marked the first observation of synchrotron radiation [48, 49]. In the years that followed, scientists began to capture and utilize this previously untapped form of radiation from electron storage rings initially designed for high-energy physics applications. Beam ports were deployed in a strategy known as "parasitic operation" [50, 1].

The demand for more tailored delivery of synchrotron radiation, covering the higher energy end of the electromagnetic spectrum, quickly increased. Radiation with wavelengths down to the Angstrom level offered fruitful opportunities for scientists pursuing absorption measurements in metals and solid-state physics studies. In 1981, more than 30 years after the first observation of synchrotron radiation, the world witnessed the inauguration of the first dedicated X-ray synchrotron radiation facility at Daresbury Laboratory in England [51, 50]. Japan and the United States soon followed suit, with the Photon Factory in Tsukuba and the National Synchrotron Light Source (NSLS) at Brookhaven National Lab in New York both becoming operational by 1984. The development of these dedicated facilities, unlike those where synchrotron light experiments were performed parasitically, marks the transition from first-generation to second-generation light sources.

Soon, the desire for increased spectral brightness intensified among the expanding and diverse user community. Brightness, a key measure of light source quality, is correlated with the ability to study systems at the atomic level and achieve high yields in spectroscopic experiments. As user interests broadened, covering fields from biology to disordered materials, the demand for better light sources became more pressing.

To meet these growing needs, facilities developed and optimized "insertion devices", such as undulators and wigglers (see **Section 2.2**). Examples of third-generation facilities include European Synchrotron Radiation Facility (ESRF) in Grenoble, France, the Advanced Photon

Source (APS) in Illinois and Super Photon ring-8 GeV (SPring8) in Japan[50].

In the same decade, Claudio Pellegrini proposed constructing an X-ray Free Electron Laser (XFEL) at SLAC National Laboratory [52]. This came 20 years after John M.J. Madey's seminal paper introduced the idea of using a Free Electron Laser to generate fully coherent X-ray radiation [53]. Such a scheme would eventually be classified as a fourth-generation light source, alongside diffraction-limited storage rings. The early 21st century saw the realization of this vision, marking the dawn of the XFEL. These facilities provide scientists with peak brightness that surpasses other sources by ten orders of magnitude (see Fig. 2.1) [54], delivering fully coherent, femtosecond-duration pulses. This enables the study of ultra-fast and ultra-small phenomena across a wide range of disciplines.

Despite its name, an X-ray Free Electron Laser differs significantly from a conventional optical laser. In a conventional laser, lasing occurs when a gain medium is pumped into an excited state, allowing it to undergo stimulated emission. For example, a solid-state laser may use a titanium-doped sapphire (Ti:Sapph) or ytterbium-doped yttrium aluminum garnet (Yb:YAG) gain medium. A gas laser might use a CO2 gain medium, and a fiber laser could have a thulium-based gain medium. In contrast, an XFEL uses free or unbound electrons as the gain medium. These electrons are emitted from an electron gun and accelerated to relativistic speeds, enabling the generation of X-ray radiation.

Furthermore, since an XFEL operates in a high-gain regime, an optical cavity is not needed, and full radiative power can be achieved in just a single pass, through a long hall of undulators, see **Section 2.3**). This necessitates the use of large-scale facilities because an optical cavity is impractical; the mirrors would be significantly damaged by the high-energy photons. An XFEL oscillator, currently under development (described more in **Chapter 5**), would use reflective X-ray crystals at reflectivity near 100 %, thereby significantly reducing potential damage.

The impact of the XFEL across scientific disciplines has been profound, with facilities in five different countries becoming operational on the hard X-ray regime within the past twenty years (Fig. 2.1). In structural biology, the barrier to accessing atomic-scale details has significantly lowered with the utilization of hard X-rays from XFELs [14]. Notably, XFELs have played a crucial role in SARS-CoV-2 research, highlighting their versatility and vital contributions to understanding complex systems [55]. Chemistry has also experienced progress by leveraging the soft X-rays generated from XFELs for studies related to electron dynamics [56]. Additionally, the high energy density community has using XFELs to study solid matter under extreme temperature and pressures [57].

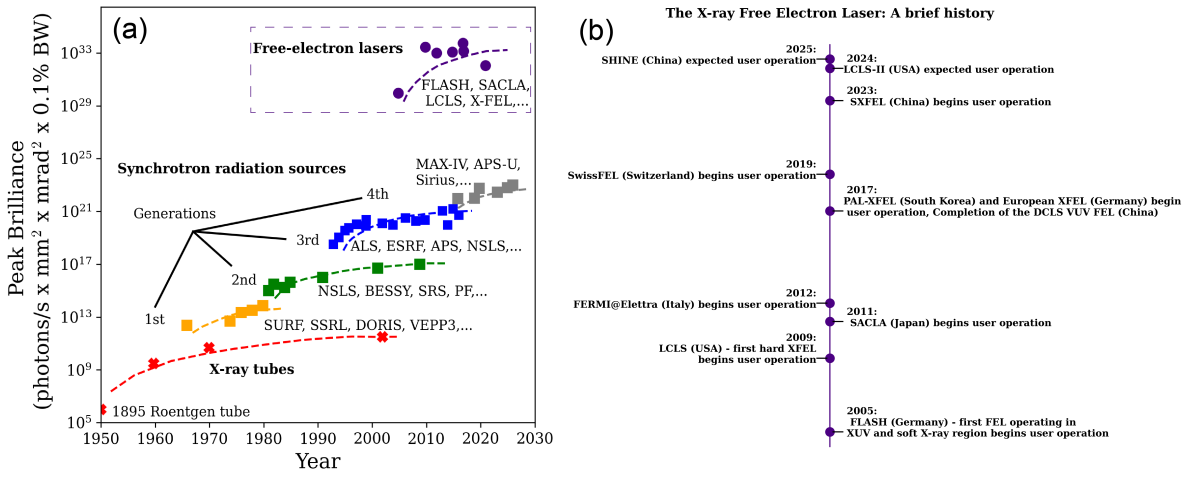


Figure 2.1: An historical overview of X-ray radiation sources. (a) Evolution of peak brilliance across generations of synchrotron radiation sources. (b) A concise timeline of X-ray Free Electron Laser history. Adapted from [58] and [59].

In the subsequent sections, a concise overview of XFELs will be presented, starting with a discussion of synchrotron radiation (**Section 2.2**). This will be followed by a detailed description of XFEL physics in more detail, including an outline of the SASE process (**Section 2.3**) and an examination of seeding techniques (see **Section 2.4**, which will be particularly relevant for the work presented in **Chapter 5**). This chapter will conclude with a comparison between XFEL and synchrotron time structure (**Section 2.5**), as well was a perspective on the future of XFEL science (**Section 2.6**). Most of the content follows Ref [1].

2.2 Synchrotron Radiation

When a charged particle undergoes acceleration - speeding up, slowing down, or changing direction - it emits radiation. Synchrotrons utilize this phenomena by altering the trajectory of relativistic electrons using bending magnets and insertion devices (undulator or wigglers) (Fig. 2.2), thus producing synchrotron radiation [1]. These three types of magnetic structures are described as follows:

1. Bending magnets: These induce a single curved trajectory, resulting in a fan of radiation. They generate a broad spectrum as the angular deviation of the electron is large.

2. Wigglers: In these, a periodic magnetic structure causes electrons to oscillate. The strength of the magnetic field results in large oscillation amplitude and high radiated power. While the radiation spectrum is akin to that of a bending magnet, a wiggler

produces a broader radiation cone in both space and angle [1]. However, it yields less peak brightness over a larger area.

3. Undulators: These are similar to wigglers but with a weaker magnetic field, which confines the angular excursions of the electrons, resulting in a smaller deflection angle. This leads to the coherent addition of emitted radiation and a narrow spectral width. This contrasts with wigglers, which have a much larger deflection angle, resulting in a broader spectrum [60].

Each of the aforementioned structures has its own set of advantages and drawbacks. For instance, a bending magnet generates excellent photon flux and is cost-effective. However, it has limited brightness compared to undulators and wigglers. While a wiggler has higher flux, it is more expensive than a bending magnet and necessitates cooled optics. It also provides limited access. Undulators achieves brighter radiation, smaller spot size, and partial coherence due to their shorter period and weaker magnetic field compared to wigglers. However, similar to wigglers, undulators entail higher costs and provides restricted access. They also require cooled optics as the radiated power is constrained within a narrow beam.

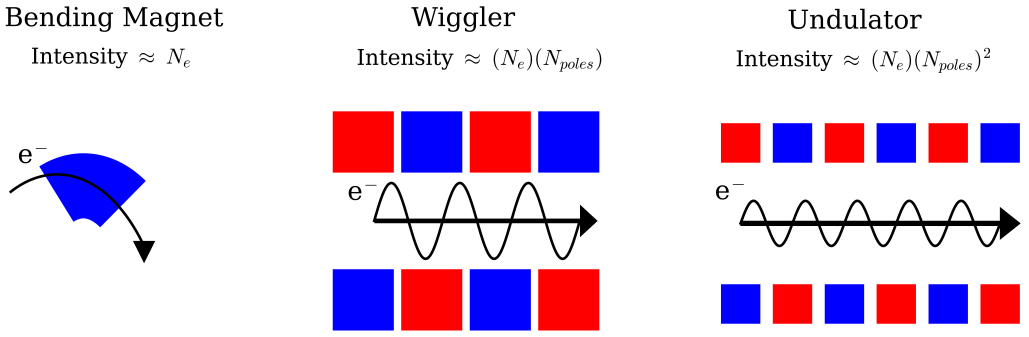


Figure 2.2: Diagram illustrating the three types of magnetic structures generating synchrotron radiation, along with how their respective intensity scales with number of electrons N_e and number of poles N_{poles}, adapted from [1].

A key parameter for the performance of synchrotron facilities is K, also known as the "undulator parameter". This dimensionless parameter characterizes the strength of the undulator and is given by:

$$K = \frac{eB_0\lambda_u}{2\pi m_e c} \tag{2.1}$$

where e is the electric charge, λ_u the undulator period, B_0 is the on-axis magnetic field strength, m_e is the electron mass, and c is the speed of light. For K $\gg$ 1 (for example, with wiggler radiation), a bright and broad spectrum of harmonics is produced, reaching higher

photon energies. These harmonics eventually merge into a continuum. When K $<$ 1 (e.g., with undulator radiation), the resulting spectrum is characterized by narrow spectral lines, high spectral brightness and partial coherence.

To mathematically describe the generation of short wavelength radiation by relativistic electrons, it is convenient to consider the radiation emitted in multiple reference frames. Fig. 2.3 illustrates in three parts how, as an electron traverses through an undulator (a) the radiation pattern of the oscillating electron in the frame of reference moving with the electron appears toroidal, with no shape and there is no radiation in the direction of acceleration. Consequently, so that the angle between the acceleration direction and the propagation direction in the electron's frame of reference, $\Phi^{'}$, is 0 (b). Moving into the laboratory frame (c), however, results in the radiation being contracted through relativistic effects into a narrow radiation cone, with the following angle:

$$\theta = \frac{1}{2\gamma} \tag{2.2}$$

where γ is the Lorentz contraction factor. For more mathematical details of this process, please refer to [1].

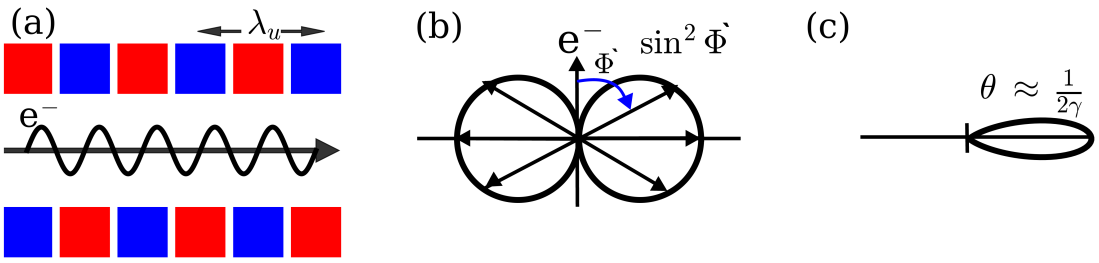


Figure 2.3: Generation of narrow cone undulator radiation. As electrons traverse the magnetic structure (a) they oscillate and generate a toroidal $sin^2\Phi^{'}$ radiation pattern in the electron's frame of reference (b). In the laboratory frame of reference, the shape is contracted and folded into a narrow forward cone with an angle of $\theta = \frac{1}{2\gamma}$ (c). Adapted from [1].

One benefit of the free-electron gain medium employed in XFEL operation is that users are not limited by specific atomic transitions. Instead, the radiation wavelengths are tunable by adjusting K. In certain cases, even two different colors can be generated by utilizing two sections of undulators (**Chapter 5**) [61]. Typically the undulator period λ_u is fixed, so to adjust K, one must manipulate the magnetic field strength B_0 by varying the undulator gap length. A larger gap yields weaker B_0 between the two poles. The process of magnetic tuning through K to achieve the desired radiation is mathematically described by the undulator equation (2.3), which explains the generation of the first harmonic wavelength λ from an undulator through $\frac{\lambda_u}{2\gamma^2}$, where the Lorentz factor γ accounts for relativistic effects, magnetic

tuning is accounted through $\frac{K^2}{2}$, and off-axis wavelength variations are accounted through $\gamma^2\theta^2$.

$$\lambda = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2} + \gamma^2\theta^2\right) \tag{2.3}$$

As such, shorter wavelength radiation (increasing photon energy) can be achieved when K is smaller. Based on equation (2.1), a smaller K is only achieved when B_0 is smaller.

This is made possible physically by increasing the undulator gap (Fig. 2.4). Doing so allows electrons to travel through more quickly and achieve higher photon energy, resulting in higher frequency radiation. The average power $\bar{P}$ radiated by the electrons at average current $\bar{I}$ is related to K by the following

$$\bar{P} = \frac{\pi e \gamma^2 \bar{I}}{\epsilon_0 \lambda_u} \frac{K^2}{\left(1 + \frac{K^2}{2}\right)^2} \tag{2.4}$$

where ϵ_0 is the permittivity of free space.

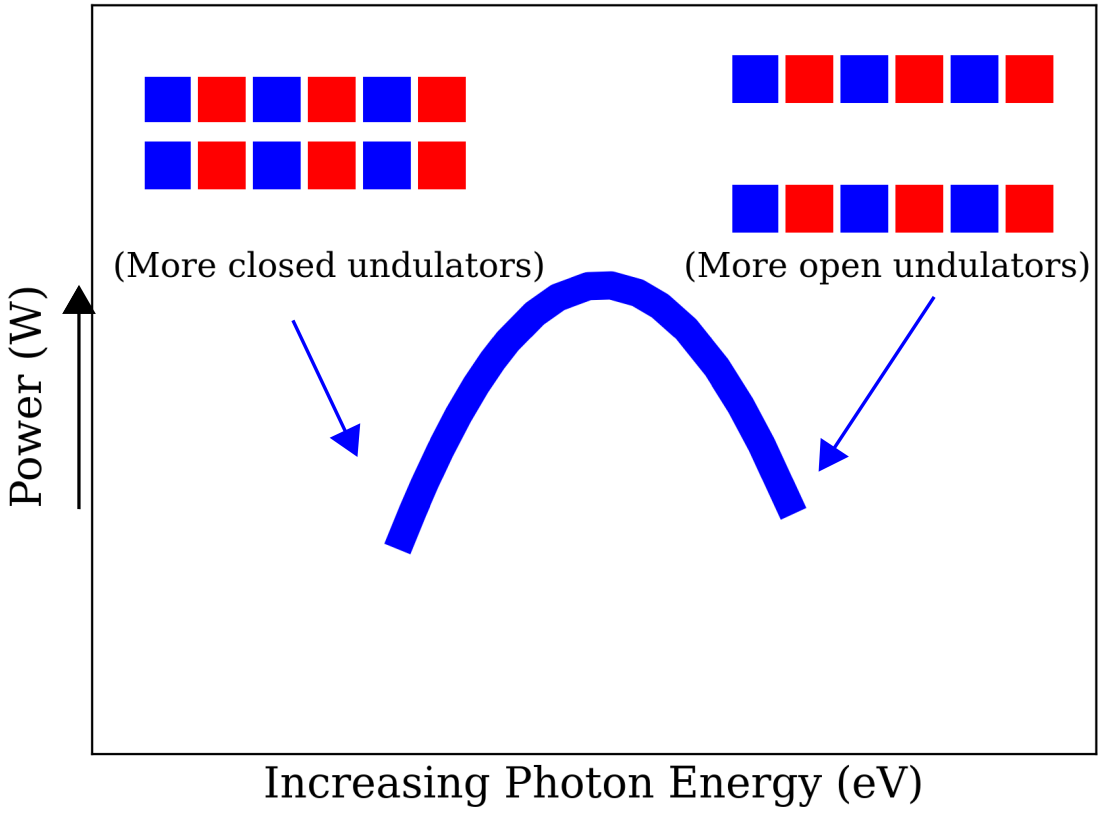


Figure 2.4: Undulator tuning curve schematic, adapted from [1].

At higher K (smaller undulator gap, left side of of Fig. 2.4), more energy is lost to harmonics, causing the electron beam to slow down. Reduced acceleration leads to decreased power. At smaller K (larger undulator gap, right side of Fig. 2.4), power increases as K^2 and peaks around K= $\sqrt{2}$.

2.3 From Undulator to XFEL

Having introduced undulators, we now direct our attention to XFELs. Although the underlying physics is similar, noteworthy distinctions exist between undulators and XFELs which is best summarized by Fig. 2.5. The main differences come down to length and electron beam characteristics In XFELs, the magnetic structures are typically extended to lengths exceeding 100 meters. The electron beam in an XFEL has a more compact and circular profile with reduced divergence, approaching the diffraction limit. Additionally, XFELs operate with significantly higher current, leading to a substantial increase in the electric field and facilitating enhanced spatial coherence.

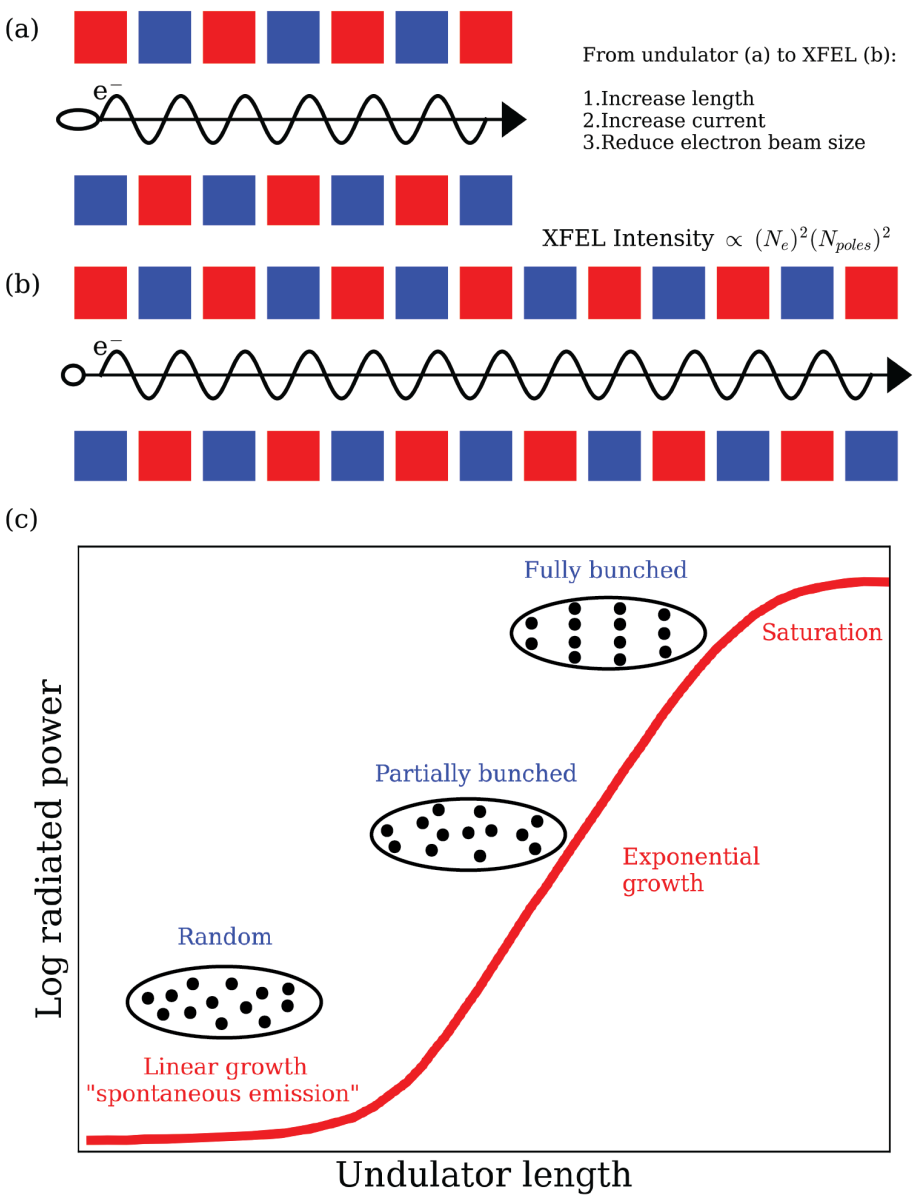


Figure 2.5: Evolution from undulator (a) to XFEL (b) accompanied by the resultant gain curve (c). As the undulator length extends, the initially randomly distributed electrons evolves, exhibiting an increasing degree of order before reaching the highest degree of electron modulation, otherwise known as "saturation." Once this point is reached, continued passage through the undulator will not be beneficial. The gain curve visually represents the amplification achieved in the XFEL process. Adapted from [1], and originally from [62].

The radiated wavelength from an XFEL is similar to that provided by the undulator

equation (2.3). When considering the on-axis case ($\theta = 0$), where most of the FEL radiation resides, the undulator equation simplifies to equation (2.5). The fundamental wavelength (λ) for forward radiation is described as follows:

$$\lambda_s = \frac{\lambda_u}{2\gamma^2}\left(1 + \frac{K^2}{2}\right) \tag{2.5}$$

As previously mentioned, the amplification process in XFELs involves larger electric fields, which cause the electronic structure to transform into well-correlated bunches. This leads to the constructive interference of radiated fields, initiating a phenomenon known as Self-Amplified-Spontaneous-Emission (SASE). The intricacies of the SASE process will be detailed below.

2.3.1 Self-Amplified Spontaneous Emission

A unique capability of XFELs is the fact that the full radiation power can build up from spontaneous emission in a single pass, through a process known as SASE. The seven steps of SASE, summarized from [1], are outlined below:

1. Electron injection: A beam of electrons is delivered into an undulator. These electrons can be generated through a variety of methods, such as heating a cathode, applying a strong field to a cathode, illuminating a photocathode with a laser pulse or using a radiofrequency gun.

2. Spontaneous Emission: As the electrons move through the undulator, radiation is emitted via spontaneous emission.

3. Transverse coupling: A transverse coupling emerges between E_{rad} and the transverse velocity v_x.

4. Energy Exchange: There is then an energy exchange between the electrons and the co-propagating field, described by:

$$\frac{dE_e}{dt} = mc^2\frac{d\gamma}{dt} = -eEv_x \tag{2.6}$$

5. Velocity Modulation: The energy exchange leads to electron velocity modulation. This then leads to bunching on axis. The electron density modulation results in stronger radiation.

6. Radiation enhancement: Electron density modulation leads to stronger radiation, so that P_{rad} is proportional to $(N_e)^2$.

$$P_{rad} \propto (eN_e)^2 \tag{2.7}$$

7. FEL action: A stronger radiated field drives stronger transverse velocity v_x. Stronger v_x drives stronger bunching which leads to stronger field. Consequently, FEL action occurs, resulting in substantial gain.

The gain length L_g (equation (2.8)), which determines the distance over which significant power is achieved and thus sets the efficiency and cost of the device, is described as follows:

$$L_g = \frac{\lambda_u}{4\sqrt{3}\pi\rho_{FEL}} \tag{2.8}$$

Equation (2.8) contains another important descriptor of FEL performance, ρ_{FEL}, which determines a combination of factors including the relative spectral bandwidth and saturation power of the instrument. At the Linac Coherent Light Source (LCLS) in California, ρ_{FEL} is approximately $\frac{1}{2300}$ [1]. ρ_{FEL} includes an effective undulator parameter $\hat{K}$ (which accounts for energy transferred to higher undulator harmonics), the electron radius r_e, number of electrons n_e, undulator radiation wavelength λ_u and lorentz factor γ.

$$\rho_{FEL} = \left(\frac{\hat{K}^2 r_e n_e \lambda_u^2}{32\pi\gamma^3} \right)^{\frac{1}{3}} \tag{2.9}$$

2.4 Seeding Techniques

One of the challenges introduced by the stochastic nature of the SASE process is that numerous microbunches can independently form within the longer electron bunch, each exhibiting a slightly different wavelength and phase. This results in a complicated radiated spectrum of uncorrelated spikes.

To mitigate this challenge, a seeding approach can be employed. One effective method, particularly at longer EUV wavelengths, involves coherent seeding using a high harmonic of a conventional femtosecond-duration laser. However, successful seeding depends on meeting several conditions. Firstly, the harmonic wavelength of the laser must be precisely matched to that of the undulator for optimal interaction between the seed and the electron beam. Therefore, a tunable seed source is critical. Additionally, the laser harmonic must possess sufficient intensity to dominate spontaneous emission in the initial sections of the undulator. This ensures that the seeded radiation controls the lasing process. If the seeding power falls below the level of the spontaneous radiation, the XFEL performance would resemble that of SASE, limiting the effectiveness of the seeding approach. Successful coherent seeding leads to enhanced control and predictability in the generation of coherent and intense X-ray radiation. This, in turn, enhances the efficiency and performance of the XFEL by reaching saturation at a shorter undulator length.

While achieving controlled conditions is feasible in the extreme ultraviolet (EUV) range, challenges emerge when attempting to implement coherent seeding at X-ray wavelengths.

To enhance coherence at X-ray wavelengths, one viable alternative is the technique of self-seeding, currently employed for hard X-rays [63, 64]. The self-seeding process typically involves two undulator sections within the XFEL and a frequency filter positioned at the midway point.

In this process, the first undulator operates in the SASE mode, generating relatively wide bandwidth of SASE radiation. Following the first undulator section, a magnetic electron chicane separates the radiation and electrons into two branches. The electrons travel through one branch, the chicane, where dispersion is introduced, disrupting the electron beam density modulation and resetting it to its initial shot noise distribution. This effectively eliminates microbunching and resets the electrons for the next set of undulators.

Simultaneously, in the photon branch, the X-rays pass through a tunable monochromator that selects a narrow bandwidth from the broad SASE spectrum. To ensure temporal overlap of electron and photon beams following the seeding monochromator, the time interval between them is adjusted by delaying the electron bunch through the magnetic chicane. Once overlapped, the monochromatic X-ray seed pulse and the electron beam co-propagate through a second undulator line. The seed pulse now dominates the bunching process within a narrower spectrum, resulting in X-ray pulses with a much narrower bandwidth relative to those generated through the SASE process.

2.5 Time Structure

There are several important implications stemming from the difference in time structure of radiation generated at an XFEL vs. at a synchrotron. At a synchrotron, with each orbit, electrons pass through cavities containing electromagnetic fields oscillating at radio frequencies (RF cavities). Such cavities provide a boost of energy to the electrons as long as the electron arrival is synchronized to the cavity's field [65], thus significantly influencing the time structure of synchrotron radiation. This leads to pulse duration on the order of picoseconds with bunches arriving every two nanoseconds at facilities such as SSRL.

XFELs, however, were developed to decrease time duration while increasing pulse intensity, leading to a vastly different time structure. XFEL pulses have durations on the order of femtoseconds and contain about 10^{12} photons. At LCLS, the highest repetition rate is currently 120 Hz, which means the shortest duration between pulses is 8 ms.

The time structure differences between a synchrotron and an XFEL have the most significant consequences for damage and data collection modes. Synchrotron-induced damage is primarily caused by the diffusion of radicals, which are highly reactive species with at least one unpaired electron. To mitigate radiation damage at synchrotrons, cryocooling is often employed. While this method is effective in slowing the diffusion of radicals, it has been shown to introduce biases into protein crystal structures, leading to adverse affects [66].

On the other hand, XFEL-induced damage primarily results from Coulomb explosion, where the intense laser field extracts electrons from the sample, leading to the accumulation of excess charge. This process occurs slowly enough for femtosecond XFEL pulses to outpace

structural damage due to their shorter duration. Consequently, data at XFELs can be collected at room temperature in a "probe-before-destroy" mode, where a diffraction pattern or spectrum is captured in a single shot before radiation damage occurs[67]. Richard Neutze and colleagues first proposed this concept in 2000 [33], and it has since been widely adopted by many research groups [68, 69, 70, 37, 71].

2.6 Outlook

Since operation of the first hard X-ray FEL just 15 years ago, this technology has had a profound impact on areas of science ranging from nanoscale dynamics and femtochemistry to structural biology [72]. This is despite the fact that the user community is relatively small, primarily due to limited access. Access is constrained because there are only a few hard X-ray facilities worldwide: LCLS in the USA (the first hard X-ray FEL facility), SACLA in Japan (the second hard X-ray FEL facility), PAL-XFEL in South Korea (third hard X-ray FEL), European XFEL in Germany and SwissFEL in Switzerland (both hard X-ray FELs), with another facility under construction (SHINE in China, hard X-ray FEL under construction) [73]. Additionally, each facility has a limited number of beamlines. For example, the European XFEL can manage around three experimental setups per week, totaling about 100 per year [74].

To address these limitations, there is potential for constructing additional XFELs around the world. For example, the UK is exploring a potential XFEL program, which would certainly reinforce the country's role in advanced science and innovation while also increasing access [75].

For now, however, submitting proposals and securing beamline time is a highly competitive process. Beamtime is often awarded to users who have already demonstrated their ability to work successfully at XFELs, creating a feedback loop that makes it harder for newcomers to gain access. Expanding the user base can prevent XFEL facilities from becoming exclusive clubs [76]. Efforts to achieve this could include programs that facilitate discussions and connections between new users, beamline scientists, and experienced users, fostering mentorship-based ties. Making XFEL science more routine through advanced scripts prepared by beamline staff, as done at SACLA in Japan, can also increase user independence. This approach would lower the barrier to entry while also reducing staffing strain.

The next decade of XFEL science promises advancements such as higher repetition rates. An example of this is the LCLS-II at SLAC, designed to deliver photons in the MHz repetition range using a superconducting RF linac [77]. Once LCLS-II is available for users, a number of downstream developments on the sample delivery, detector, and data analysis end will be required in order to take full advantage of the upgrade. Higher repetition rates offer benefits like improved signal-to-noise ratios, resulting in the ability to detect smaller signals. Upgrades such as this one can also enhance XFEL economics by allowing X-ray pulses to be multiplexed across more experimental stations, thereby increasing the number of experiments conducted simultaneously [74].

Another key area of XFEL-based science focuses on obtaining fully coherent and stable hard X-ray pulses. While XFEL pulses exhibit high spatial coherence, their temporal coherence varies. Fully coherent, transform-limited pulses could unlock new research areas in imaging, spectroscopy, and X-ray quantum optics [78]. In 2008, Kwang-Je Kim proposed an X-ray Free-Electron Laser Oscillator (XFELO) [79] which uses diamond Bragg reflecting crystals to form an optical cavity around an undulator, thereby reducing the bandwidth of X-ray pulse wavelengths and shot-to-shot fluctuations. This concept is currently being explored at the European XFEL in Hamburg [80]. Concurrently, a population-based X-ray Laser Oscillator (XLO) (see **Chapter 5**) is being investigated at SLAC [78]. An additional breakthrough area is the emergence of attosecond hard X-ray pulses, which allows for studies of real-time electron dynamics. While this specific topic is beyond the current scope of this dissertation, the interested reader can learn more about this more about this in Ref. [81].

Chapter 3

Navigating Data-Related Challenges

3.1 Introduction to Data Reduction at XFELs

3.1.1 Importance of Data Reduction

In the "probe-before-destroy" approach deployed at XFELs [82, 33], fresh sample must be introduced for each incoming X-ray pulse. Due to the stochastic nature of the SASE pulses (see **Chapter 2**), every event is, in essence, a unique experiment. Furthermore, depending on the efficiency of the sample delivery method and the behavior of the beam, the "hit" rate (hit = successful X-ray interaction with sample) may be quite low, resulting in a dataset that contains more than 50% empty or nearly-empty shots. This makes it necessary to collect and treat the data on a shot-by-shot basis, which can lead to extraordinarily large and cumbersome data sets. This is expected to become an even more prominent issue with upgrades to higher repetition rate facilities in the future (see **Section 3.4**).

LCLS currently operates at a repetition rate of 120 Hz. In optimal conditions, a given experiment has the capability of generating data on the order of tens of terabytes [83]. By the end of 2024 or 2025, user operations of LCLS-II are anticipated to begin. This upgrade will lead to a significantly higher frequency of data acquisition in the MHz range. This will not only increase the incoming data velocity, but also the complexity of the analysis which may hinder real-time feedback.

Scientific discoveries at XFELs rely heavily on adequate computational power. During a beamtime, updates regarding sample behavior and turnover, or potential contamination concerns, should be provided regularly. These updates may have consequences for the sample prep team, the setup, or the overall strategy of the shift. However, in order to have the opportunity to provide fast feedback, users typically need tools to handle the massive influx of data, making efficient management of large data loads all the more crucial. Moreover, because only a fraction of the data is actually useful for downstream analysis, the remaining surplus must be managed efficiently.

For example, consider collecting X-ray fluorescence data on an ePix detector [84], which records the positions and approximate energy of incoming photons. These detectors are

designed to have noise levels below 360 eV r.m.s. and a small pixel size (50 μm by 50 μm) [84]. These detectors are composed of 4 panels, each of which is (384 x 352) pixels in size. They are capable of collecting 7200 frames per minute or 432,000 frames per hour if running continuously at 120 Hz. This illustrates the need for efficient data handling and reduction.

In this chapter, the data reduction pipeline that is utilized during the XES experiments described in (**Chapter 4**) of this dissertation are highlighted (**Section 3.1.2 - 3.1.3**). Then, a few special cases, which involve additional algorithms for data classification (**Section 3.2**) and curvature correction (**Section 3.3**)(relevant to some of the data presented in **Chapter 5**), are detailed. I will end this chapter by describing future considerations for data analysis as facilities move to higher repetition rates (**Section 3.4**).

3.1.2 Data Reduction Pipeline for XES

The general data collection flow during an XES experiments at LCLS is illustrated in Fig. 3.1. During data collection, all device data is initially recorded on the detector. The ePix camera, and an example constructed detector image, can be seen in Fig. 3.2.

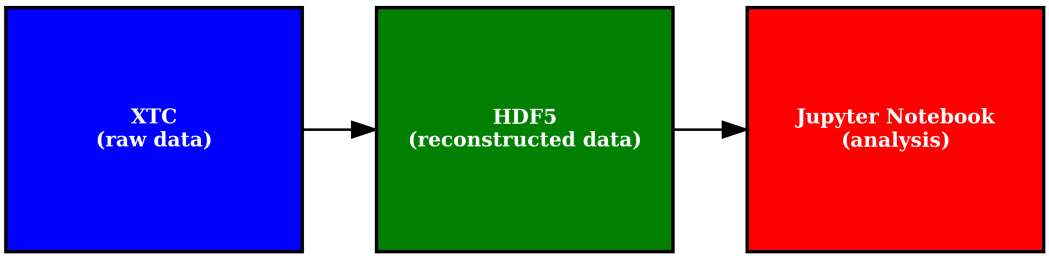


Figure 3.1: Flowchart outlining the typical data reduction steps for XES data collected at an XFEL. First, all raw data is written out to XTC format. The resulting files are quite large and must be reduced. They are then translated into HDF5 files. Finally, analysis is taken care of in a Jupyter notebook environment.

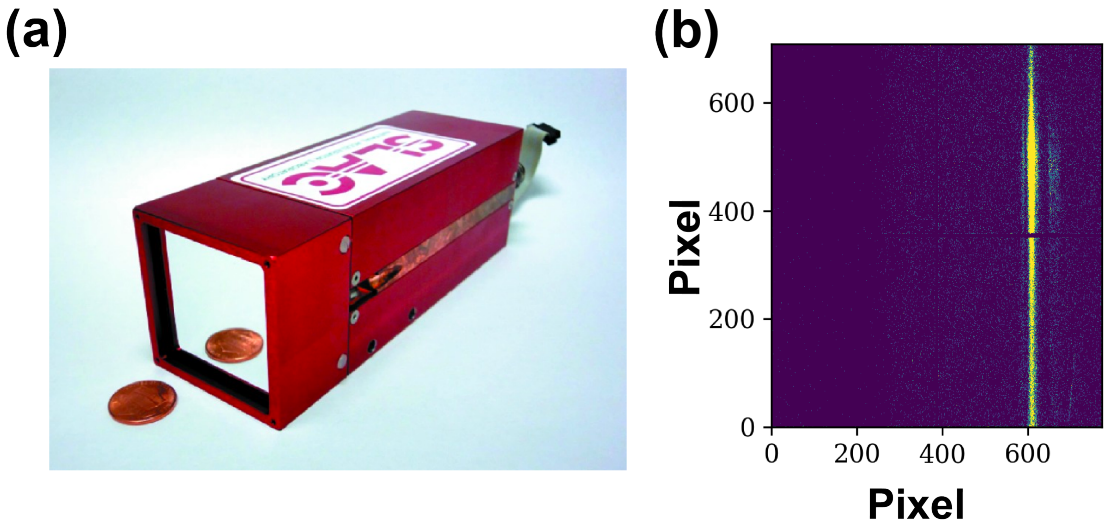


Figure 3.2: (a) ePix camera, reproduced from [84]. (b) Example detector image reconstructed from the recorded data (Mn K$\beta_{1,3}$). The spectral direction is shown along the vertical axis, and the spatial direction is shown along the horizontal axis.

This raw detector data is written out into XTC file format. A set of such files is generated for each "run" which can have a variable number of events contained within it (for our experiments, typically this number does not surpass 10,000). The experimental operator typically is the one to manually trigger the start and stop of each run. The number of runs collected during a given beamtime will vary depending on how smoothly things are running.

This XTC file type, the format of which depends on the detector and facility, contains all raw data. As a result, these files are extraordinarily large (up to 100 GB each) and need to be reduced. During the translation process from XTC to HDF5 format, typically a few corrections are done to the raw data behind the scenes.

The first is a "pedestal correction", which involves subtracting off the signal level from a run with no beam. In addition to this, a mask, which removes pixels with unreliable signal, can also be applied. A gain correction can also be used, which accounts for slightly different scaling in read-out signal of individual pixels with photon energy [85]. A common mode correction, which corrects for low-frequency read-out noise affecting detector areas, can also be applied [84].

The resulting HDF5 file can be considered an object with a hierarchical structure comprised of both groups and datasets. The groups in an HDF5 file will organize the data sets. The structure of an example h5 file generated during one of our XES experiments is provided in Fig. 3.3, where the following data sets are contained within the root group:

1. **x** and **y** coordinate of every photon hit / droplet
2. signal strength of every photon hit / droplet (**adu**)

3. **mask** for bad pixels on the detector
4. timing related information, such as event codes (**evr**)
5. number of droplets/photon hits per image (**ndroplets**)
6. intensity position monitor information (**ipm2**)
7. gas detector monitor information (**pulse energy**)

The HDF5 files can be loaded into a Jupyter notebook environment, and analysis tools in the form of python scripts are executed to visualize and further process the data.

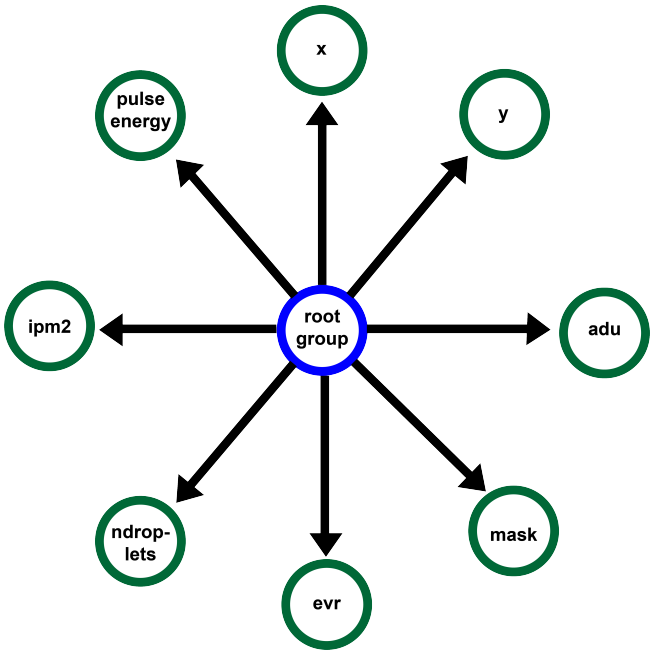


Figure 3.3: Schematic of the structure of an HDF5 file generated during an XES experiment. The blue node represents the root group within the HDF5 file. Green nodes represent datasets within the group. Arrows between nodes represent relationships between groups and datasets - with the direction of the arrow indicating the parent-child relationship.

3.1.3 Challenges and Considerations

There are a few challenges to consider in the reduction process. The first of which involves the distribution of charge on the detector and sharing between neighboring pixels. A schematic of this phenomenon is shown in Fig. 3.4. This issue has been relevant to the XES experiments described in **Chapter 4** and can be addressed through a reconstruction scheme based on a "droplet algorithm" [86]. This algorithm is typically used in the translation

process from XTC to HDF5. This technique involves iterating through all raw detector images, identifying pixels which are at a level surpassing a specific threshold, then inspecting the nearest neighbors and aggregating the signals if they surpass a second threshold. The resulting reconstructed positions and signal levels are then saved out for downstream analysis.

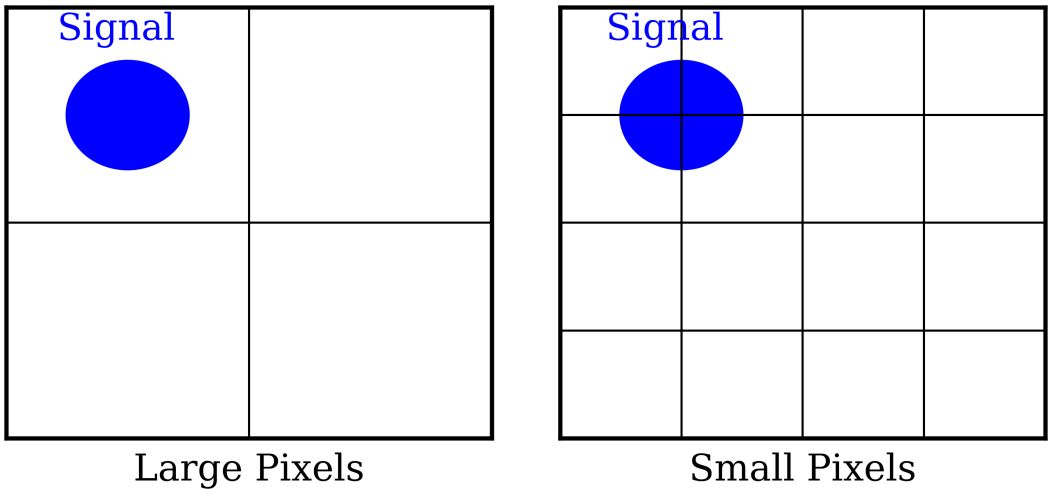


Figure 3.4: When the signal is collected on a detector with small enough pixel size, the charge cloud gets spread across adjacent pixels in what is known as "charge-sharing." A reconstruction algorithm is required to correct this.

This corrective approach encounters complications when photon density is too high, as the signals from two photons may erroneously combine as one. Thus, in order to ensure accurate reconstruction, thresholds must be adjusted based on the specific characteristics of the sample and the detector. In some cases, it is more advantageous to turn the droplet reconstruction algorithm off.

The major benefit from this reconstruction scheme is that it allows for better separation between the types of signals on the detector, so that artifacts and elastic scattering can be selected out. Values collected by the detector are typically called analog-to-digital units (ADU's). These correspond to the intensity at the pixel, and their values map roughly into the amount of energy a photon imparted to the detector. When generating a histogram of ADU values from the droplet data, at least three clear smooth peaks are visible (Fig. 3.5), allowing for better isolating the artifact, XES or elastic signals through ROI selection.

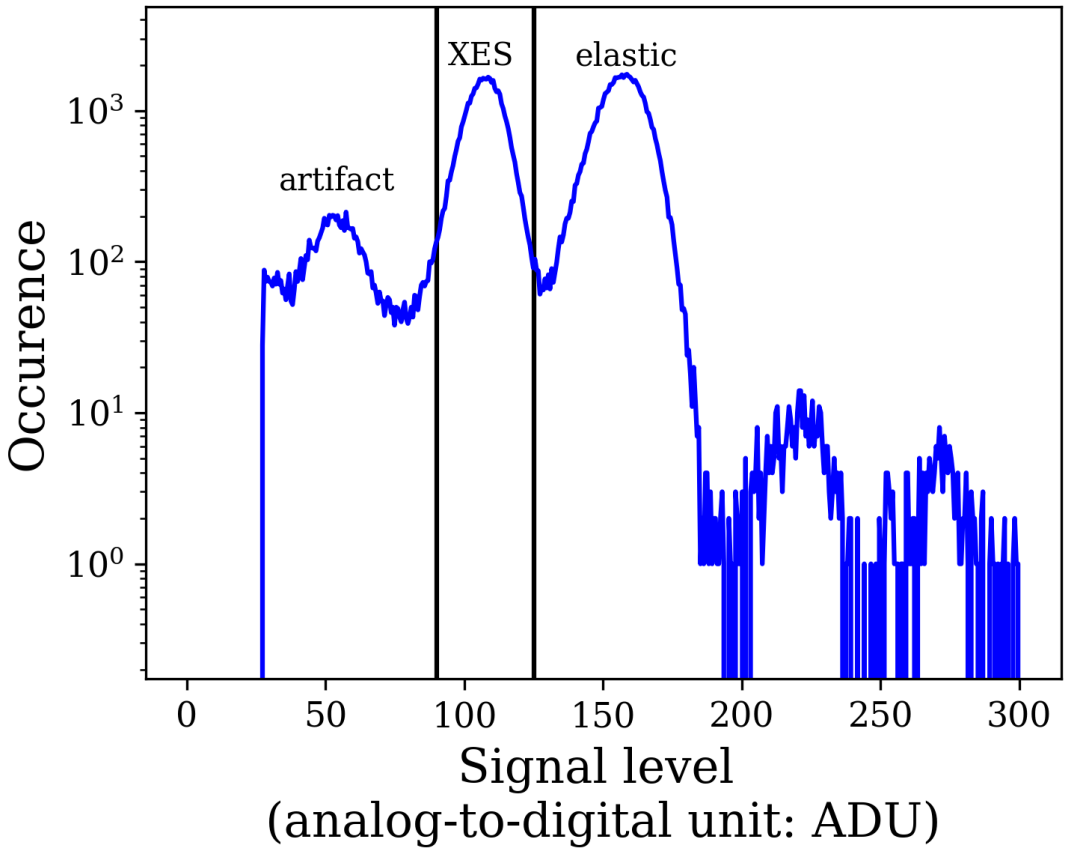


Figure 3.5: ADU histogram with three prominent features below 200 ADU consisting of an artifact of the detector, the X-ray fluorescence signal, and elastic scattering. The black vertical lines indicate the region of interest. There are two weak peaks at higher energies corresponding to multi-photon hits (around 220 and 280 ADU) which can be incorporated into analysis. The peak around 220 ADU comes from two XES photons, and the peak around 280 ADU comes from one XES photon plus one elastically scattered photon.

One can further sort the detector images for "hits" and "misses." If SFX is being collected simultaneously, then the indexed images from crystallography can be used to help sort XES data. This is beneficial as it allows for the best images to be identified; however, by using this method, shots which contain good signal for XES and poor signal for SFX are lost. Another option involves sorting images based on incoming beam intensity, and removing shots where beam intensity is too low.

Typically, the data ends up getting sorted according to the number of XES photons counted on the detector. Two thresholds are set for this kind of sorting. The first threshold specifies the minimum number of XES photons which are detected inside of the ROI (red dashed lines in Fig. 3.6), while the second threshold is related to the ratio of this number

to the number of photons identified in a region outside of the ROI (orange dashed lines in Fig. 3.6).

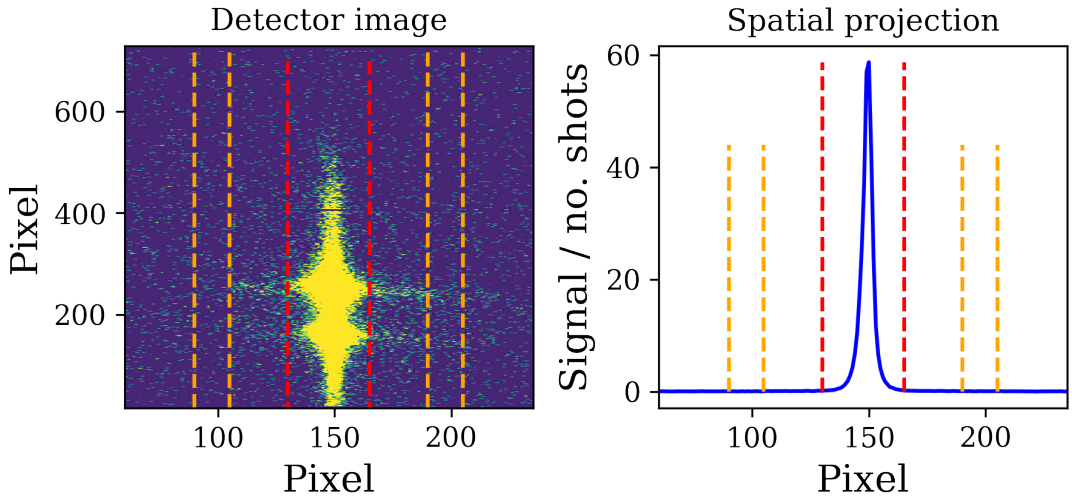


Figure 3.6: Example 2D detector image (left) and corresponding spatial projection (right) of a nickel Kα emission signal. The signal ROI (red dashed lines) and background ROI (orange dashed lines) are marked as well.

For PS II, we typically set a threshold such that only shots which contain at least three photons in the ROI, and at least twice as many counts inside the ROI compared to outside the ROI, are considered. The results from employing multiple techniques to identify hits have been compared previously and show similar results [87].

3.2 Exploring Machine Learning Models for Spectral Classification

XES data, especially when collected from low concentration metalloenzymes at XFELs (see **Chapter 4**), typically exhibits very low signal-to-noise levels on a shot-to-shot basis. This is further complicated by the often heterogeneous nature of XFEL data sets, due to sample heterogeneity along with pulse-to-pulse fluctuations [88].

Machine learning (ML) is a sub-field of artificial intelligence that involves computer algorithms trained from large amounts of raw data in order to construct models that provide advanced feature extraction as well as pattern and image recognition [89]. As a result, it has great potential as a workhorse at facilities that produce large, complex, multidimensional datasets, such as XFELs [90].

In this section, I will outline how I constructed an ML model from a library of model compounds containing a range of different metal centers for the purpose of analyzing and classifying individual low-signal shots. First, I assembled an XES dataset collected from diverse nickle, iron and manganese standards across a range of XFEL experiments and consolidating the data into HDF5 format [91].

I gathered over 19,000 distinct single shot images that were 50 pixels wide and 400 pixels tall centered very roughly on the emitted signal. The collection of data which I used to train, test and validate my model comprised eight distinct classes, with at least two options available for each of the three metals (Fig. 3.7). Additionally, a separate class was allocated for "dark" images (detector images with no beam.)

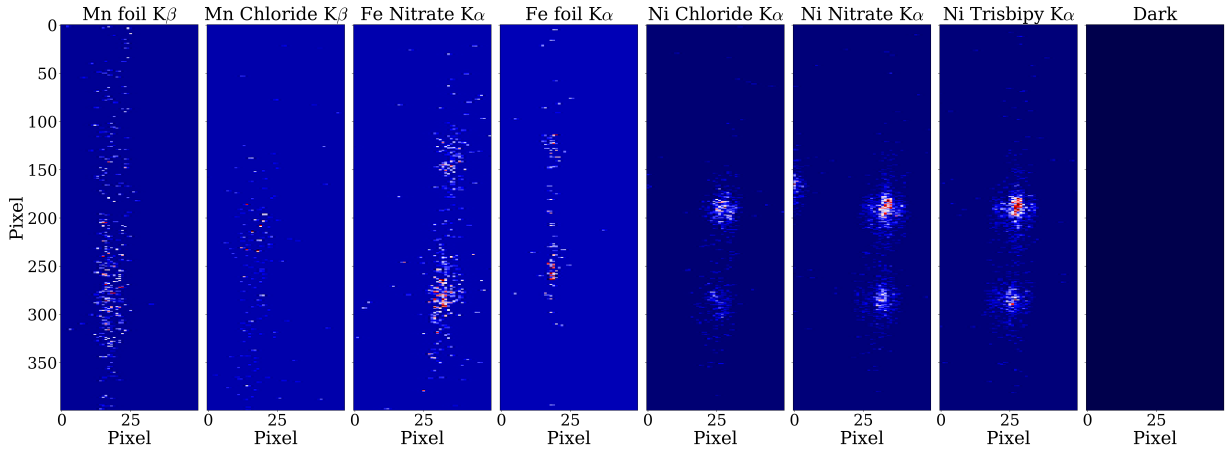


Figure 3.7: Example single shot detector images from each class used constructing my model.

The distribution of the data used for the ML algorithm is shown in Fig. 3.8. The dataset is not perfectly balanced and represents an uneven representation across classes, with manganese data dominating in total counts and "dark" data serving as the minority. Handling imbalance is a common topic in classification-based problems and has many solutions related to randomly re-sampling the data [92]. However, since all classes were on the same order of magnitude, none of these re-sampling methods were used for the work presented in this dissertation.

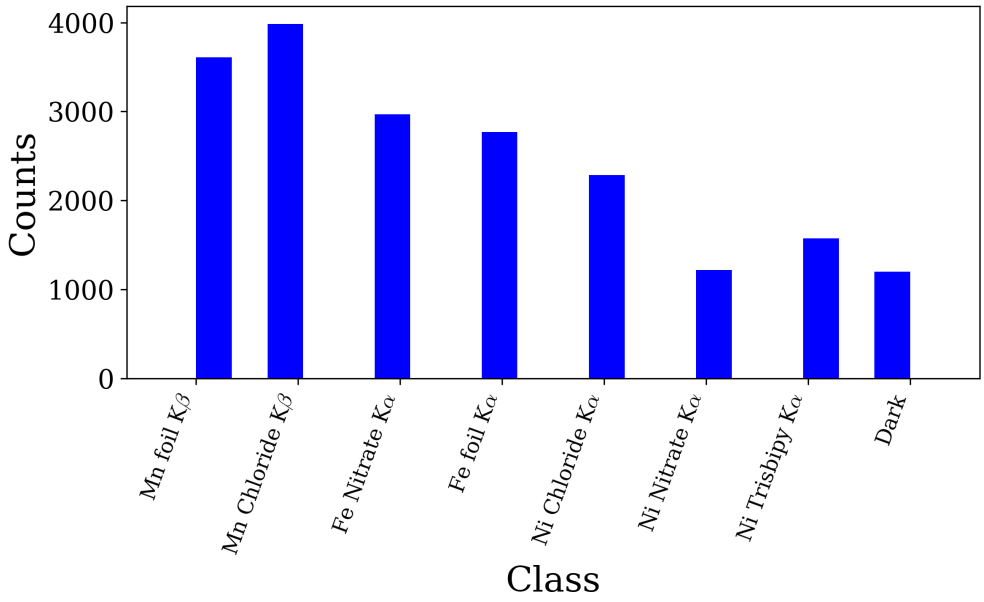


Figure 3.8: Histogram showing distribution of images for each class used.

These images were used to build an ML model which, when fed a 2D detector image, could accurately predict the metal center, the emission line, and (eventually) more detailed chemical information such as oxidation state or ligand environment. Eighty percent of the data was set aside for training, while ten percent was used for testing, and the remaining 10 percent was used for validation. A convolutional neural network (CNN) was then designed through a series of convolutional layers (for feature extraction [93]) and max pooling layers (in order to reduce the spatial dimensions of the input, thus reducing the computational complexity and improving efficiency) [94], batch normalization (enables higher learning rates and dense layers (otherwise known as fully-connected layer) [93]). This left me with 409,360 parameters, of which 409,232 were trainable.

To test my model, I constructed a confusion matrix which is a table containing different combinations of predicted and actual values [95]. The resulting confusion matrix is shown in Fig. 3.9. While the model experienced some trouble with mistaking iron nitrate Kα emission spectra for iron foil Kα emission spectra, and also confused nickle chloride Kα for nickle trisbipy Kα, it appeared to overall accurately predict the metal and spectral lines relatively well.

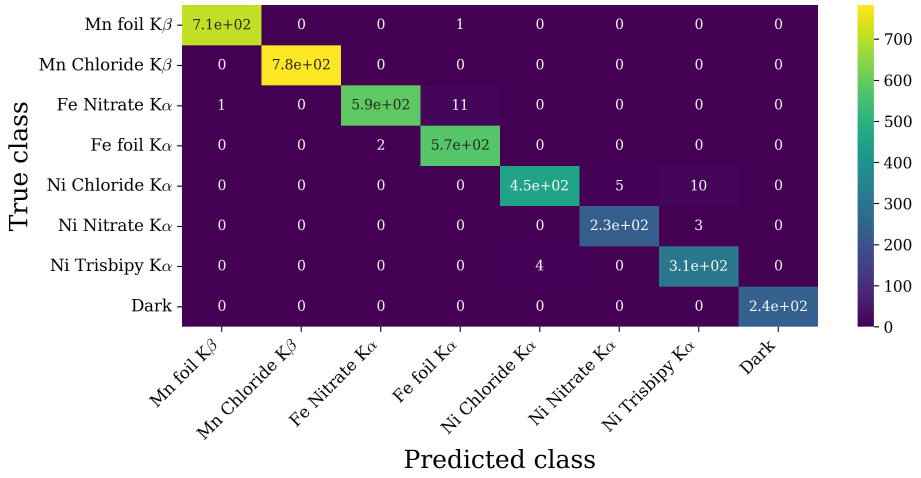


Figure 3.9: Confusion matrix demonstrating performance of the ML model with predicting spectral line and probed metal.

The third part of this work involves integrating the model into a simple Flask-based web app [96], where users may upload their own spectra and obtain a prediction based on the standards in my standards library used for this work. More details can be found here: *https://github.com/mdoyle17*.

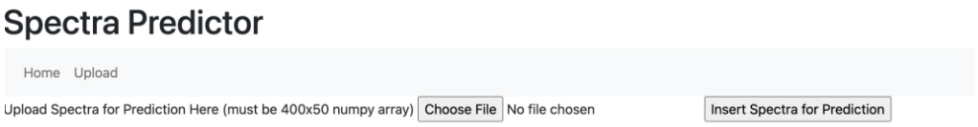


Figure 3.10: Screen shot of the resulting flask-based web application, showing the page where interested users can upload a 2D detector image and obtain a prediction of the metal probed and transition line measured.

This work, if expanded further, has the potential to significantly reduce data collection times in the future. Often, XES collection time at XFELs can be a bottleneck during measurements that are coupled with XRD. Fast feedback on sample oxidation state is critical in order to inform decision making and also identify sources of background contamination in certain cases. However, this type of information is not quickly resolvable from noisy data. The ML algorithm described in this section holds promise for classifying single images with few photons, thus significantly reducing collection time.

3.3 Image Curvature Correction Algorithm

3.3.1 Origins and Impact of Image Curvature

When data is collected, it can be corrupted by some kind of aberration, curvature or distortion. In this section, I will describe the design and validation of an image-correction algorithm which linearizes signal curvatures in 2D spectroscopic data. The development of this was made possible by significant contributions from Joshua Everts.

This algorithm was applied to a significant portion of the data presented in **Chapter 5**, which displayed noticeable energy-dependent quadratic distortion (see Fig. 3.11). This was attributed to imperfect optical alignment during data collection. In post-processing, it became imperative to devise an algorithm capable of independently correcting different sections of the same image.

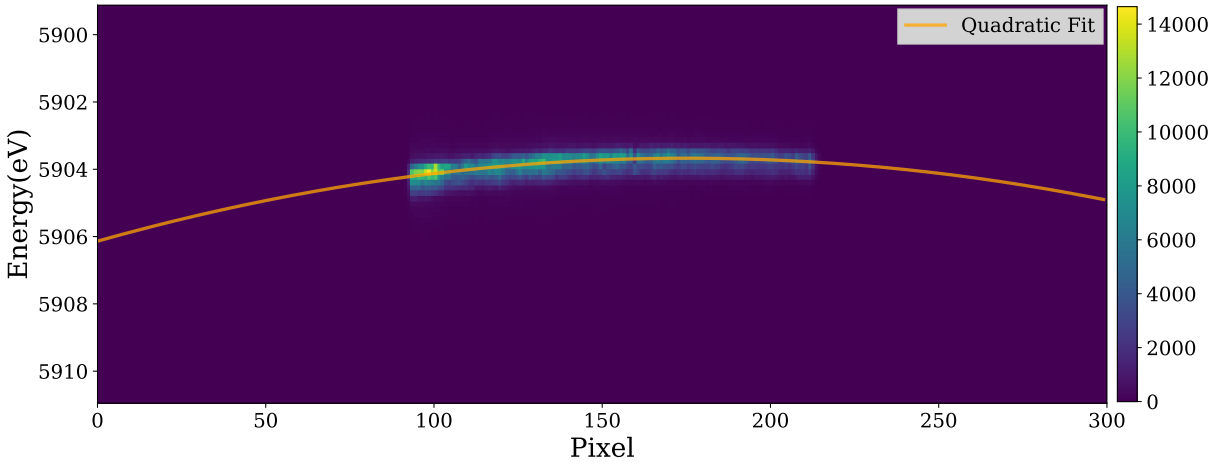


Figure 3.11: Identifying curvature of the raw detector image. In this image, only a seed pulse is present.

In order to do this, the distortion was first quantified across the detector using a set of "calibration scans," each of which contained the spectrum of a monochromatic seed pulse (generated using self-seeding, see **Chapter 2**), at a different photon energy. This allowed us to identify and characterize the image curvature, with no sample present, across a range of energies. For this, a quadratic shear function of the form

$$\text{shear}(x) = k_2(x - x_0)^2 + k_1(x - x_0) + k_0 \tag{3.1}$$

was fit to each seed pulse using a center-of-mass (COM) algorithm (see **Section 3.3.2**). This resulted in a set of k_2, k_1 and k_0 for each seed energy. A quadratic interpolation function for each parameter $[k_2, k_1, k_0]$ as a function of energy was generated and could then be applied to correct the full detector image.

3.3.2 Center-of-Mass Algorithm

The approach used for shearing 2D image data is to first reduce it to a set of 1D data and then apply a correction from there. This is the approach that the COM method employs. First, a 2D image is read in, and the necessary background-correction and dark corrections are applied. Next, the COM algorithm starts by splitting the image into a series of vertical ribbons, each with the same slice width, and extracts a projection from each slice.

From there, two different approaches can be used to find the COM. One approach is to simply use the `numpy.argmax()` function to return the pixel value in y which contains the maximum, but this is limited to integer pixel values and is not well-equipped to deal with a thin seed pulse ranging over only a few pixels to begin with. Instead, for a finer estimate of the true COM, a Gaussian fit can be made to the y-projection, where the centroid value of the fit is the COM for that y-projection. This concept can be visualized Fig. 3.12.

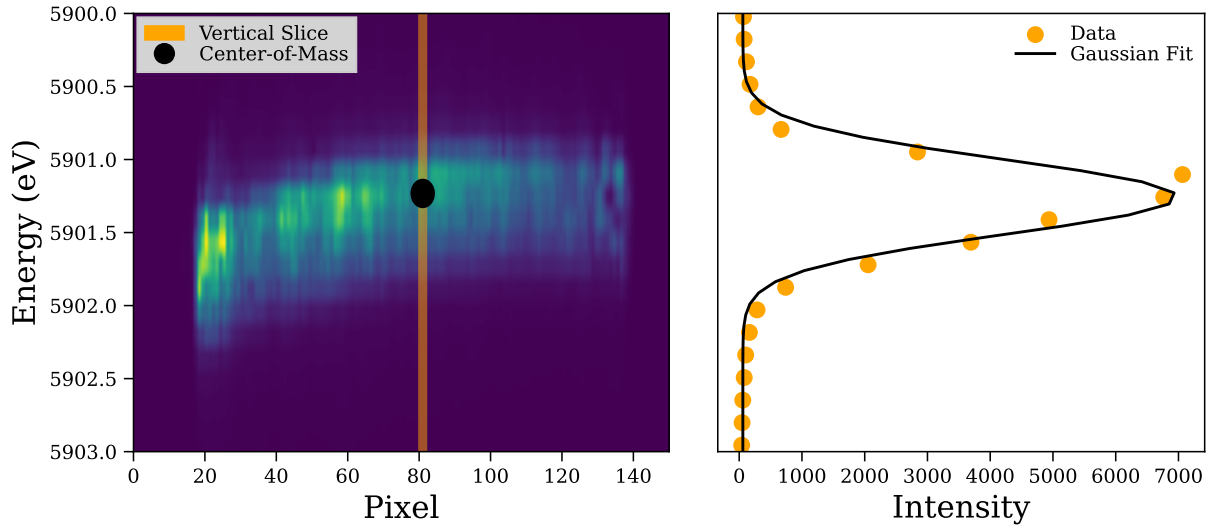


Figure 3.12: 2D detector image of monochromatic seed (left) and its y-projection (right) with the corresponding Gaussian fit.

However, one issue with this approach is that many regions of the seed pulse either (a) do not have enough data points or (b) simply do not fit a Gaussian distribution. In order to combat this, a `try/except` structure is used so that only those fits which do converge are included in the COM data that is output. Finally, even amongst the fits that technically converged, many did so with large chi-squared values, and these resulted in COM outliers that hindered further analysis. In order to remove these outliers, which typically existed at the lower-intensity regions of the seed pulse, a "threshold value" was added as a function parameter. The threshold value ensures that only those COMs that resulted from a Gaussian of an amplitude greater than the threshold will be included in the final output. By

adjusting this parameter, along with the slice width, it was possible to obtain significantly more consistent and outlier-free COM data.

Image curvature was then identified by outputting the COM data for a single calibration run at a given seed pulse photon energy, the optimized shear parameters for that run's corresponding energy must then be calculated in order to eventually apply a correction across the full energy range of an entire image. To do this, a quadratic function is fit to the seed pulse of a given energy by identifying and fitting the COM of the data. Then, the parameters $[k_2, k_1, k_0]$ are extracted from the fit and output into an array. The process is repeated for each seed pulse. A quadratic shear was chosen because it offered the best quantitative correction statistics when compared to higher ordered polynomials.

The signs of these parameters are then reversed before they are input into a shear function which multiplies every column of the image by the given quadratic shear. In order to shear an entire image, not just a seed pulse of a given energy, it is necessary to differentially shear slices of the image with their corresponding parameters. To correct an entire 2D image, a set of shear parameters are generated from each calibration scan and then a quadratic interpolation function for each parameter $[k2, k1, k0]$ as a function of energy is generated using `scipy.interp1d` - see Fig. 3.13.

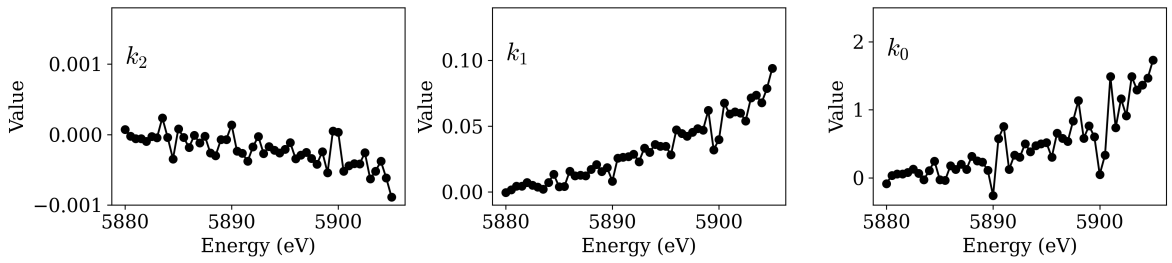


Figure 3.13: Linear interpolation to optimized constants in quadratic fit which are used to correct the detector images.

Once the interpolation functions are known, they can be used to generate shear-parameters for any given energy along the y-axis of an image. A full image is corrected in the following manner. First, it is divided into horizontal slices 5 pixels in length. Next, the energy corresponding to the center of each slice is used as the input for retrieving the appropriate shear parameter set $[k_2, k_1, k_0]$. A shear correction using the selected parameters is then applied across the full image. Finally, the slice from the now-corrected image is transferred into a new image array, and the process repeats for all slices until the final image is rebuilt in full. A schematic of this process is provided in Fig. 3.14.

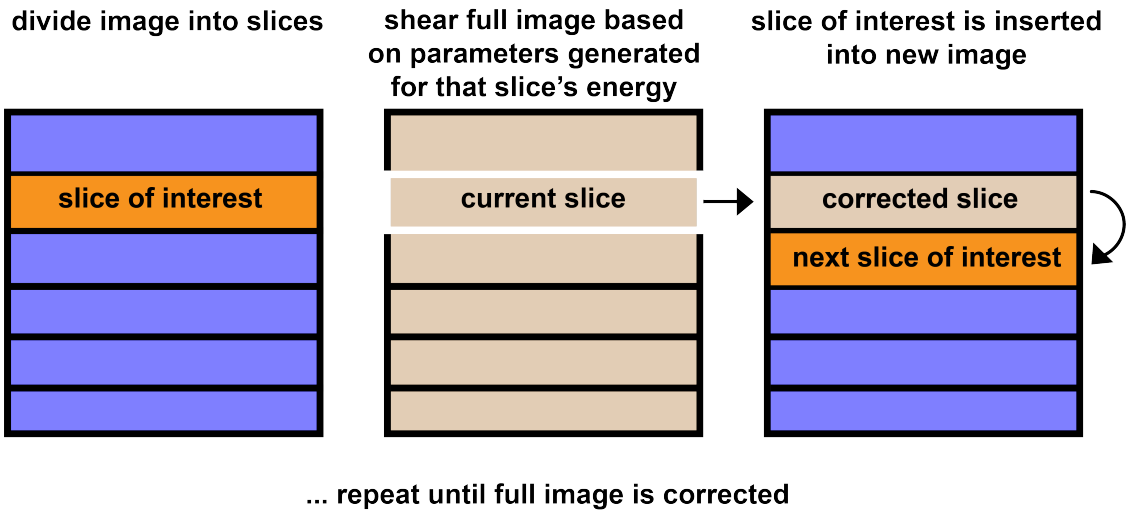


Figure 3.14: Schematic of full image correction procedure. First, each image is divided into a set of horizontal bars of defined slice width. Next, the interpolation functions are applied for that slice's energy, generating shear parameters. The entire image is sheared, and the slice of interest is placed into a brand new image. The process is repeated for every horizontal slice until the entire image is corrected.

3.3.3 Image Correction Results

There were several quantitative conditions that had to be met before a correction was considered "good". Most importantly- the amplitude from the y-projection of the corrected image had to increase with respect to that of the original. This condition would indicate a straighter image, in addition to a qualitative "by-eye" estimate. The latter can be best visualized in Fig. 3.15, where the full-image shear correction described above was applied to a run containing a stimulated emission signal (see **Chapter 5**) from $NaMnO_4$ solution. A noticeable, visual improvement after the correction is observed.

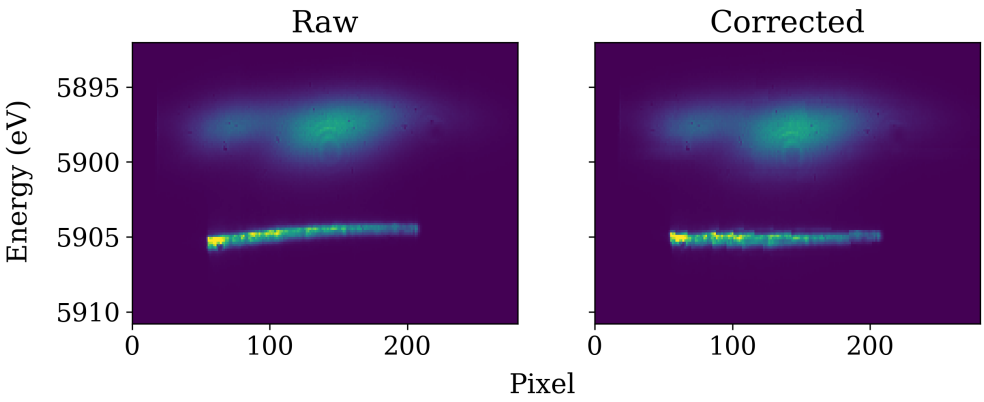


Figure 3.15: Visual representation of a detector image before (left) and after (right) correction. The broad "cloud" is a stimulated XES signal (see **Chapter 5** for more details), and the intense and narrow feature is the monochromatic seed. The two features become more parallel to each other, as they should be, after the correction is implemented.

The image correction presented meets the initial goal: To linearize energy-dependent curvature in the data resulting from non-perfect optical alignment. The COM method proved successful across a range of different energies and images, and was fast enough to be applied across multiple images in series. This method can be applied to other cases of image curvature in the future.

3.4 Outlook

As XFELs upgrade to repetition rates in the MHz regime (LCLS-II [97]), similar to what is already available at the European XFEL [98], there will be a substantial leap in data collection rate. This has the potential to present a significant computational challenge to users, in addition to the slew of other difficulties posed for sample injection and detector technology [98]. Larger files and higher acquisition rates makes real time feedback, flexibility and efficient storage even more crucial to consider as the sample is also being used up at a higher rate. To mitigate this, facilities such as LCLS have designed their own adaptable data reduction pipeline and online monitoring analysis software for users that is continuously evolving in preparation for the upgrade [99].

Some facilities have also proposed incorporating ML methods into their data reduction pipeline. This has the potential to reliably categorize useful and non-useful images, further streamlining the data handling process [83, 100]. Furthermore, in a large collaboration involving scientists from Argonne National Laboratory (contains the Advanced Photon Source which is currently undergoing a comprehensive upgrade), ML has been deployed in efforts to parse through large volumes of particle collision data generated from the Large Hadron Col-

lider and identify anomalous events [101]. This type of analysis could serve as the blueprint for detecting rare events in XFEL studies down the road.

Chapter 4

X-Ray Emission Spectroscopy Measurements at XFEL

4.1 Introduction

XES is a powerful element-selective technique for probing the electronic structure, charge and spin states, and ligand environment of 3d transitional metal complexes [40, 102]. It is a useful tool for investigations into the electronic structure of metalloenzymes, which contain a metal center that is integral to catalysis. Better understanding the redox changes experienced by the metal center is a crucial step towards more thorough knowledge of the enzyme's biological function.

XES measurements were initially performed using X-ray tubes [103, 104, 105]. Then, around the late 1990's and early 2000s, synchrotron radiation sources started to become a popular option for collecting high-resolution XES spectra. One of the noteworthy experiments around this time involved measuring Kβ X-ray fluorescence from dilute metalloproteins using the high flux provided by a synchrotron [106].

A significant challenge arises in trying to study anything other than the ground state at a synchrotron. For this to take place, the intermediate needs to be cryo-trapped before measurement [102]. This can be a complicated task for especially short-lived intermediates. Cryo-trapping also carries with it the implication that the biological system of interest is no longer being studied under functional conditions.

A solution would involved room temperature data collection. However, room temperature data collection at a synchrotron is challenging due to the secondary radiation damage through the generation of unwanted radicals. Since the time scale of the diffusion of this radical is on the order of the pulses, it is often difficult to collect a damage-free room-temperature synchrotron measurement.

However, the ultra-short and ultra intense pulses provided by XFELs open up the possibility for collecting room temperature time resolved spectra on metalloenzymes in order for electronic structural dynamics to be tracked while the sample is under biologically relevant

conditions.

One of the early concerns of using such intense pulses for XES studies centered on potential damage to, or alteration of, the sample during data collection. In one of the first experiments conducted at LCLS, XFEL pulses were demonstrated to generate non-linear effects at high peak power densities [107], which can diminish the integrity of the sample and affect spectral features.

While the probability of X-ray induced damage through sequential photon absorption within the time scale of the XFEL pulse is quite low, the extent and time scale of XFEL pulse-induced Coulomb explosion is more difficult to understand [38]. For XES studies, there was concern that electronic rearrangements associated with Coulomb explosion would affect the collected spectra and, in the worst case, lead to misinterpretation of the data.

To address this, Alonso-Mori and coworkers [38] carried out femtosecond XES experiments at the coherent X-ray imaging (CXI) instrument at LCLS using redox-active aqueous manganese complexes, measuring data in shot-by-shot mode with a multicrystal energy dispersive spectrometer based on a von Hamos geometry (more in **Section 4.2.1**). The team demonstrated that the spectroscopic XFEL data was in good agreement with the undamaged synchrotron measurements - implying that room temperature XES spectra can be collected at XFELs before the electronic structure is modified. This seminal work opened the door for more probe-before destroy spectroscopy-based techniques to be carried out on other samples at XFELs.

The advent of serial femtosecond crystallography measurements at XFELs has uncovered structural details revealing the transient states of catalytic reactions [108]. This has accelerated the development of in-crystallo spectroscopy methods to complement this [109]. Combining spectroscopy with crystallography is a powerful way to correlate electron and protein dynamics, evaluate associated kinetics, while also verifying the integrity of the sample.

Furthermore, by collecting XES, one can immediately tell whether radiation damage has occurred. This is challenging to do using crystallography alone. Experimentally-speaking, the two methods also fit together nicely because XES is a non-resonant technique - not requiring monochromatic or tunable X-ray beams - unlike XAS. All that is required is for the incident X-ray energy to be tuned above the absorption edge of the metal of interest. This fact enables the full XFEL SASE pulse to be deployed, efficiently combined XES with XRD.

In this chapter, I primarily focus on K-edge XES experiments conduced at the LCLS. This technique involves an incoming photon getting absorbed by a 1s electron, which is then subsequently excited into the continuum. This excited state has a short lifetime, before it experiences radiative decay which gives rise to a fluorescence spectrum (see **Chapter 1 for more details**)

Photosystem II, the metalloenzyme mentioned in this chapter contains an active site comprised of one transition metals: Manganese. It also has a K-edge that lies in the hard x-ray regime [40]. Hard X-ray probes are quite advantageous for a number of reasons, in

particular their greater penetration depth that enables experiments to be conducted in air and out of vacuum.

I will first describe our experimental setup (**Section 4.2**), which is deployed for our XES (solution), and simultaneous XES + XRD (crystal) measurements at an XFEL. Next, I will detail how XES is used to provide real-time in situ diagnostics on sample and measurement quality (**Section 4.3**). I will then discuss how reaction kinetics are analyzed through time-resolved X-ray Emission experiments (**Section 4.4**), before going into results on PS II (**Section 4.5**). Finally, I will discuss the limitations of this work and future directions (**Section 4.6**).

4.2 Experimental Overview

Collecting XES and XRD data simultaneously can be a complicated by the fact that optimal sample delivery conditions for each, such as crystal size, sample path length, and overall sample consumption, are different [110]. Moreover, engineering how to efficiently position samples into the beam, (otherwise known as "Sample Delivery") is an art form of its own that generally requires dedicated personnel. In brief, sample delivery at XFELs can take on many forms, involving liquid jet injection, a viscous medium, or fixed-target scanning. The last option, fixed-target, involves spraying crystals into a sample holder and delivering them to the X-ray interaction point with motion stages - this is similar to what is done at a synchrotron. A more in-depth discussion on sample delivery forms is beyond the scope of this work, please see Park et al. for more [111].

In our measurements, we deploy a hybrid sample delivery method, called the "Droplet on Tape" (DOT) method. This consists of a larger reservoir of sample (similar to that of an injection method) that deposits a single drop of a crystal suspension or solution onto a continuously moving kapton conveyor belt. This is done with an Acoustic Droplet Ejector (ADE). The droplet is then transported through an interchangeable excitation region where the sample is activated before being delivered to the focused location of the XFEL beam (similar to the fixed target method). The sample is blasted by the X-ray pulse and the cycle repeats with a new droplet. The timing between reaction initiation (pump) and probe is carefully set so that time-resolved data can be collected and reaction intermediates can be better studied. Reactions can be initiated via gas incubation, light activation or droplet mixing. For studies on PS II, which is the primary focus of this dissertation, light activation is the method for initiating the reaction cycle, i.e. the sample is pumped into the different S-states of the Kok cycle using 527 nm ns light flashes from a fiber-coupled laser.

Geometrically-speaking, for our multi-modal measurements, the X-ray diffraction is collected in the forward direction, along the beam axis, and the fluorescence data is collected on the side, although fluorescence photons are emitted in all directions. The most efficient von Hamos spectrometers, however, are operated above 75 degree Bragg angles [37]. In our emission setup, it is typically set at 90 degrees in order to minimize background contributions of elastic scattering from the sample. The detector position generally depends on the

direction of polarization of the XFEL beam. An example schematic of the setup can be seen in Fig 4.1.

There are a few advantages to this method, the first of which involves its high stability, which allow for conditions to be maintenance-free for hours at a time. This method also keeps sample consumption levels much lower than other methods, such as Rayleigh jets. Previous work by Chatterjee et al. demonstrated that the switch from a Rayleigh jet sample delivery system to the DOT system was able to reduce sample consumption by up to two orders of magnitude [112].

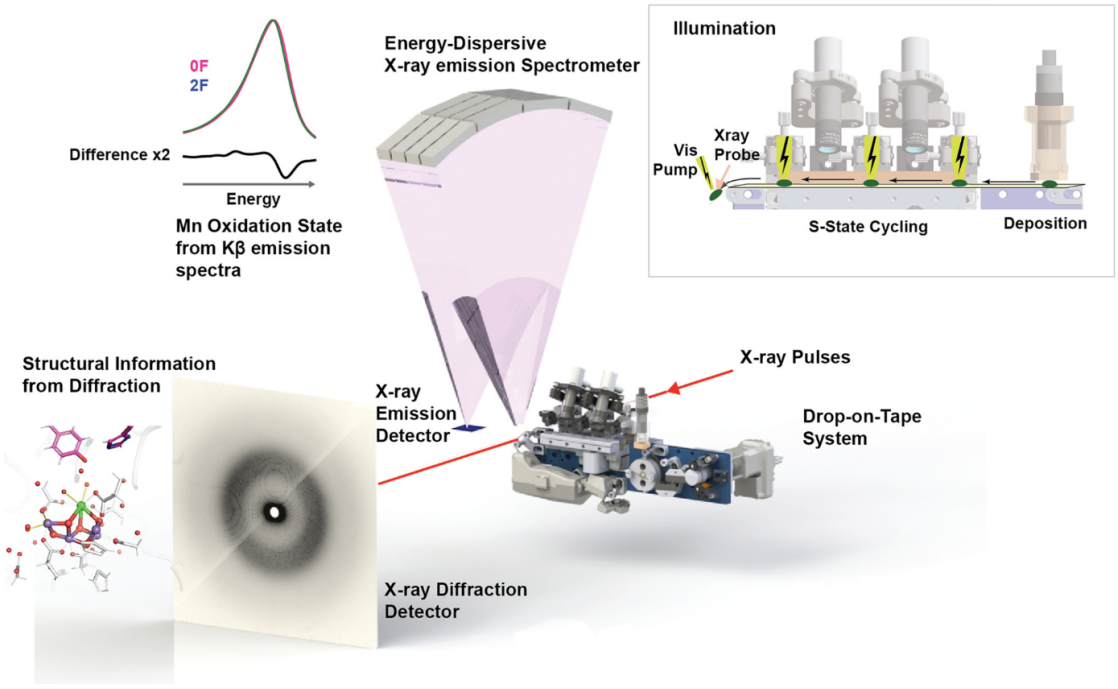


Figure 4.1: Schematic of the drop on tape setup for simultaneous collection of XES and XRD data, reproduced from [113]

4.2.1 Spectrometer

Multi-modal, time resolved measurements involving XES data collection are made possible through the development of spectrometers which enable a full emission spectrum to be captured in a single shot. For this task, we take advantage of the von Hamos geometry which uses cylindrically cut analyzer crystals that diffract emitted signals in accordance with Bragg's law (equation (4.1)):

$$\lambda = 2dsin\theta \tag{4.1}$$

onto a 2D position-sensitive detector, which is mounted on a horizontal plane. The chosen crystals typically have a lattice d-spacing, or distance between adjacent lattice planes, that can select for specific energies before reflecting them onto the detector [114].

One drawback to the von Hamos geometry, however, is that the signal-to-noise ratio is quite low due to all solid angle being collected into an extended area. This is in contrast to the Rowland geometry used at synchrotrons, for example, which involves a scanning monochromator paired with a spherically curved Bragg crystal and involves the analyzer angle and detector position to be scanned, thus ruling out the possibility of single shot measurements [40]. This configuration results in all solid angle getting collected into one point, resulting in higher signal to noise. In order to best minimize background scattering during our experiments, the entire tape drive system is typically housed in a helium environment within an acrylic box which contains openings for collecting the fluorescence and diffraction.

The crystal arrangement will typically differ depending on the sample. For our studies of the Kβ emission spectrum from PS II, we use an array of 4 x 4 Si(440) analyzers with a radius of curvature of 500 mm perpendicular to the scattering plane [114]. In other experiments, by installing appropriate analyzer crystals on the spectrometers, we are able to collect different combinations of Kα and K$\beta_{1,3}$ XES on Mn, Fe, and Ni containing metalloenzymes simultaneously with our XRD data collection [115, 116, 117, 118].

While increasing the number of crystals has the potential, in theory, to improve the signal-to-noise ratio, the increase can also cause alignment-related complications. This is especially the case when utilizing the 4 x 4 configuration for PS II, which requires precise tuning to focus the spectrum on the horizontal axis such that all 16 spectra are aligned onto the same x-axis in the energy domain. Aberrations in alignment can be caused by a tilt in the entire spectrometer, imperfections on analyzer crystal surface, or a mechanical tilt of the crystal stages that cannot be corrected. If not all crystals can be aligned, one can judge whether or not it is straightforward enough to work with one of the XES signals, or merge all of the spectra together. The latter option improves signal to noise at the cost of the energy resolution. The schematic in Fig. 4.2 is an example of the experimental setup geometry for XES which relates the detector, sample, Bragg angle, and radius of curvature all to one another.

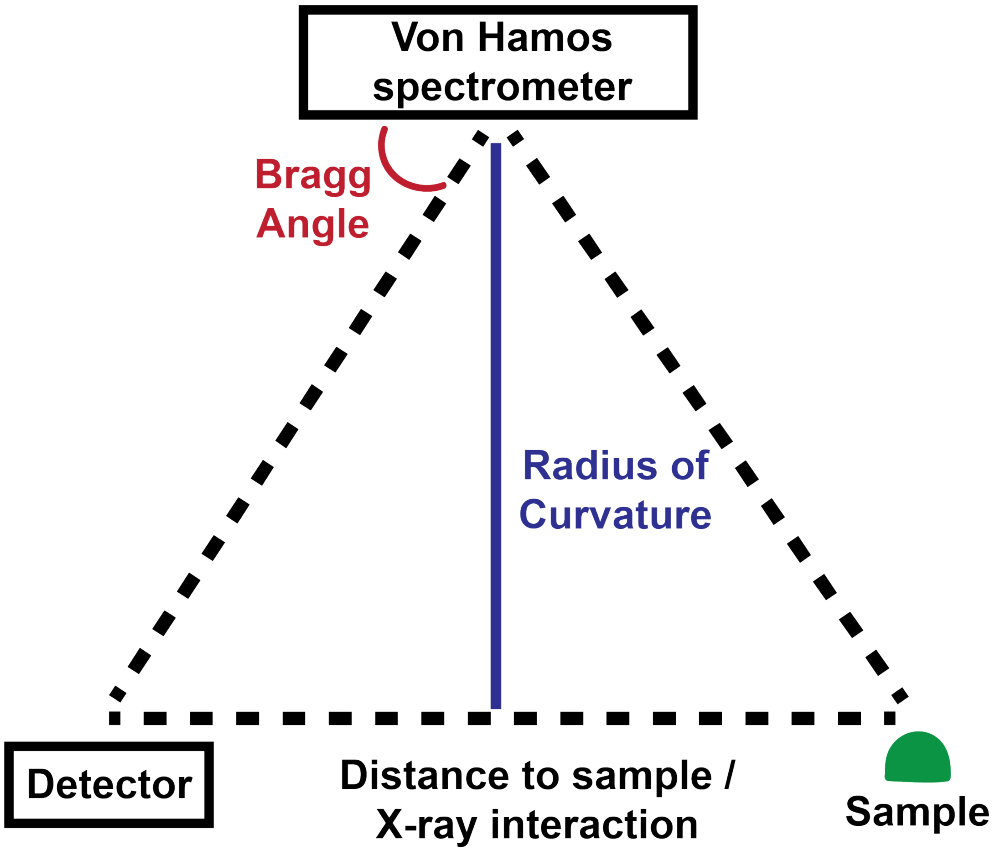


Figure 4.2: Experimental setup geometry for XES.

4.2.2 Experimental Conditions

For our combined XES and XRD experiments which are carried out at the Macromolecular Femtosecond Crystallography (MFX) instrument at LCLS, our experimental conditions are typically as follows: We deploy an incident beam at 9.5 keV, with a repetition rate between 20-30 Hz. Our pulse length is typically around 40 fs, with a focus around 4 μm FWHM.

4.3 In Situ Diagnostics

Since the samples used in XFEL experiments are destroyed immediately after data collection, many different batches are typically prepared beforehand in order to fill the full

experimental period. This underscores an inevitable aspect of SFX experimentation: The data we collect can often be heterogeneous.

Each incoming X-ray probe pulse is stochastic and exhibits shot-by-fluctuation, the sample itself also can vary, in part because we collect from a variety of crystals that differ in both size and quality. Crystals are also sensitive to factors such a temperature, light, oxygen, and environmental conditions, leading to various levels of degradation. Sample degradation can result in a decrease in activity. For all of these reasons, in situ diagnostics for quality control purposes is critical.

In addition to monitoring sample quality, in situ monitoring is also crucial for pump-probe experiments, such as the ones described in this chapter, to confirm the generation of chemical intermediates. During our SFX experiments, we deploy XES as our in situ diagnostic tool to monitor the chemical nature of the metal catalytic center under study. During these experiments, online data analysis and fast feedback regarding sample integrity and turnover of the catalytic reaction is a critical component. In this section, I will focus primarily on using XES for in situ diagnostics of PS II during crystallographic experiments, although what I am presenting can be applied, and has been applied, for studies on other metalloenzymes.

In order to maintain an active, live in situ feedback system during a beamtime, it is important to have some kind of simple descriptor of the spectra that can be generated quickly. Since the emission data we collect from PS II ($K\beta_{1,3}$) is typically low in signal levels due to both (1) the dilute nature of the sample and also (2) the 8x weaker signal generated from $K\beta_{1,3}$ relative to $K\alpha 1, 2$ emission, simple interpretation from difference spectra is not always straightforward. As a result, we resort to tracking the first moment (equation (4.2)), which can also be thought of as the center-of-mass, during our manganese ($K\beta_{1,3}$) measurements:

$$1^{st}\text{Moment} = \frac{\sum_i E_i I_i}{\sum_i I_i} \tag{4.2}$$

where E_i and I_i correspond to X-ray energy and emission intensity of data point i, respectively [87]. This descriptor is highly sensitive to the spin state [119] (hence, oxidation state in the case of high-spin complexes such as PS II). An added benefit of this descriptor is the fact that it is less susceptible to noise than other parameters, while also maintaining spectral sensitivity [87].

During a PS II beamtime, the first moment can be tracked for individual batches, thus monitoring progressive Mn(II) accumulation in the sample over time. The significance of this comes from the fact that excess Mn(II) can be released from the OEC during on-site sample handling. As a consequence, a gentle rise in 0F (ground state) first moment can indicate that the batches are becoming increasingly contaminated [87].

For such experiments on PS II, the first moment is also plotted as a function of flash state (0F, 1F, 2F, 3F) to check that the sample is functional and turning over. Fig. 4.3 illustrates the first moment trends collected across nine different experiments spanning several years, where variables such as sample preparation, detector, end-stations and personnel all vary.

Despite all of these variations, high levels of reproducibility are maintained, thus demonstrating an incredibly robust experimental protocol. The overall trend across experiments in Fig. 4.3 involves a dip by 0.1 eV between 0F and 2F, followed by a return between 2F and 3F. The progression of these trends can serve as a powerful indicator of whether or not the sample, or flash scheme, is corrupt.

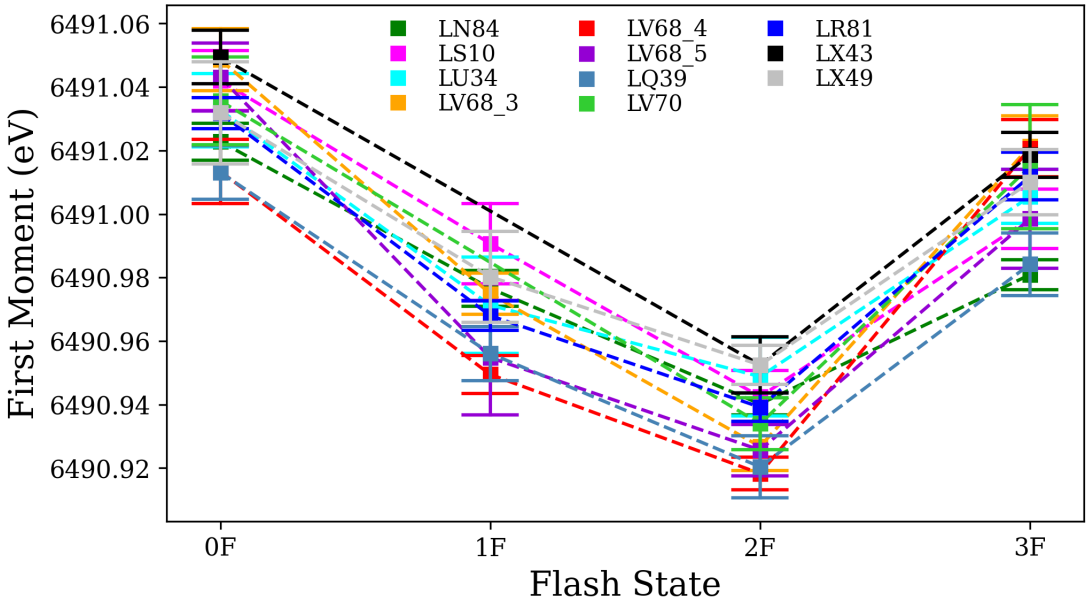


Figure 4.3: First moment trends from PS II collected for both crystal and solution, collected across nine different experiments using different experimental stations, samples, sample delivery systems, laser set-ups, detectors, and other parameters.

During a beamtime, XES can also be used to track other important parameters - such as the focus - by monitoring the spatial projection of the signal.

4.4 Studying Reaction Kinetics Through X-ray Emission Spectroscopy

The data collected from time-resolved XES experiments can be utilized to extract time constants for the study of reaction kinetics. This information can then be correlated with our structures to aid in telling a more complete story. This was done previously by Ibrahim et al. [118] in a study that combined SFX with XES at room temperature using an XFEL. From the data, the first moment (equation (4.2)) changes of Mn K$\beta_{1,3}$ spectra collected on PS II during the S_2 to S_3 transition appeared to be highly correlated with the insertion of an

additional water. Single exponential kinetics were assumed and the resulting time constant τ of the oxidation state was found to be around 350 μs.

In order to extract useful reaction kinetics, there are a few factors to consider. I will summarize them in this section, drawing upon many discussions with, and resources from, Thomas Fransson. The first thing to consider is the fact that the raw spectra (and the corresponding first moment values) are not pure. Instead, they are from mixed state data. As a result, one must take into account the various populations that are included within the Mn K$\beta_{1,3}$ spectral envelope. This results from a number of factors, which include the following:

1. Some amount of sample not getting affected by the flash (miss, p_m)
2. Some amount of sample getting affected by two flashes (cross-illumination, p_c) during data collection (see Fig. 4.4 for clarity)
3. The S_0 population in the dark (0F) state (p_{dark})

As a note, miss (p_m) and cross-illumination (p_c) are technically state specific. However, modeling them in this way has the potential to result in over-fitting. Therefore, for the work presented in this dissertation, they are assumed to take on state-independent values.

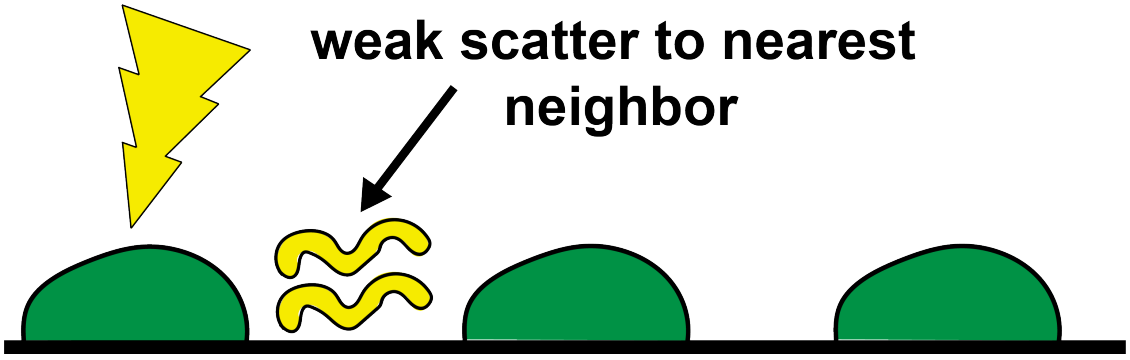


Figure 4.4: Schematic of "cross-illumination", which typically results from light scattering off a neighboring droplet which "sees" the laser first. It does not correspond to a droplet that gets a double hit.

p_{dark} can be used to estimate the meta-stable populations for 0F PS II data :

$$\overline{\rho}_{0\mathrm{F}} = \begin{pmatrix} \rho_{0\mathrm{F}}^{\mathrm{S}_1} \\ \rho_{0\mathrm{F}}^{\mathrm{S}_2} \\ \rho_{0\mathrm{F}}^{\mathrm{S}_3} \\ \rho_{0\mathrm{F}}^{\mathrm{S}_0} \end{pmatrix} = \begin{pmatrix} 1.0 - \rho_{\mathrm{dark}} \\ 0.0 \\ 0.0 \\ \rho_{\mathrm{dark}} \end{pmatrix}, \tag{4.3}$$

Next, equation (4.3) can be propagated to compute populations at the other flash states ($\overline{\rho}_{1\mathrm{F}}$ - $\overline{\rho}_{3\mathrm{F}}$) using the miss ($p_m$) and cross illumination (p_c) parameters:

$$\overline{\rho}_{nF} = \begin{pmatrix} \begin{bmatrix} \rho^{\mathrm{S}_1}_{(\mathrm{n-1})\mathrm{F}} \times p_m \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_0}_{(\mathrm{n-1})\mathrm{F}} \times (1.0 - p_m - p_c) \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_3}_{(\mathrm{n-1})\mathrm{F}} \times p_c \end{bmatrix} \\ \begin{bmatrix} \rho^{\mathrm{S}_2}_{(\mathrm{n-1})\mathrm{F}} \times p_m \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_1}_{(\mathrm{n-1})\mathrm{F}} \times (1.0 - p_m - p_c) \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_0}_{(\mathrm{n-1})\mathrm{F}} \times p_c \end{bmatrix} \\ \begin{bmatrix} \rho^{\mathrm{S}_3}_{(\mathrm{n-1})\mathrm{F}} \times p_m \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_2}_{(\mathrm{n-1})\mathrm{F}} \times (1.0 - p_m - p_c) \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_1}_{(\mathrm{n-1})\mathrm{F}} \times p_c \end{bmatrix} \\ \begin{bmatrix} \rho^{\mathrm{S}_0}_{(\mathrm{n-1})\mathrm{F}} \times p_m \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_3}_{(\mathrm{n-1})\mathrm{F}} \times (1.0 - p_m - p_c) \end{bmatrix} + \begin{bmatrix} \rho^{\mathrm{S}_2}_{(\mathrm{n-1})\mathrm{F}} \times p_c \end{bmatrix} \end{pmatrix}. \tag{4.4}$$

The relation between the measured first moment (μ_{nF}) and pure state first moment (μ_{S_n}) can be described by the following:

$$\begin{pmatrix} \mu_{0\mathrm{F}} \\ \mu_{1\mathrm{F}} \\ \mu_{2\mathrm{F}} \\ \mu_{3\mathrm{F}} \end{pmatrix} = \begin{pmatrix} \overline{\rho}^{\mathrm{T}}_{0\mathrm{F}} \\ \overline{\rho}^{\mathrm{T}}_{1\mathrm{F}} \\ \overline{\rho}^{\mathrm{T}}_{2\mathrm{F}} \\ \overline{\rho}^{\mathrm{T}}_{3\mathrm{F}} \end{pmatrix} \begin{pmatrix} \mu_{\mathrm{S}_1} \\ \mu_{\mathrm{S}_2} \\ \mu_{\mathrm{S}_3} \\ \mu_{\mathrm{S}_0} \end{pmatrix} \Longrightarrow \overline{\mu}_{\mathrm{F}} = \underline{\rho}\overline{\mu}_{\mathrm{S}} \tag{4.5}$$

Using the above parameters, which are unique and differ across sample and laser setup, one may deconvolute the mixed state data and extract pure-state (S_1,S_2,S_3,S_0) first moments. Pure state first moments μ_S can then computed as follows:

$$\overline{\mu}_{\mathrm{S}} = \underline{\rho}^{-1}\overline{\mu}_{\mathrm{F}}. \tag{4.6}$$

Fig. 4.5 compares the flash and pure state first moment trends from the same data presented in Fig. 4.3.

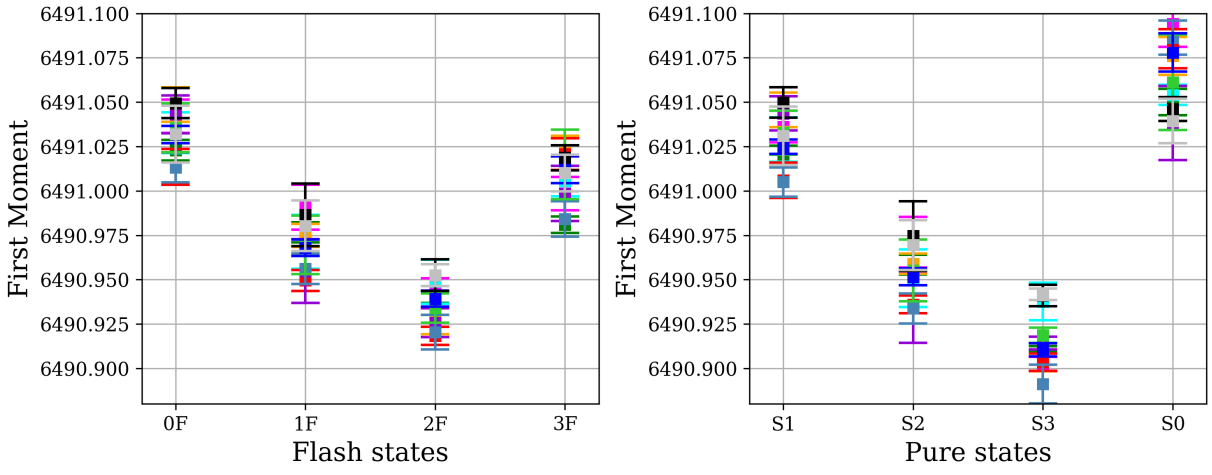


Figure 4.5: Comparing flash and pure state first moment trends from the experiments in Fig. 4.3

.

When carrying out kinetics modeling on PS II, we assume a sequential set of steps, resulting in a total sequence of events that can be modelled as follows:

$$S_1 \rightarrow S_1' \rightarrow S_2 \rightarrow S_2' \rightarrow S_2'' \rightarrow S_3 \rightarrow \cdots \rightarrow S_0 \rightarrow S_0' \rightarrow S_1 \tag{4.7}$$

S_n' designates the fraction of the population which is to propagate towards S_{n+1}, i.e. hits from S_n and cross-illumination from S_{n-1}. Each transition is assumed to occur with a simple decay model, with associated time constant τ.

For $S_1 \rightarrow S_2$, $S_2 \rightarrow S_3$, and $S_0 \rightarrow S_1$ we use simple models with direct transition ($S_1 \rightarrow S_2$ and $S_0 \rightarrow S_1$, with time constants τ_{12} and τ_{01}), and for $S_2 \rightarrow S_3$ we include a lag. This lag is modelled by including a step from S_2' to a lag phase S_2'', which has the same first moment as S_2. We can obtain time resolution of the population of pure states by first constructing a population vector for the time immediately after the flash (for example $\overline{\rho}_{0\mathrm{F}}(t=0)$ would be considering propagation from S_1 to S_2 after one flash). Next, a transition matrix $\underline{T}$ using the rates from proposed time constants $r_n = 1/\tau_n$ can also be constructed, and the populations are computed by propagating the transition matrix on the initial population

$$\overline{\rho}_{n\mathrm{F}}(t) = e^{-t\underline{T}}\overline{\rho}_{n\mathrm{F}}(0) \tag{4.8}$$

This gives us access to the populations as a function of time. The time-resolved first moment can be computed from the vector product between the population vector $\underline{\rho}$ and a vector with the first moments of corresponding pure states $\overline{\mu}_{\mathrm{S}}$.

Kinetics modeling can be an iterative process which is described by Fig. 4.6 and involves starting with a set of models, rates and sample parameters. Next we take our computed population trends and our observables, which in this case are our measured first moments, and calculate first moment trends from there. Typically a few rounds of this, involving further refinement of the model, are required.

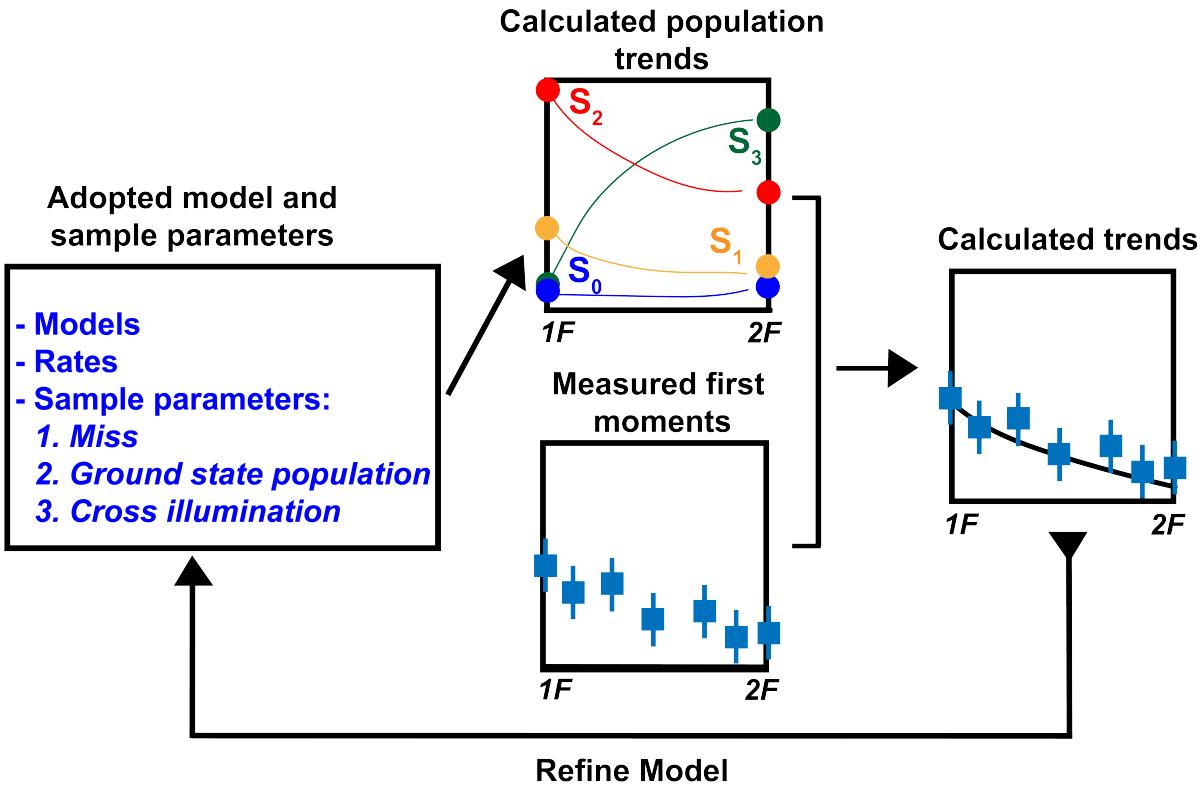


Figure 4.6: Procedure for calculating first moment trends.

At the end of the day, the chosen model will never provide the entire picture. For example, when using K$\beta_{1,3}$ first moments in the case of PS II, for example, manganese oxidation is the observable. This may be correlated with, but will not necessarily be equivalent to, the timescale of structural changes, so one must keep this in mind when interpreting the data. More details of how kinetics modeling has been applied to describes the reaction kinetics of PS II's catalytic center will be shown later on in this chapter.

4.5 Results

In this section, I will present results from room temperature XES experiments carried out at LCLS. I will start by describing the results from time-resolved measurements of manganese K$\beta_{1,3}$ XES collected along PS II's water oxidation reaction cycle. Subsequent kinetics modeling helps advance our understanding of the intermediate steps and associated time constants for all S-state transitions, with particular emphasis on the S_3 to S_0 step (**Section 4.5**).

4.5.1 Kinetics of Mn Redox Chemistry During Water Oxidation in Photosystem II Monitored Using an X-ray Free Electron Laser

The content and figures of this section are adapted from:

Thomas Fransson, Margaret D. Doyle, Philipp S. Simon, Mun Hon Cheah, Isabel Bogacz, Hiroki Makita, Ramzi Massad, Ruchira Chatterjee, Medhanjali Dasgupta, Rana Hussein, Mohamed Ibrahim, Julia Gätcke, James M. Glownia, Sanghoon Song, Clemens Weninger, Franklin D. Fuller, Asmit Bhowmick, Aaron S. Brewster, Nicholas K. Sauter, Dimosthenis Sokaras, Roberto Alonso-Mori, Uwe Bergmann, Athina Zouni, Johannes Messinger, Jan Kern, Vittal K. Yachandra, Junko Yano. "Kinetics of Mn redox chemistry during water oxidation in Photosystem II monitored using an X-ray free electron laser."

which is in preparation for publication.

4.5.1.1 Abstract

A complete understanding of the reaction mechanism of the photocatalytic cycle by which photosystem II (PS II) in plants, algae and cyanobacteria splits water and generates molecular oxygen has eluded much effort. This is because the final $S_3 \rightarrow [S_4] \rightarrow S_0$ step, where O-O bond formation and O_2 evolution occurs, has been highly challenging to follow with sufficient temporal resolution under operando conditions. The advent of X-ray free-electron lasers (XFELs), which enable X-ray spectroscopic and X-ray diffraction data collection at room temperatures, while outrunning radiation damage, has made it possible to follow such a process. Here, we employed Mn K$\beta_{1,3}$ X-ray emission spectroscopy (XES) to collect time-resolved data throughout the photocatalytic cycle to monitor the Mn redox changes. Through kinetic modeling, we extracted intermediate steps and associated time constants for all the S-state transitions, with particular emphasis on the $S_3 \rightarrow S_0$ step. When including the delay ('lag' phase) between photoexcitation in the S_3 state and the formation of S_4 into these kinetic models, only simulations with a transient reduced state ($Mn^{III}Mn^{IV}{}_3$), i.e. a peroxide-like intermediate between the S_4 and S_0 states described the XES data satisfactorily. Adding a short-lived Mn^V intermediate into the simulations, although possible, did not improve their quality, indicating that either such an intermediate is too short-lived to be resolved, or that the S_4 state features a substrate water oxidation intermediate, such as an oxygen radical.

4.5.1.2 Introduction

Embedded within the thylakoid membrane of plants, algae and cyanobacteria is an enzyme, Photosystem II (PS II), that catalyzes the light-induced oxidation of water to molecular oxygen (Fig. 4.7a) [120, 121, 122, 123]. The catalytic site for this reaction is the

oxygen-evolving complex (OEC) that consists of a Mn_4CaO_5 cluster and the surrounding protein and water-ligand environment. This cluster forms molecular oxygen through a 4-step sequence known as the Kok cycle (Fig. 4.7 b) [124]. Starting from the dark stable S_1 state, the S-state intermediates (S_2, S_3, and S_0 states) are reached by single turnover light flashes. Each flash creates a charge separation in the chlorophyll-containing reaction center of PS II that is connected to the OEC by a redox-active tyrosine side chain known as Y_Z [120, 122]. The formal oxidation states of Mn change from $Mn^{III}{}_3Mn^{IV}$ (S_0) to $Mn^{IV}{}_4$ (S_3), with a further oxidized species, the S_4 state, proposed to occur transiently before O-O bond formation and O_2 release during the S_3 to S_0 transition.

The kinetics of this reaction have long been studied through an array of complementary time-resolved methods, including electron paramagnetic resonance (EPR) spectroscopy [125], infrared (IR) spectroscopy [126, 127, 128], X-ray absorption spectroscopy (XAS) [129], and photothermal beam deflection (PBD) [130], see SI Table 1. Each technique monitors a specific process; for example, PBD measurements track expansion and contraction of the protein that are linked to charge variations at the quinone acceptor and around the OEC, Mn K-edge XAS specifically probes changes in oxidation state and ligand environment of the water-splitting catalyst, EPR measures the reduction kinetics of the $Y_Z^{\cdot}$, and IR spectroscopy detects changes in vibrational modes of specific chemical bonds around the Mn cluster.

Among the X-ray based techniques, inner-shell X-ray emission spectroscopy (XES) is useful for studying the evolution of transition metal electronic structures. It involves the creation of a 1s core hole via photoionization, which is rapidly ($\sim$fs) filled by a higher-lying electron, emitting excess energy through Auger decay or fluorescence. The fluorescent decay yields characteristic emission lines, and Mn K$\beta_{1,3}$ XES, which is the method used here, provides insight into the 3p and 3d orbitals of Mn, [40] and thereby about chemical properties such as spin and oxidation state, ligand type, structure, and covalency [119]. XES has also been used previously, together with crystallography, to probe oxidation states of the metal catalytic center in PS II in its (meta)stable intermediates [131, 118, 87, 110, 36].

In this study, we use an XFEL to collect transient Mn K$\beta_{1,3}$ XES at room temperature, as the catalytic reaction proceeds through the Kok-cycle. Special focus is given to the $S_3 \rightarrow S_0$ transition, where the O-O bond formation, oxygen evolution and resetting of the cluster to the S_0 state occurs. By kinetic modeling, we find that a fast, unresolved four-electron reduction from S_4 to S_0 cannot simultaneously describe the short and long time points in our data. By kinetic modeling, we find that a fast, unresolved four-electron reduction from S_4 to S_0 cannot simultaneously describe the short and long time points in our data. A satisfactory correspondence to the data was instead achieved by the inclusion of a reduced intermediate, probably a peroxide-like species.

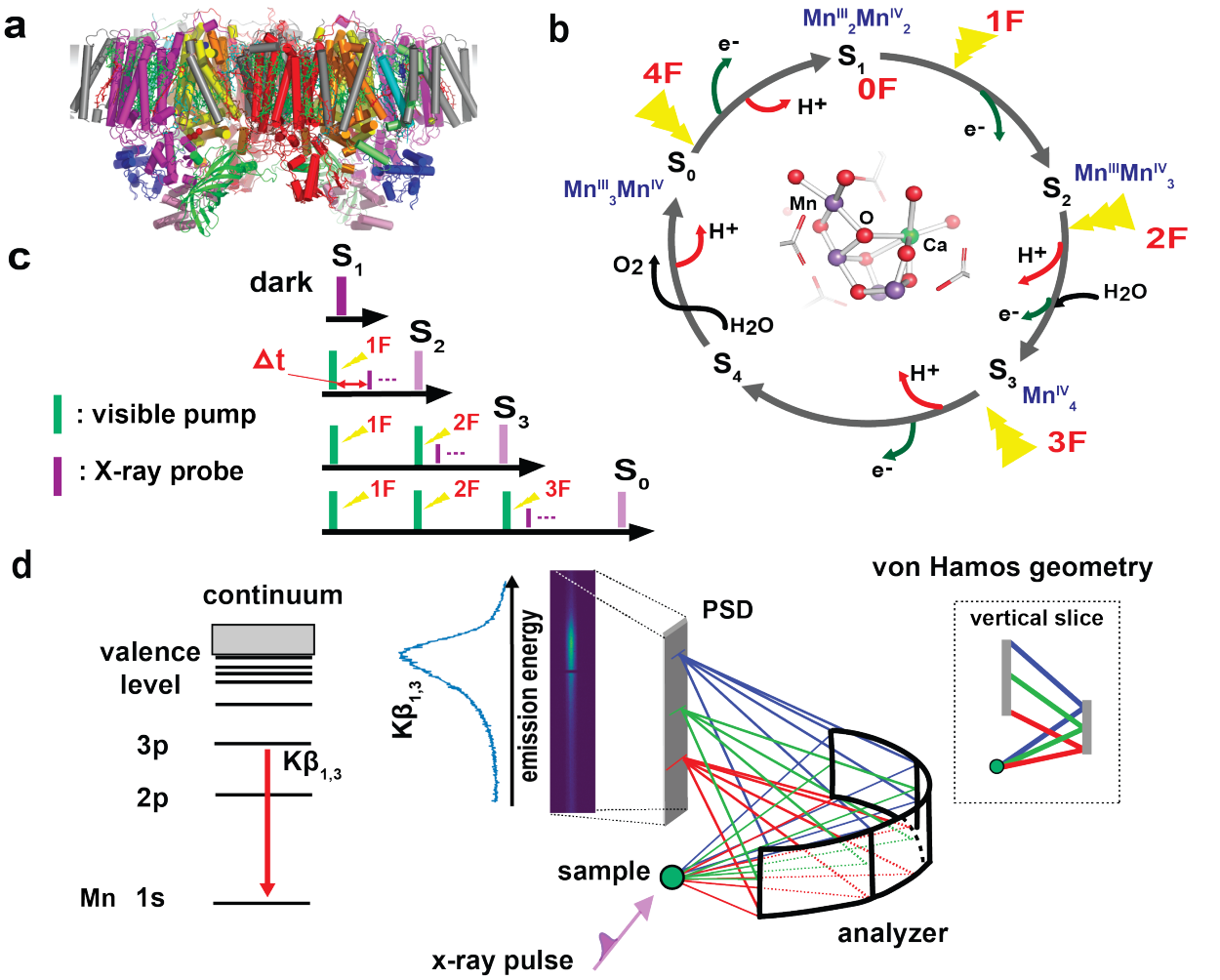


Figure 4.7: Schematic representation of the XES measurement setup at an XFEL. (a) Structure of PS II (pdb ID 6DHO). (b) The Mn_4CaO_5-O_x cluster (in S_3 state) and its reaction (Kok) cycle. (c) Schematic depiction of the pump-probe sequences for measuring time-points between S_i states. (d) K$\beta_{1,3}$ transition that gives rise to the XES emission spectrum measured by employing a von Hamos spectrometer and a position sensitive 2D detector (PSD). Data were obtained with *T. vestitus* PS II core complexes in solution at room temperature. See SI for details on the measurement conditions.

4.5.1.3 Results and Discussion

4.5.1.4 Mn K$\beta_{1,3}$ XES During the Reaction Cycle

Mn K$\beta_{1,3}$ XES spectra were collected using previously established protocols [87], with the drop-on-tape sample delivery system [110] and a von Hamos spectrometer [114], illustrated

in Fig. 4.7 c,d. In Fig. 4.8, Mn K$\beta_{1,3}$ XES spectra of PS II solution samples are shown along with their difference spectra and first moment changes at various time points during the reaction cycle (0F (dark state) to 3F (3 flashes)) (also see Fig. 4.13). We use the nomenclature nF(Δt) to describe time-resolved data collected during a transition. For example, 3F(1 ms) refers to data 1 ms following a third flash, in order to probe the $S_3 \rightarrow S_0$ transition process. If no delay time is given, the data is collected 200 ms after the flash. Fig. 4.8a shows the Mn K$\beta_{1,3}$ XES and corresponding difference spectra of the meta-stable S-states, as well as one example of a time point (3F(150 μs)) measurement. The first moments calculated from these Mn K$\beta_{1,3}$ spectra are employed as sensitive descriptors of the spectral shifts and plotted in Fig. 4.8b. The negative shift of the first moment from 0F to 1F and 2F to 3F represent the oxidation of Mn (Fig. 4.8b), while the positive shift from 2F to 3F indicates the reduction of Mn, demonstrating the good catalytic turnover in our experiments. The shift from the 0F to the 1F state corresponds mainly to the S_1 to S_2 transition (see e.g. lower panel of Fig. 4.11), and those of 1F to 2F and 2F to 3F correspond primarily to the S_2 to S_3 and the S_3 to S_0 transitions (note that the S_2 to S_3 transition data was discussed in Ref. [[118]]).

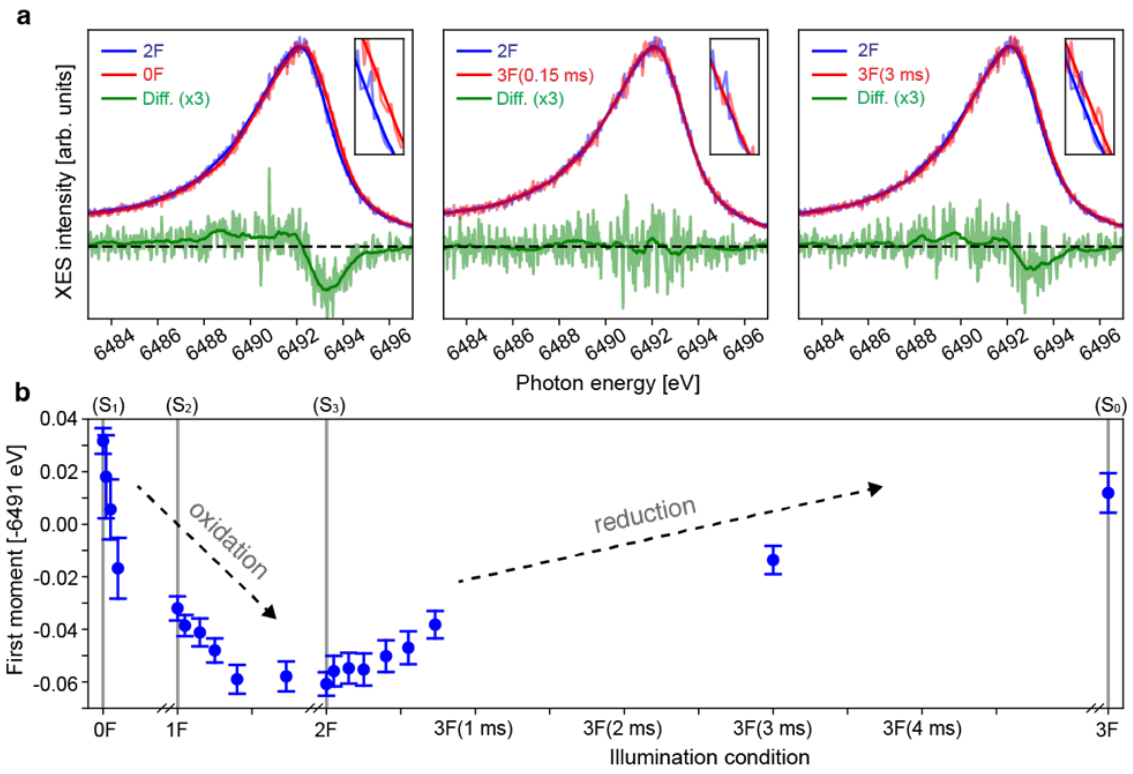


Figure 4.8: (a) Mn K$\beta_{1,3}$ XES of PS II in different flash states: raw spectra and smoothed spectra are shown. Data are area-normalized over the region from 6483 to 6497 eV. Insets show the zoom in view for 6493.5-6494.0 eV. (b) First moments of flash states (with prevalent S-state indicated). Error bars were calculated from bootstrap estimates. The first moment energy is offset by 6491 eV [87].

4.5.1.5 Kinetic Models

To understand the nature of the S_3 to S_0 transition from the Mn redox changes displayed in Fig. 4.8b, we hypothesized, based on the literature, eight possible reaction sequences (ordered by increasing complexity i.e. increasing number of intermediates): (i) the presence or absence of a delay ('lag' phase) between photoexcitation in the S_3 state and the formation of S_4, (ii) the nature of the last oxidation step of the OEC (S_4), i.e. whether Mn or substrate oxygen is oxidized, and (iii) two different options for the O-O bond formation and release of O2, namely an apparent single-step (4-electron reduction with no observable delay between O-O bond formation and release of O_2) or a two-step (including a peroxide-like intermediate) reduction of the catalytic center.

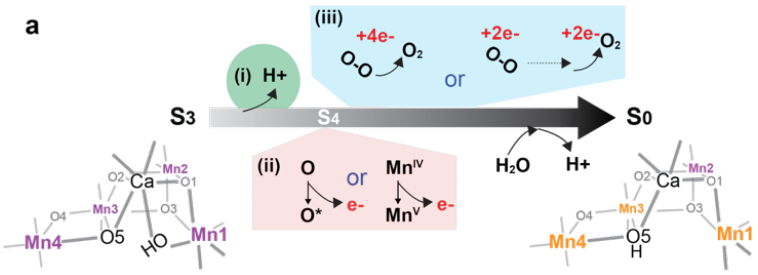
The presence of a lag phase after the 3rd flash has been described in the literature with a time-constant of 150 to 380 μs [127, 129, 132, 133, 134, 135]. This could occur if, for example, a rate-limiting proton release from the OEC happens prior to the final oxidation

event of the OEC ((i) in Fig. 4.9 a,b), [128] and is explored in models B, D, F and H.

The hypothetical S_4 state ((ii) in Fig. 4.9 a,b) represents the last oxidation step of the OEC. This may either involve oxidation of Mn^{IV} to Mn^{V}, or the oxidation of an oxygen (substrate) ligand to an O radical ($O^{\cdot}$). If the process involves Mn^{V}, it should be observable in XES, which is explored in models E through H ((ii-b) in Fig. 4.9b). However, if a ligand−centered oxidation occurs instead, the effective charge density on Mn likely remains close to that of the S_3 state (Mn4IV). Alternatively, such oxidized species (either Mn^{V} or $O^{\cdot}$) are highly reactive and thus never significantly populated at any time point rendering them non-observable.

After the final oxidation of the Mn_4CaO_5-O_x cluster to the S_4 state, O-O bond formation, water binding and the release of one proton and of molecular oxygen occur, and the most reduced S_0 state ($Mn_3{}^{III}Mn^{IV}$) is formed. The reduction of S_4 may happen in one step by a simultaneous transfer of 3 or 4 e^- from the substrate oxygens to Mn (options A, B, E, F, depending on the nature of the S_4 state, see above), or in two steps with a postulated observable $Mn^{III}Mn^{IV}{}_3$ state featuring a bound peroxide-like intermediate (options C, D, G, H). Fig. 4.9b shows the Mn redox steps that are anticipated in each model, which can be directly evaluated with the experimental XES data. More details on the kinetics modeling can be found in the SI.

a



b

Model	(i)* $Mn^{IV}_4Y_z$* lag phase	(ii-a)* S_4 not observable	(ii-b)* S_4 $Mn^{IV}_3Mn^{V}$	(iii) $Mn^{III}Mn^{IV}_3$ intermediate	S_0 = $Mn^{III}_3Mn^{IV}$	Mn oxidation state changes
A		x				
B		x				
C		x				
D		x				
E						
F						
G						
H						

Figure 4.9: Scenarios modeled in this work for the S_3 to S_0 transition. (a) Schematic depiction of the three variables in the eight scenarios considered in the modeling. (b) Tabulated breakdown of models used in this work, including Mn redox steps anticipated in each model.
* Indistinguishable from S_3 oxidation state (Mn^{IV}_4) by XES, only indicated if long lived due to H^+ release
** Two options: Mn oxidation states are Mn^{IV}_4-$O^{\cdot}$ and thus indistinguishable by XES from S_3 state (Mn^{IV}_4) and/or state is so short-lived that it does not accumulate

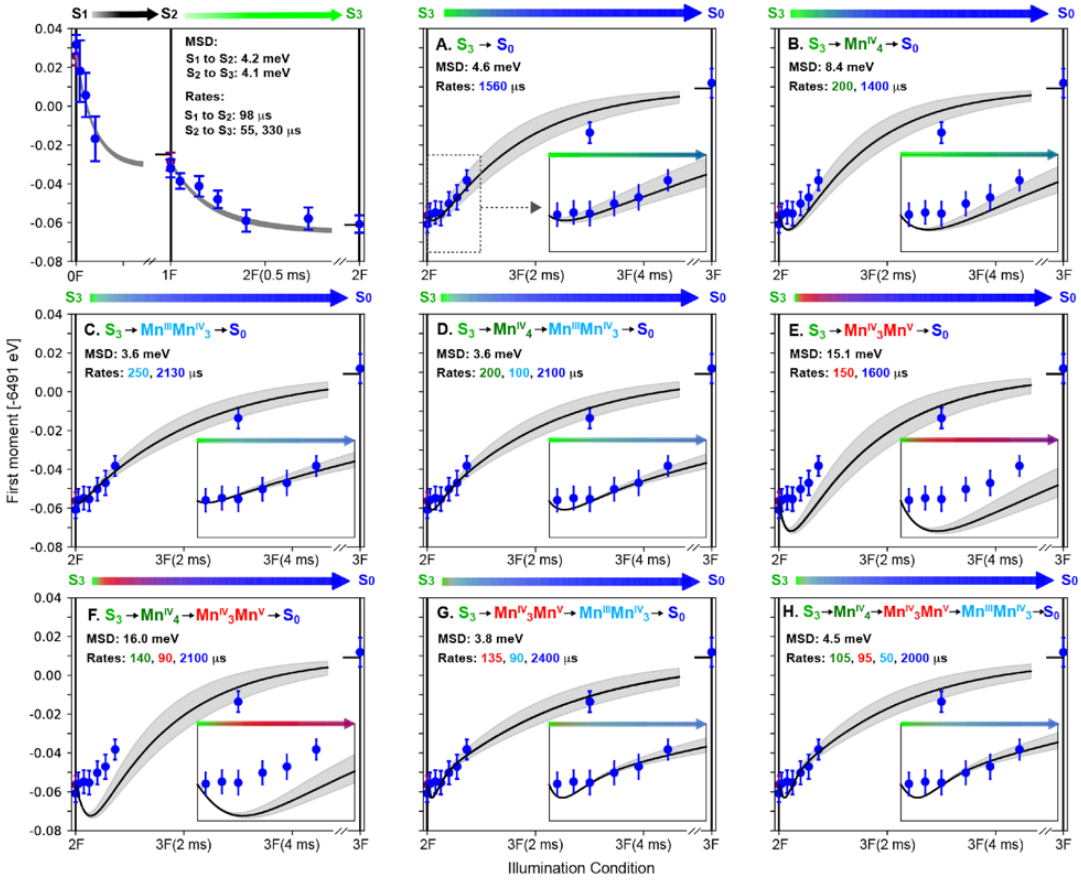


Figure 4.10: Calculated Mn XES first moment changes for the eight scenarios for the $S_3 \rightarrow S_0$ transition (A through H, solid lines) (see Fig. 4.9 b), compared to experimental data (blue markers), with small off-set in first moments at nF(0 ms) shown in dark purple (see the SI for details). Upper left panel shows the calculated trends of all models (overlapping gray lines) for $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$, and the remaining panels show $S_3 \rightarrow S_0$ for each model (A through H) individually. The $S_3 \rightarrow S_0$ panels show the oxidation states and include insets showing times up to 1 ms, and rates $\pm$ 25 % are shown as a gray area. Colored horizontal arrows illustrate time-resolved populations of intermediates – see Fig. 4.14 for a detailed explanation. Kinetic rates and mean-squared deviation (MSD) are provided for each model in the form of time constants (τ), which are related to the half-life ($\tau_{\frac{1}{2}}$) by $\tau_{\frac{1}{2}} = \ln(2)\tau$. Error bars were estimated for each data point by bootstrapping as described in the reference [87].

4.5.1.6 Evaluation of the Models

Fig. 4.10 shows the comparison of the Mn K$\beta_{1,3}$ XES first moment changes with the calculated results using the eight models, A through H, for the S_3 to S_0 transition described in Fig. 4.9, with estimated rate constants shown in the form of time constants in each panel.

The corresponding S-state population trends for the eight models are shown in Fig. 4.14.

As each flash state (0F to 3F) is a mixture of S-states in different ratios, the entire cycle was modeled simultaneously to extract the rate constant for each S-state transition. This includes the $S_0 \rightarrow S_1$ transition, which is not directly probed in this study. For details of the process, see the SI. From 0F to 1F, the $S_1 \rightarrow S_2$ transition is predominant and therefore modeling here is insensitive to the model chosen for the $S_3 \rightarrow S_0$ transition. The time constant of the $S_1 \rightarrow S_2$ transition we obtained is $\sim$98 μs, which is in line with the rate reported in the literature ($70 - 160$ μs; Table 4.1) [125, 128, 129, 135, 136].

Similarly, the 1F to 2F transition is dominated by the $S_2 \rightarrow S_3$ transition; it has been proposed that a proton release from the OEC precedes the electron release[130] in this process. In the current fitting procedure that considers all the flash states, we obtained values of $\sim$55 μs for the proton release and $\sim$330 μs for the following oxidation of one Mn (Mn^{III} to Mn^{IV}). These values are similar to the number we reported previously ($\sim$50 μs, and $\sim$300 μs) [118].

To obtain rate constants for the $S_3 \rightarrow S_0$ transition, a variety of plausible models and parameters were considered (see Figs. 4.14 - 4.18). For the 2F to 3F data, model A assumes a single-step model for the $S_3 \rightarrow S_0$ transition, i.e. the direct formation (no lag phase) of a short-lived, non-observable S_4 state, immediately followed by the O-O bond formation and release of molecular oxygen (3 or 4 e^- reduction of S_4). Addition of the lag phase (model B) made the model deviate more from the observed data. Similarly, the models that involve formation of an observable S_4 intermediate containing MnV (E and F) followed by the 3 or 4 e^- reduction lead to poor agreement with the experimental data at the earlier time points, largely independent of the rates selected (Fig. 4.15). It is clear that none of these models works satisfactorily. In contrast, models C, D, G, and H all yield reasonable trends. Models C and D account best for the early time points, while models G and H, which both contain an observable S_4 ($Mn^{IV}{}_3Mn^{V}$) state, were slightly less favorable (see Fig. 4.10). What models C, D, G and H all have in common is an observable reduced $Mn^{III}Mn^{IV}{}_3$ intermediate, suggesting the accumulation of a measurable quantity of a peroxide-like species on the way from S_4 to O_2 formation (see also Fig. 4.15).

The presence or absence of the early lag-phase is difficult to discriminate from the current XES data. However, our recent structural studies [131] as well as several reports in the literature [129, 132, 135, 137, 138, 139] suggest the presence of such a lag prior to the oxidation of the OEC. Altogether, the result suggests that the redox changes of Mn during the $S_3 \rightarrow S_0$ transition involve multiple steps. Therefore, we favor model D, which involves a lag phase followed by a non-observable S_4 state and a long-lived peroxide-like intermediate before O_2 formation. The estimated first moment trend and S-state population of model D are illustrated in Fig. 4.11.

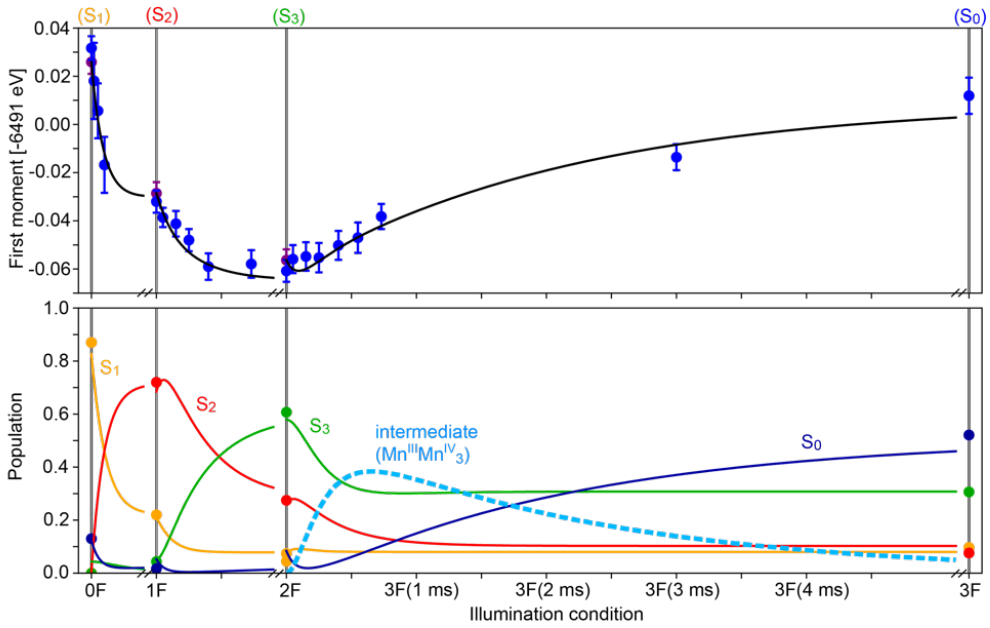


Figure 4.11: Estimated first moment and population trends using model D. Top: Calculated trends (black) compared to experimental measurements (blue), with small off-set in first moments at nF(0ns) shown in dark purple (see text for details). Bottom: Time-resolved populations, including the proposed reduced intermediate ($Mn^{III}Mn^{IV}_3$). The populations of intermediate $Mn^{III}Mn^{IV}_3$ for the $S_0 \rightarrow S_1$ and $S_2 \rightarrow S_3$ transitions have been removed, for clarity (see Fig. 4.14).

4.5.1.7 Reaction Sequence Leading to O_2 Formation

Several spectroscopic techniques have been deployed previously to capture reaction steps and intermediates during the water oxidation reaction in PS II (see above and Table 4.1). The kinetic data reported here are highly sensitive to Mn redox changes and were obtained with *T. vestitus* core complexes in solution at room temperature using Mn K$\beta_{1,3}$ XES as a probe.

Fig. 4.12 summarizes our proposed reaction scheme for the S_3 to S_0 transition based on the primary findings from this work, i.e. the Mn oxidation state changes during this transition. Additionally, it considers our recent structural study by Bhowmick et al. [131], which suggests a multistep S_3 to S_0 transition, with an initial lag between photoexcitation and S_4 formation and a delay between the O-O bond formation and the release of O_2. Starting from the S_3 state with a Mn^{IV}_4 oxidation state, there is a 200 μs lag phase (τ_1) (Fig. 4.10) which is linked to an initial proton release. Our data cannot identify the electronic state of a possible S_4 intermediate (Mn^{IV}_4-$O^{\cdot}$ or $Mn^{IV}_3Mn^{V}$), but exclude a significant accumulation of a state containing Mn^{V}. The rate constants into and out of the reduced $Mn^{III}Mn^{IV}_3$

intermediate are 100 (τ_2) and 2100 μs (τ_3), respectively. The location of the lone Mn^{III} in this reduced intermediate depends on the mechanism of O-O bond formation, which cannot be determined from our present data (a, b in Fig. 4.12). However, we interpret this reduced intermediate to feature a peroxide-like intermediate, as it is oxidized by two electrons compared to S_0 + O_2. It is important to note that we cannot exclude a state in which $Y_Z$ox is not yet reduced and a one electron-oxidized water intermediate is formed by electron donation from a substrate oxygen to one Mn^{IV}, thus forming the same $Mn^{III}Mn^{IV}{}_3$ oxidation states of the metal cluster (c in Fig 4.12) [140].

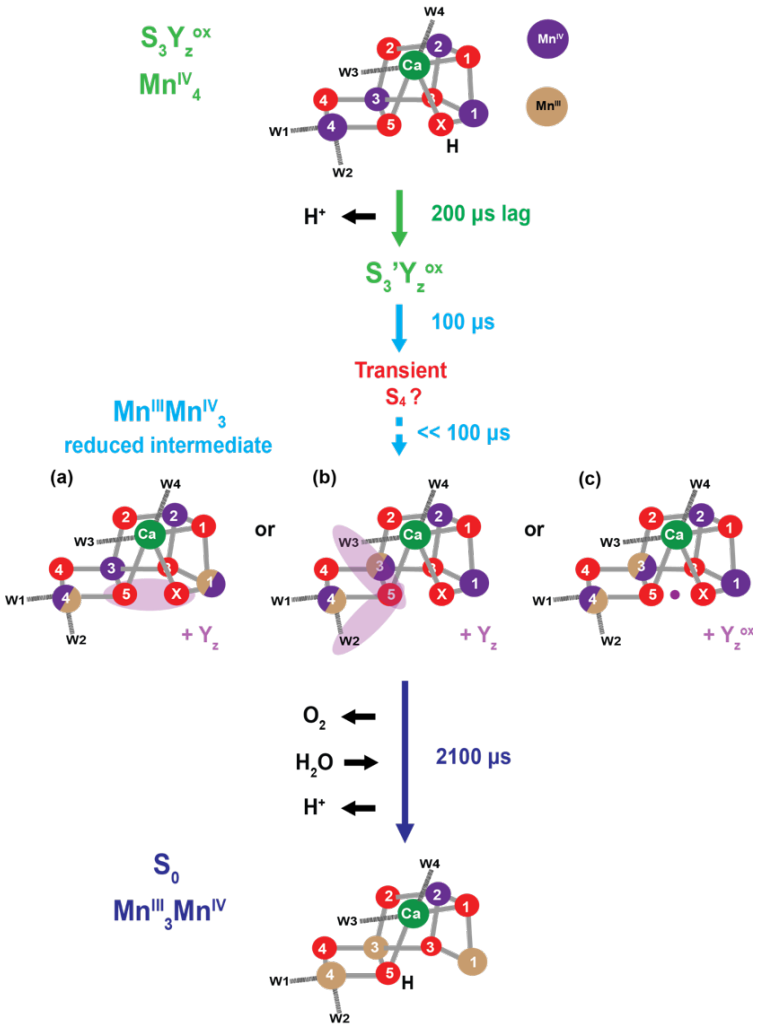


Figure 4.12: Proposed S_3 to S_0 chemical process based on the kinetics of the Mn redox changes, in the context of the structural data by Bhowmick et al [131]. The rate constants of the transition into and out of the peroxide-like intermediate, after a 200 μs lag (τ_1), are estimated as 100 (τ_2) and 2100 μs (τ_3), respectively. The proposed structural models of the $Mn4CaO_{5/6}$ cluster at each step and oxidation state of each Mn are shown. The location of the MnIII in the peroxide-like intermediate depends on the model for O-O bond formation.

4.5.1.8 Conclusions

Shot-by-shot data collection of metal XES at room temperature at XFELs is a powerful method for capturing the kinetic information of the metal redox states, which can be correlated with atomic structure information obtained similarly by crystallography [131, 118, 87, 110, 36]. Using Mn K$\beta_{1,3}$ XES, we tracked here the light-induced catalytic cycle of PS II. For the $S_0 \rightarrow S_1$, $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$ transitions, we obtained rates in good agreement

with other measurements in literature. For modeling of the $S_3 \rightarrow S_0$ transition, eight different scenarios have been tested. Importantly, our results provide experimental support for a two-step reduction mechanism of the Mn_4CaO_5-O_X cluster leading to O_2 formation with a detectable intermediate, most likely a bound peroxide. This is an important step forward towards understanding the chemistry of the photosynthetic water oxidation reaction.

4.5.2 Supplementary Information

4.5.2.1 Materials and Methods

Details on experimental setup and data analysis have been described in previous works, [118, 69, 87, 110]and we here summarize the most important aspects. Photosystem II (PS II) was extracted and purified from *T. vestitus*. Solution samples were in a buffer comprising, 0.1 M MES, pH 6.5, 5 mM calcium chloride and 0.005% (w/v) β-dodecyl maltoside (βDM) at 4.0-5.0 mM chlorophyll concentration (40-50 mg/ml protein). The solution samples were measured using a drop-on-tape setup [110], using up to three 527 nm ns light flashes from a fiber-coupled laser for advancing the sample from zero flash (0F, dark) to three flashes (3F), each spaced by 200 ms. An additional ns-pulsed free-space laser was used for measuring time-points, and details on the illumination conditions can be found elsewhere.[69] The Mn K$\beta_{1,3}$ X-ray emission spectra were measured using a 16-crystal von Hamos spectrometer [114, 38, 37] providing a dispersive line focus on an ePix100 2D detector[84] situated under or sideways from the interaction point (with exact position depending on the polarization of the XFEL beam). Data from different experimental runs at the Linac Coherent Light Source (LCLS) [141] at SLAC National Accelerator Laboratory was used for our study, conducted at the XPP[142] and MFX[143] beamlines for experimental runs R1 and R2, respectively. Aqueous solutions of $MnCl_2$ were used to calibrate the spectrometer setup.

The recorded data was corrected for detector artifacts (such as charge sharing[144]), and reduced by use of protocols described in Ref. [87].The corrected data was then sorted for sample hits, with a shot being considered a hit if the total number of photons inside the region-of-interest (ROI) exceed 3, and the ratio of photons per unit area inside versus outside the ROI was above 2.0. Previous studies suggest that such hit selection improve the signal-to-noise ratio, and introduce no discernible artifacts [87].

First moments were calculated on spectra splined to a resolution of 0.01 eV. The first moment measures the spectral center-of-mass and is thus primarily used to quantify small shifts in energy. We used an interval of 6485.5 - 6495.5 eV around the K$\beta_{1,3}$ peak, where the peak height is approximately equal. Absolute first moments and first moment shifts vary in the literature [145, 146, 147, 106, 148, 149] as a result of differences in energy calibration, background subtraction, spectrometer resolution and first moment integration range. However, within our experiments, the first moments and difference spectra computed across different beamtimes and sample types (solution or microcrystals) show consistency and establish the integrity of our protocol – see, e.g. Ref. [150] which demonstrates this stability across seven different experiments. First moment error bars were estimated using

photon statistics, using a previously established bootstrap method [87]. Other descriptors, such as first moment with variable energy intervals or integrated absolute differences, have been seen to provide similar information as the first moments used here [119].

Please note that we use the term fit for finding parameters which give weight to plausible or physically reasonable models, or models that have been discussed in the literature, rather than carrying out a pure, unconstrained, optimization. Models A and C were derived from constrained optimizations, while the other models were further adjusted manually, accounting for plausible rates and data from other experiments. The range of accepted rates for each model can be seen in the gray area included in Fig. 4.10. Furthermore, Figs. 4.16 - 4.18 demonstrate the relative stability of the models under fluctuating input parameters. As data sets from 1F and beyond include a mixture of all pure states, all transitions need to be considered simultaneously when fitting rates and models to experimental data. This has been done both by using models informed by previous studies and considering each transition at a time. The final values have been obtained by first modeling all rates simultaneously, and then regarding them in more detail one at a time. Model C was utilized for estimating rates of the other transitions, however the exact model chosen had very little impact. These rates were later applied to the other transitions of all models. This yields small differences for the estimated first moment trend, primarily affecting the earlier time-points when studying the $S_3 \rightarrow S_0$ transition. While earlier transitions are affected by the model used, the impact is quite weak (Fig. 4.18) and the background reaction during the $S_3 \rightarrow S_0$ transition is small and slow (Fig. 4.14); therefore, there is minimal impact from the $S_3 \rightarrow S_0$ transition.

4.5.2.2 Time Constants from Other Studies

Transition	Reference	Method	Time Constant (μs)	Year
$S_0 \rightarrow S_1$	Pace	EPR	58-87	1997
	Dau	XAS	43	2005
	Noguchi	TRIR	45 (H_2O) and 56 (D_2O)	1997
	Sauer	EPR	144	1975
	Van Gorkum	UV/vis	4.3	1993
	Lavergne	Optical Absorbance	361	1994
	Van Gorkum	Optical Absorbance	43	1984
	Haumann	XAS	52 (H_2O) and 66 (D_2O)	2016
	Dau	PBD	40 + 100	2012
$S_1 \rightarrow S_2$	Pace	EPR	123	1997
	Dau	XAS	101	2005
	Noguchi	TRIR	74 (both H_2O and D_2O)	2018
	Sauer	EPR	144	1975
	Van Gorkom	Optical Absorbance	137	1993
	Lavergne	Optical Absorbance	72	1994
	Van Gorkom	Optical Absorbance	159	1984
	Haumann	XAS	89 (H_2O) and 108 (D_2O)	2016
	Noguchi	TRIR	12 + 72 (H_2O) and 16 + 85 (D_2O)	2017
	Dau	PBD	100	2012
$S_2 \rightarrow S_3$	Pace	EPR	202	1997
	Dau	XAS	274	2005
	Van Gorkom	Optical Absorbance	548	1993
	Lavergne	Optical Absorbance	418	1994
	Van Gorkom	Optical Absorbance	505	1984
	Haumann	XAS	26 + 317 (H_2O) and 117 + 568 (D_2O)	2016
	Noguchi	TRIR	13 + 104 + 352 (H_2O) and 14 + 122 + 663 (D_2O)	2017
	Dau	PBD	30 + 300	2012
$S_3 \rightarrow S_0$	Pace	EPR	1082	1997
	Dau	XAS	361 + 1587	2005
	Sauer	EPR	1443	1975
	Van Gorkom	Optical Absorbance	6492	1993
	Lavergne	Optical Absorbance	1731	1994
	Van Gorkom	Optical Absorbance	1876	1984
	Haumann	XAS	153 + 1538 (H_2O) and 380 + 2208 (D_2O)	2016
	Dau	PBD	200 + 1600	2012

Table 4.1: Time constants reported from other studies in the literature. Note that time constants are used, not half−lifes. Also note that exact sample conditions vary between experiments, which will also have an effect on estimated time constants.

4.5.2.3 Kinetics Modeling

The kinetics of the photocatalytic cycle is modeled as a set of steps for each transition, with each pair coupled with an exponential transition function. These transitions are initiated by having a fraction of the metastable state population moved to an intermediate (decay) population, and then propagated in time. Prior to modeling the kinetics, we estimated the steady S-state population of the sample in each flash state (0F, 1F, 2F, and 3F),

using estimates from a combination of XES, MIMS, and EPR. The first moments for the metastable S-states are then calculated by deconvoluting the static flash state first moments (0F to 3F).

Here, there are three parameters that influence the meta-stable populations: 1) the S_0 population in the dark state (0F), 2) the probability of a miss, and 3) the probability of cross-illumination. The S_0 population in 0F (S_1) is generally around 10%, and depends primarily on sample and dark adaptation conditions prior to the XES data collection. The miss parameter is the fraction of the sample which does not progress to the next step. This is generally expected to be dependent on the transition which is to be undertaken (see Ref. [151]), but is here estimated as the same value for each step (typically around 0.1-0.2) in order to avoid over-fitting. The cross-illumination parameter is a result of a fraction of the sample proceeding to the second-next transition due to the double hit, as opposed to the next transition. Since the laser pulses used to progress the sample are ultrashort (<10 ns), this is not caused by the sample being hit twice by the same laser flash, but it results from light scattering off a neighboring droplet which sees the laser first. This cross-illumination is usually negligible in the DOT setup, but it can contribute when using a small droplet spacing, as was done for experiment run 1 (R1). The estimated sample turnover parameters for the two experimental runs considered in this study are provided in Table 4.2.

Experiment	Miss	S_0 in dark	Cross-illumination
R1	0.13	0.13	0.050
R2	0.10	0.10	0.015

Table 4.2: Sample turnover parameters for the included experiments.

While we do not have any measurements on the first moment of the (hypothetical) S_4 state, we adopt a value 0.025 eV lower in energy than for S_3. This corresponds to approximately 30% of the shift between S_1 and S_3, and is here designated as $Mn^{IV}{}_3MnV$. The impact of larger (and smaller) shifts are illustrated in the Fig. 4.15, and it can be noted that the (non)linear relation between the formal oxidation state and the first moment has been discussed, and it has been suggested that formal oxidation state is not sufficient to understand the detailed electronic structure changes.[152]

In this study, lag phases are modeled as a transition from one population (S_n) to one with the same first moment ($S_{n'}$). We stress that this second intermediate can have a different atomic structure, but the manganese oxidation state is (approximately) unchanged.

The comparatively large diameter of the spot size of the fiber laser, which is employed for illuminating the droplet near the x-ray intersection during time point measurements, leads to a weak excitation of the subsequent droplet. As a result of this 'cross-illumination',

there is a small difference between the metastable nF first moments and those of nF(0ms). The estimated first moments of nF(0 ms) are inserted as dark purple circles and lines, and we note that the modeled first moments start there, rather than at nF. This also affects populations, as visible in the bottom panel of Fig. 4.11 and in Fig. 4.14.

4.5.2.4 Results from Different Experiments

Sample turnover parameters for the considered experiments are shown in Table S_2, with resulting measured and modeled first moments shown in Fig. 4.13. The calculated first moments use model D for $S_3 \rightarrow S_0$, and rates from Fig. 4.10. The cross-illumination parameter is highest for R1, as this experiment was run at the highest repetition rate, necessitating the smallest droplet spacing.

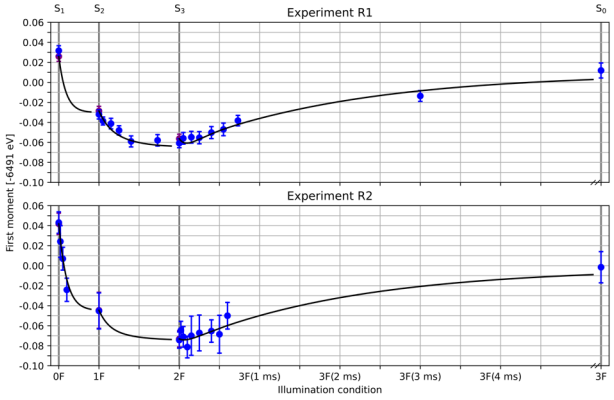


Figure 4.13: Measured and modeled first moment trends for experiments R1-R2, using model D with rates from Fig. 4.10. Each nF corresponds to data collected at 200 ms after n flashes, such that the sample is predominantly in the intermediate S-state reached by this time (S_1, S_2, S_3, and S_0, for n=0,1,2,3).

4.5.2.5 Population Trends of Models

In Fig. 4.14 the population trends of each model are illustrated, using rates from Fig. 4.10. Note a small off-set between the populations at the meta-stable time-points, as compared to at t=0 for a transition - this is due to the influence of cross-illuminations, as discussed above.

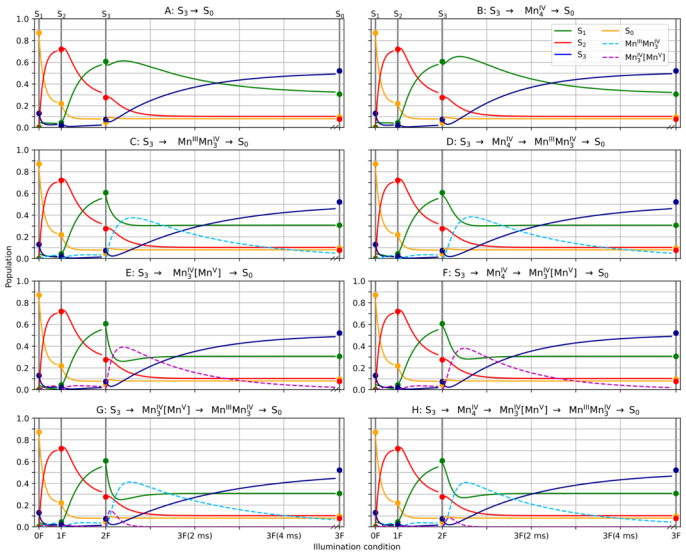


Figure 4.14: Population trends of the different models, using rates from Fig. 4.10.

4.5.2.6 Sensitivity to Parameter Choices

Figs. 4.15 - 4.17 illustrate the variability in modeled first moments when varying different parameters. In Fig. 4.15 the first moment of a proposed [S_x] intermediate is varied, using models with only one such intermediate (top), and [Sx] followed by a partially reduced intermediate (bottom). The first moment is varied around that of S_3, ranging from 0.05 eV lower (oxidized by more than one step) and 0.10 eV higher (reduced by more than two steps). These intermediates thus span scenarios ranging from further oxidized (S_4 state with MnIV), to reduced by two electrons (similar to S_1). Using only one intermediate, we see a large spread in the estimated first moments, with reasonable agreement to measured values only for the unchanged or slightly reduced intermediates. This is also the case for models including a lag phase, where this phase only delays the onset of the increase in the first moments. Using an [S_x] intermediate followed by a partial reduction, the variations are smaller, but generally favors [S_x] with first moments around S_3.

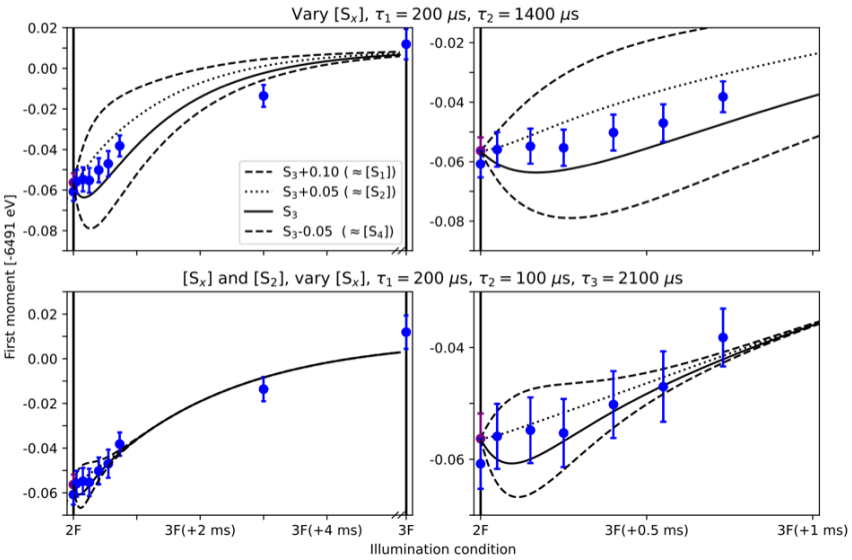


Figure 4.15: Using $[S_x]$ intermediates with different first moments for $[S_x]$ (top), or this intermediate followed by partial reduction (bottom). Varying $[S_x]$ around $[S_3]$, with rates from models B and D.

In Fig. 4.16 we illustrate the influence of variations in $S_3 \rightarrow S_0$ rates (using model D), including changing rate ratios (top two panels, for which the sum of rates are preserved) and varying the short or long rates by ±25%. The impact of simultaneously increasing or decreasing the rates by ±25% is seen in the gray area of Fig. 4.16. We see that varying rate ratios between the short- and long-lived intermediates has impact in particular for short times, but has limited impact when considering the short-lived intermediates in isolation. There is furthermore a very small difference between τ_1=200 and τ_2=100 as compared to τ_1=100 and τ_2=200, as is seen by the overlapping full and dashed lines in the corresponding panels. Varying early rates yield relatively little difference to the modeled first moments, while varying $S_3 \rightarrow S_0$ rates leads to large variations.

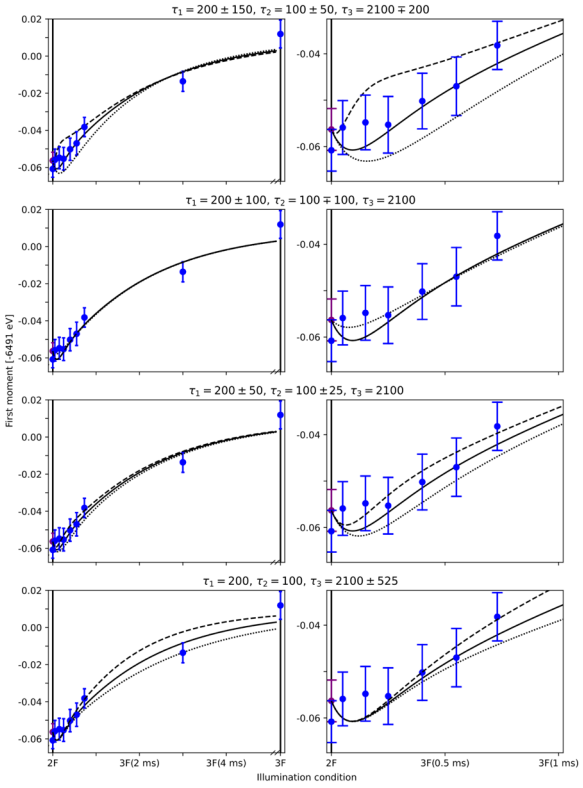


Figure 4.16: Varying rates for model D, including varying the ratio between long- and short-lived intermediates (first panel), the ratio between the two short-lived intermediates (second panel), and varying the total short (third panel) and long (fourth panel) rates by ±25%. Increase in (first) rate shown as dotted line, decrease as dashed line.

The impact of variation in sample turnover parameters is shown in Fig. 4.17 (using model D), with impact on both long and short times. Variations of misses and S_0 population primarily affect shorter times, while variations in cross-illumination impact both shorter and longer times. For the cross-illumination parameter, there are noticeable shifts in initial first moments, due to the impact of cross-illumination on the nF(0ms) estimates (see discussion above).

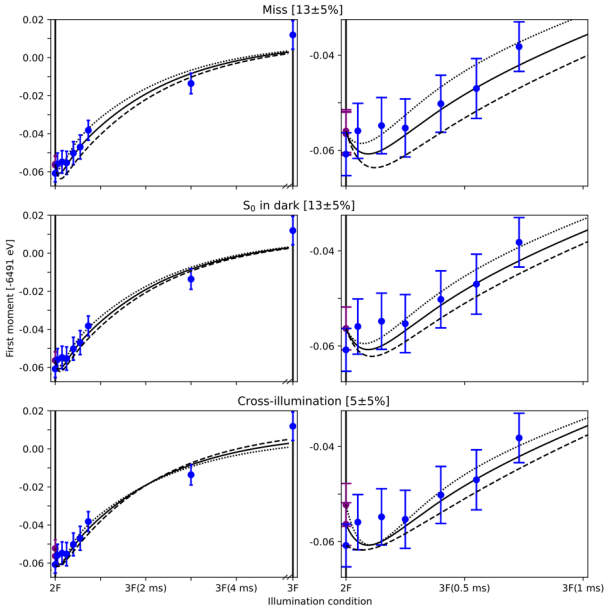


Figure 4.17: Modeled first moments for model D, varying rates and sample turnover parameters. Increasing and decreasing the parameters, with results shown as dotted and dashed lines, respectively.

Finally, Fig. 4.18 shows the stability of the models over simultaneous variations in several different parameters. The impact of varying sample turnover parameters, rates, and metastable first moments are shown for models A, D, and H, illustrating the resulting mean first moment trends and standard deviations. The first moments are calculated by sampling a Gaussian distribution for each parameter, with mean and standard deviation as follows:

Sample parameters

- Misses: 13 ± 5 %
- S_0 in dark: 13 ± 5 %
- Cross-illumination:5 ± 5 %
- Rates: adopted ± 25 %

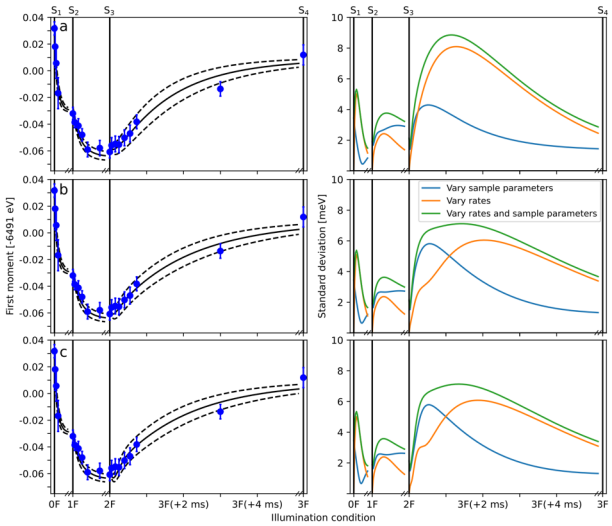


Figure 4.18: Sensitivity of models A (a), D (b), and H (c). Calculated from 10 000 iterations, varying sample turnover parameters and rates (left), as well as standard deviations for all combinations (right).

The first moment spread is largest for S_3 to S_0, with different variations impacting the spread for different time regions. Changing sample turnover parameters primarily impacts the calculated first moments for shorter times, and has some of the smallest impact. Varying rates give larger spreads for intermediate times, and varying both yields rapidly increasing uncertainty for short times, followed by a relatively flat uncertainty at $\sim$ 0.5-3.0 ms, which then decreases. Importantly, we note that the uncertainties are not additive, and stay below 7 meV for variations in sample parameters and rates for model D. It is clear that the spreads are larger for model A, and the observed first moments are clearly off-set from the estimates. For models D and H we obtain similar trends.

4.6 Outlook

When carrying out conventional XES at an XFEL, we are only ever able capture a fraction of the isotropic emission with the von Hamos spectrometer. We also are typically working with low concentration samples (mM range), so signal levels can be quite small. Conventional XES also comes with a relatively large core-hole lifetime broadening and small spectral sensitivity to electronic structure changes in the stronger lines, such as K$\alpha_{1,2}$ and K$\beta_{1,3}$. One way around this is to measure weaker valence-to-core lines, which also are much easier to model theoretically [153]. However, the emission cross sections for these lines are quite low, requiring high flux, which is not always feasible.

The problems listed above are a reality for all low concentration metalloenzyme XES experiments carried out at XFELs. As a result, collecting interpretable data can be quite

time consuming. If beam flux is too low, these limitations cannot even be overcome by collecting longer, and one walks away from an experiment with unusable data.

In **Chapter 5**, I will describe a completely new method of collecting XES data which involves using the stimulated XES method [102]. This approach is analogous to the stimulated Raman method in the optical regime and has the potential to collect XES data highly effectively. It also may be used for probing the less intense spectral transitions that have more information about the surrounding ligand environment.

Chapter 5

Enhancing Chemical Sensitivity With Seeded Stimulated Emission

5.1 Introduction

The previous chapter focused on the use of time-resolved XES at XFELs in order to enable kinetics studies and in-situ sample quality monitoring. However, conventional XES at XFELs has limitations including, but not limited to, the following:

1. The relatively large core-hole lifetime broadening, causing weak spectral sensitivity [154].
2. Only a fraction of the sample's spontaneous XES signal gets collected in shot-by-shot XFEL measurements due to the small solid angle of the analyzer crystals (a few % of 4π [37]).
3. The metalloenzymes we study typically contain a low concentration ($<$1 mM) of the transition metal-of-interest. In addition to this, typically just one metal ion is changing oxidation state at a time; however, the collected spectrum is an average over all metal ions. These facts put together result in a relatively small observable change. This will be exacerbated if the charge is fairly delocalized.

Taken together, all of this means that our spontaneous X-ray emission data is often low in signal, requires long collection times and is sometimes difficult to interpret. This final point is unfortunate because underneath a broad spectral envelope lies a treasure-trove of information (Fig. 5.1). The dependence of the Kα and Kβ XES spectral shape of 3d transition metal complexes is given by the spin state as well as the covalent mixing strength [155, 119, 156]. The experimental spectrum is resulting from a number of mainline transitions between the photo-ionized core-hole intermediate states and all final states [155], and their individual intensities. Points #1-3, which are laid out above, make resolving these informative features quite difficult.

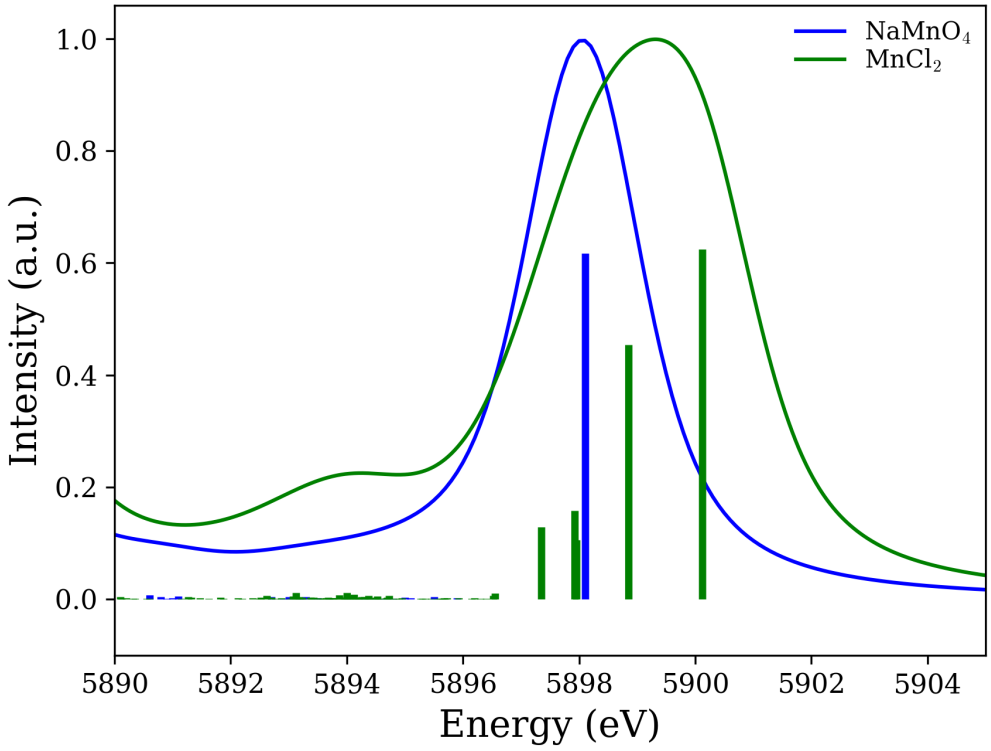


Figure 5.1: Simulations of spontaneous Kα emission from $MnCl_2$ and $NaMnO_4$, including transition lines, which build each spectrum, shown as vertical sticks.

In this chapter I introduce a new approach to collecting XES data which attempts to resolved the aforementioned issues by moving from the linear regime, where the photon–matter interaction considers one photon at a time, to the nonlinear regime [42]. In the nonlinear regime, the photon output is no longer proportional to the photon input, and the high peak power from an XFEL is used.

This approach is called Amplified Spontaneous Emission (ASE) or stimulated X-ray emission spectroscopy (S-XES) and can result in significant gain and enhanced chemical sensitivity. In order to obtain ASE, a short ($<$ 10 fs) and highly focused ($\sim$ nm) beam which is tuned above the K-edge of the metal-of-interest, is directed into the sample. This pump must be intense enough to create a population inversion of 1s core-hole excited states along the pulse direction. Inevitably, the excited states will decay after a short time ($\sim$ fs), resulting initially in photons which are spontaneously emitted in all directions. The few emitted along the forward direction will stimulate the emission of additional photons at the same wavelength and in the same direction. This produces a signal that grows exponentially, also known as an atomic laser around 5898 eV driven by an XFEL. Most emitted photons are

confined to a small solid angle and easily collected by a flat analyzer crystal in the forward direction, resulting in more efficient data collection than in the spontaneous case.

In the soft X-ray regime, this type of atomic inner shell lasing has been carried out on neon [157] and silicon [158] and in the hard X-ray regime it has been carried out on copper [159] and manganese [160, 154]. The latter is most relevant for this dissertation, due to the manganese cluster in PS II.

Earlier stimulated emission experiments on manganese complexes tested whether chemical sensitivity was preserved at such high power densities. For this work, $NaMnO_4$ ($3d^0$) and $MnCl_2$ ($3d^5$) were chosen since they represent end points of manganese oxidation and thus exhibit large chemical shifts. Work by Thomas Kroll et al. [154] demonstrated that through stimulated emission spectroscopy in transition metal complexes, chemical shifts of Kα lines appear to be preserved at extreme peak powers (10^{20} W/cm^2). Through these experiments, Kroll et al. also observed an exponential increase in the number of photons generated and Kα line widths which could vary to a point where they were narrower than the 1s core hole lifetime broadening.

These findings have been reproduced at subsequent experiments carried out at different experimental facilities (See Fig. 5.2) and have established a framework for utilizing nonlinear X-ray spectroscopy to enhance spectral characteristics in transition metal complexes.

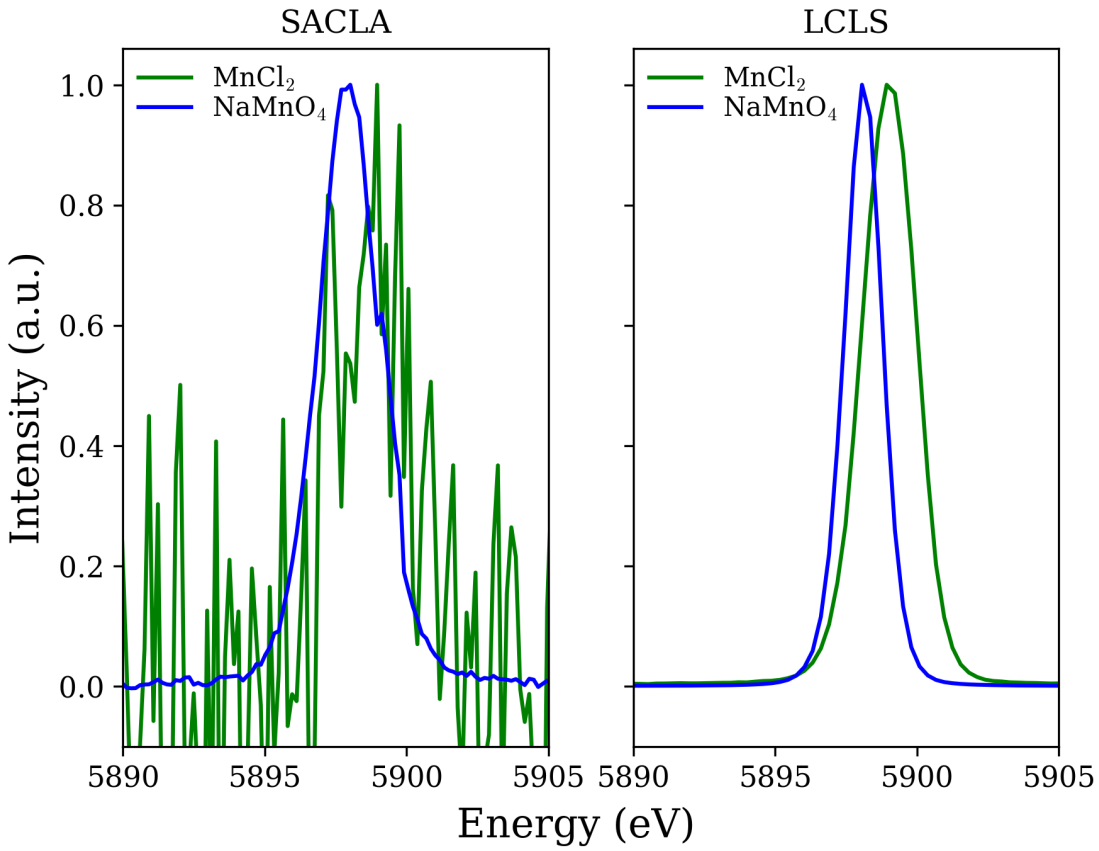


Figure 5.2: The characteristic chemical shifts observed between ASE spectra collected from $MnCl_2$ and $NaMnO_4$ across two different facilities.

Due to the strong nonlinear nature of the stimulated emission method, spectra collected without a second color seed (see Fig. 5.2) consist of mainly one strong peak due to the fact that the strongest transition gets amplified before all others. While valuable, the information about the full electronic structure is therefore limited. This limitation could be overcome by tuning a second pulse, a seed pulse, to the energy of a specific transition (see Fig. 5.1) for selection and amplification. This is called seeded stimulated emission or SSE. SSE has the potential to unlock a selection of weaker but more chemically sensitive decay channels and has implications for studying less intense spectral transitions that harbor richer chemical information regarding the ligands bound to transition metals.

This technique has been demonstrated previously [160, 159]. However, both of these experiments used SASE seed pulses which are highly fluctuating, leading to an input seed pulse spectrum that consists of a number of narrower peaks, changing shot-to-shot, within a total bandwidth of about 40 eV [159].

The research presented in this chapter, however, demonstrates the advantages of using a

monochromatic and tunable seed pulses generated by SACLA's two color mode [61]. This allows for enhanced selectivity of individual spectral features, which are characteristic of electronic structure, hidden under the broad spectral envelope (see Fig. 5.1).

By scanning the seed pulse photon energy through the spectral envelope, we can potentially select out and seed spectral features. Additionally, a seed pulse is advantageous because its use facilitates a lower pump pulse intensity (see **Section 5.2**). Deploying the lowest possible pump pulse intensity is generally preferred, as this facilitates rapid gain media replacement and reduces X-ray-induced electronic structure changes.

The experiments described in this chapter are a crucial step towards developing S-XES into a powerful tool that one day could potentially be applied towards studies on PS II and other related systems. I will first describe the required experimental parameters for carrying out a seeded S-XES experiment at SACLA on manganese solutions (see **Section 5.2**), in addition to describing details surrounding the two-color mode utilized (see **Section 5.3**). Next, I will describe how seeded S-XES can be utilized to experimentally determine the transitions that build up XES spectra (see **Section 5.4**). I will conclude by describing future applications of this work, which has the potential to turn S-XES and related nonlinear spectroscopies into common tools deployed in XFEL-based science (see **Section 5.5**).

5.2 Seeded Stimulated X-ray Emission at 5.9 keV

The content and figures of this section are adapted from:

Margaret D. Doyle, Aliaksei Halavanau, Yu Zhang Yurina Michine, Joshua Everts, Franklin Fuller, Roberto Alonso-Mori, Makina Yabashi, Ichiro Inoue, Taito Osaka, Jumpei Yamada, Yuichi Inubushi, Toru Hara, Jan Kern, Junko Yano, Vittal K. Yachandra, Nina Rohringer, Hitoki Yoneda, Thomas Kroll, Claudio Pellegrini, Uwe Bergmann "Seeded stimulated X-ray emission at 5.9 keV," Optica 10.4 (2023): 513-519.

It can be accessed at the following link:
https://doi.org/10.1364/OPTICA.485989

5.2.1 Abstract

X-ray free-electron lasers (XFELs) provide intense pulses that can generate stimulated X-ray emission, a phenomenon that has been observed and studied in materials ranging from neon to copper. Two schemes have been employed: amplified spontaneous emission (ASE) and seeded stimulated emission (SSE), where a second color XFEL pulse provides the seed. Both phenomena are currently explored for new coherent X-ray laser sources and spectroscopy. Here we report measurements of ASE and SSE of the 5.9 keV Mn Kα_1 fluorescence line from a 3.9 molar $NaMnO_4$ solution, pumped with 7 femtosecond FWHM XFEL pulses at 6.6 keV. We observe ASE at a pump pulse intensity of 1.7 $\times 10^{19}$ W/cm^2, consis-

tent with earlier findings. We observe SSE at dramatically reduced pump pulse intensities down to 1.1 $\times 10^{17}$ W/cm^2. These intensities are well within the range of many existing XFEL instruments. This supports the experimental feasibility of SSE as a tool for generating coherent X-ray pulses, new spectroscopic studies of transition metal complexes, and other applications.

5.2.2 Introduction

The advent of the X-ray Free-Electron Laser (XFEL) [161] has made possible the exploration of many femtosecond-Angstrom scale physics phenomena previously unobservable [72, 162]. One such phenomenon is stimulated X-ray emission, first observed in the form of inner-shell lasing for the neon Kα line at 850 eV[157] and then for the copper Kα line at 8 keV [159]. Stimulated emission can occur by amplified spontaneous emission (ASE) using an XFEL pump pulse [157, 154, 158] and seeded stimulated emission (SSE) using a two-color XFEL pump and seed pulse [159, 160]. The collectively emitted photons, e.g., at the Ne, Mn, Cu Kα_1 lines, can exhibit high levels of coherence and spectral resolution [154, 160, 163]. Several applications of ASE and SSE are currently under investigation, including novel X-ray spectroscopy probes with potentially enhanced chemical sensitivity [154, 160] and generating new coherent X-ray sources with properties beyond those provided XFEL pulses [157, 164, 159, 78, 163].

XFEL pulses are based on self-amplified spontaneous emission (SASE) [165], which results in limited coherence, temporal and spectral spikes [166], and large intensity fluctuations. Techniques to improve the pulse properties include self-seeding [167, 168, 169, 170, 171], fresh slice/bunch seeding [172, 173], and the conventional approach of frequency-filtering the pulses with a monochromator.

While externally seeded FELs can produce nearly fully coherent, e.g. see Ref [[174]] and stable pulses, and there are proposed schemes for increasing stability of FELs [175], high-gain, single-pass sources like XFELs have inherent stochastic fluctuations which limits the generation of stable, fully coherent, transform-limited pulses in the hard X-ray regime. This requires a multi-pass cavity-based system as, for example, an XFEL oscillator[176, 177].Stimulated X-ray emission is at the heart of a different multi-pass cavity-based approach for generating intense, stable, transform-limited, femtosecond X-ray pulses. The proposed population inversion X-ray laser oscillator (XLO) [78] uses a two-stage process where the XFEL pulse generates stimulated Kα X-ray emission in a gain medium. The XLO operates with a train of intense XFEL pulses and uses an X-ray optical Bragg cavity to recirculate the stimulated X-ray emission pulses. Critical parameters of the XLO are the pump and seed pulse intensities and the amplification factor (gain) obtained at each cavity passage. Provided that one can achieve adequate temporal and spatial overlap of pump and seed pulses, the gain is primarily dependent on the pump and seed pulse intensity, the gain medium concentration and thickness, and the number of excited atoms. These parameters are also critical when considering stimulated X-ray emission for spectroscopy applications on transition metal complexes [154, 160] or for the generation of phase-stable femtosecond

X-ray pulse pairs [163]. The general aim is to obtain stimulated X-ray emission at the lowest possible pump pulse intensity, as this facilitates the rapid gain medium replacement required for an XLO (a cavity length of ∼10 m corresponds to a pulse spacing of ∼33 ns). A lower pump pulse intensity also reduces X-ray-induced electronic structure changes for spectroscopy applications[178, 150].

In this paper we report a systematic study that establishes the parameters to minimize the pump pulse intensity for stimulated X-ray emission. This is accomplished by splitting the XFEL undulator in two parts, one tuned to the pump pulse photon energy, the other to the seed pulse photon energy. We measured stimulated X-ray emission at the Mn Kα line in a 3.9 molar $NaMnO_4$ solution where we varied the pulse parameters over more than two orders of magnitude. Calculations of pump and seed pulse intensities based on the measured signals are provided and our results can be straightforwardly extrapolated to other emission lines, various gain medium and sample concentrations and thicknesses, as well as XFEL parameters. They provide a benchmark for various applications of the exciting phenomenon of stimulated X-ray emission.

5.2.3 Experimental Setup

Experiments were performed at the SACLA XFEL facility at Experimental Hutch 5 (EH5), (see schematics in Fig. 5.3 A). The SACLA XFEL was configured in two-color mode[179, 180]. The first 7 undulators were used to generate the seed pulses at the 5.9 keV Mn Kα_1 photon energy. These SASE pulses were filtered through a Si (111) channel cut Bragg crystal, to provide monochromatic seed pulses with < 1 eV FWHM. The exact number changes slightly on a shot-by-shot basis depending on beam conditions upstream of the monochromator. The 14 downstream undulators produced the 7 fs FWHM SASE pump pulses at 6.6 keV, above the Mn K edge[181, 182, 183]. The two-color pulses were focused with Kirkpatrick-Baez (KB) mirrors[183, 184, 185] to 120 nm (horizontal) by 180 nm (vertical) FWHM spot size, as determined by knife-edge scans. The beam divergence downstream of the focus was 3.8 mrad (horizontal) and 1.9 mrad (vertical).

The gain medium was a concentrated sodium permanganate, $NaMnO_4$, solution (3.9 molar Mn) delivered in a liquid jet, with 100-micrometer diameter, located inside a helium-filled vacuum chamber. The nozzle and the recirculating pump system for the jet was designed at SACLA [186]. The estimated pump pulse energy at the gain medium is 40 μJ, corresponding to an intensity of 1.7 $\times 10^{19}$ W/cm^2, when the gain medium is fully in focus. For a conceptual figure behind how both pump and seed pulses participate in the SSE process in the $NaMnO_4$ solution, see Fig. 5.3 B.

The seed pulse spectra were analyzed with a 2% sampling grating followed by a dispersive spectrometer consisting of an elliptical mirror, Si (111) analyzer crystal, and a 2D MPCCD detector located in Experimental Hutch 1 (EH1)[187], referred to as the Seed Spectrometer (see Fig. 5.3 A). The ASE and SSE spectra were analyzed with a flat Si (220) analyzer crystal reflecting the beam in the horizontal plane onto a second 2D MPCCD detector [163], the Kα Spectrometer (see Fig. 5.3 A). For measurements where the jet is fully in focus along

the beam direction and a seed is used, a 25 μm thick Aluminum foil was placed in front of the Kα Spectrometer. With this configuration, we were able to record the seed pulse and the SSE spectra simultaneously, shot-by-shot. All relevant transmission factors from the experimental setup are detailed in Table 5.1.

Component	Thickness	Transmission (%)
KB Be windows	200 μm	91
KB reflectivity	N/A	80
Kapton windows	150 μm	56
Aluminum foil	25 μm	46
X-ray transport losses	N/A	45
Air path after gain medium	35 cm	36
Si (220) analyzer	N/A	1*

Table 5.1: Relevant experimental components and their corresponding transmission factors at 5.9 keV - used to correct for the number of photons at the gain medium. The only losses relevant to the Seed Spectrometer come from the X-ray transport losses and KB reflectivity. The remaining elements in the table are used to correct the $K\alpha$ Spectrometer.

*Resulting from the ratio of horizontal beam divergence (3.8 mrad) and Si (220) Darwin width (27 μrad).

5.2.4 Establishing the Pump and Seed Pulse Parameters for SSE

Using the same formula as described by Fransson et al. (Table 1 (e) in Fransson et al. [150]), we estimate pump and seed pulse intensities, as shown in Table 5.2. The pump pulse energy at the gain medium is estimated to be 40 μJ, as measured by a calibrated inline intensity monitor[184]. For the seed pulse, we base our estimates on the photon counts in regions (a) and (b) of the average seed pulse spectrum shown in Fig. 5.5 , (the photon count in the latter suggests the estimated seed pulse energy at the gain medium to be $\sim$ 0.01 μJ) and the assumption that the beam size of the seed pulse is the same for the cases of a tuned and detuned monochromator. To obtain the intensity, we only consider the photons that fall within the FWHM of the elliptical Gaussian focus, which is 50% of the total[150]. We then divide this by the pulse duration (7 fs)[188] and area (A) under this focus (FWHM) A = xy $\times$ $\pi/4$, where x and y are the horizontal and vertical beamsize at the gain medium.

We first placed the jet in focus at the maximum pump pulse intensity I_{pump}= 1.7 $\times 10^{19}$ W/cm^2 to generate Mn Kα_1 amplified spontaneous emission (ASE) (without a seed pulse).

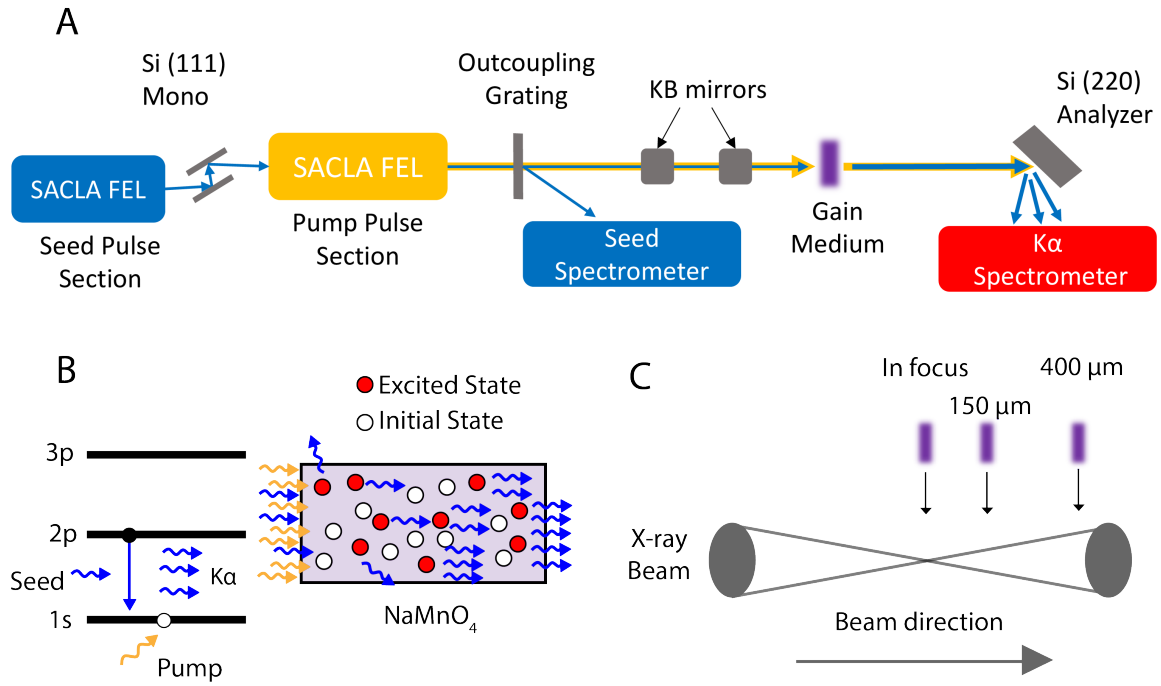


Figure 5.3: (a) Experimental layout (not to scale): SACLA XFEL radiation containing the seed and pump pulses is focused with a KB mirror system onto a liquid jet of $NaMnO_4$. 2% of the seed pulse is outcoupled and registered in the Seed Spectrometer. The resulting stimulated emission is dispersed onto the $K\alpha$ Spectrometer with a Si (220) crystal analyzer. (b) The level diagram and concept behind seeded stimulated emission (SSE). The pump pulse (yellow), which is tuned above the Mn K edge, creates a population inverted 1s core hole excited state. The seed pulse (blue) is tuned to the $K\alpha$ emission line in order to seed the SSE signal (red). (c) Schematic showing how the pump and seed pulse intensity is varied by moving the gain medium to various distances out of focus along the beam direction.

The corresponding 2D spectrum is shown in Fig. 5.4 A, in which the spectral direction is shown along the horizontal axis and the spatial direction is shown along the vertical axis. Moving the gain medium 150 μm out of focus reduces the pump to $I_{pump} = 1.1 \times 10^{18}$ W/cm^2 and results in the loss of the ASE signal. This establishes that the lowest pump pulse intensity to observe ASE from a 3.9 molar $NaMnO_4$ solution is between these two values, at some factor times 10^{18} W/cm^2. We estimate that the minimum pump pulse intensity for observing ASE from the most concentrated Mn gain medium, such as a 25 μm Mn foil, is a factor of 8 lower, corresponding to four times smaller thickness and 33 times larger concentration (78.9 Mn atoms/nm^3 for the foil, 2.4 Mn atoms/nm^3 in solution). This estimate assumes (i) that the Raleigh length of the KB mirrors is the same or more than the size of our jet diameter (100 μm), and (ii) that the onset of stimulated emission occurs when the concentration times pump pulse intensity reaches a certain value (e.g. twice the concentration requires half the

pump pulse intensity), which is an approximation.

After establishing the pump pulse intensity limit of 1.1 $\times 10^{18}$ W/cm^2 where no ASE is observed, we added the monochromatic seed pulse and scanned its photon energy through the Mn Kα_1 line. We started at δE = 7 eV above the Mn Kα_1 line, where we already observed SSE (Fig. 5.4 B). At this detuning value only a small fraction of the seed pulse intensity spectrally overlaps with the peak of the Mn Kα_1 line (see next section for detailed discussion). Fig.5.4 C shows the heat map of the complete seed pulse scan, where each horizontal line corresponds to the seed pulse photon energy as it is swept through the Mn Kα_1 spectrum. As the seed moves closer to the center of the Mn Kα_1 line, there is no strong change in the spectral shape and intensity of the SSE signal (see Fig. 5.4 E for the corresponding series of SSE spectra). This indicates that this series of SSE was close to the limit of the gain curve, where a stronger seed pulse does not result in a stronger SSE signal.

Distance out of focus (μm)	x (nm) FWHM	y (nm) FWHM	I_{pump} (W / cm^2)	I_{seed} (W / cm^2) (δE = 0 eV)	I_{seed} (W / cm^2) (δE = 7 eV)
0	120	180	1.7×10^{19}	3.2×10^{15}	7.9×10^{12}
150	690	465	1.1×10^{18}	2.2×10^{14}	5.3×10^{11}
300	1260	750	3.8×10^{17}	7.4×10^{13}	1.8×10^{11}
400	1640	940	2.4×10^{17}	4.5×10^{13}	1.1×10^{11}
500	2020	1130	1.6×10^{17}	3.1×10^{13}	7.4×10^{10}
600	2400	1320	1.1×10^{17}	2.2×10^{13}	5.4×10^{10}

Table 5.2: Distance from nanofocus, horizontal (x) and vertical (y) beamsize at the gain medium, calculated by assuming the 3.8 mrad (horizontal) and 1.9 mrad (vertical) beam divergence downstream of the focus, and the corresponding pump (I_{pump}) and seed (I_{seed}) pulses intensities at the gain medium.

Our next step was to align the seed pulse photon energy with the center of the Mn Kα_1 line (see Fig. 5.4 D) and reduce the pump pulse intensity by moving the jet progressively out of focus (downstream) along the beam direction (see Fig. 5.3 C). As the pump and seed pulses share the same beam line optics and focus, doing so also simultaneously changes the seed pulse intensity. The highest number of Kα_1 photons observed from a single shot of SSE , using the correction coefficients given in Table 5.1, is 2.6 $\times 10^8$, using a pump pulse intensity of 1.1$\times 10^{18}$ W/cm^2 (corresponding to 150 μm out of focus - see Table 5.2) and a seed pulse intensity of 7.6 $\times 10^{14}$ W/cm^2. For this shot, the seed pulse was detuned by δE = -0.5 eV below Kα_1.

An important quantity is the seeding efficiency defined as the photon ratio ρ=$N_{K\alpha_1}/N_{seed}$ where each N corresponds to the number of photons that fall within the measured Kα_1

FWHM of that shot. The experimentally obtainable value of ρ depends on the exact spatio-temporal overlap of pump and seed and is difficult to estimate for this experiment without more advanced shot-by-shot diagnostics. $N_{K\alpha_1}$ was determined by subtracting N_{seed} (from first detector, see Fig. 5.3 A) from the total photon count registered by the Kα Spectrometer in the same bandwidth. The strongest single shot enhancement of Kα_1 photons with respect to seed photons within the Kα_1 FWHM was $\rho = 4.0 \times 10^3$. For this single shot of SSE, the pump and seed pulse intensities were 1.1×10^{18} W/cm^2 and 4.3×10^{11} W/cm^2, respectively. The low I_{seed} comes from the fact that, for this shot, the seed was detuned by $\delta E = 5.5$ eV below the Kα_1 so that only 1.6×10^4 seed photons were estimated to be found within the Kα_1 FWHM. Nevertheless, 6.5×10^7 Kα_1 photons were observed.

After moving to a pump pulse as low as $I_{pump} = 1.1 \times 10^{17}$ W/cm^2, a seed pulse intensity of $I_{seed} = 7.8 \times 10^{12}$ W/cm^2 was required to observe a single shot of SSE with $\rho = 3.9$, resulting in an estimated 1.1×10^7 photons in the Kα_1 FWHM. This pump pulse intensity is comparable with or even an order of magnitude lower than typical values used in experiments at XFELs using standard μm focusing optics [150].

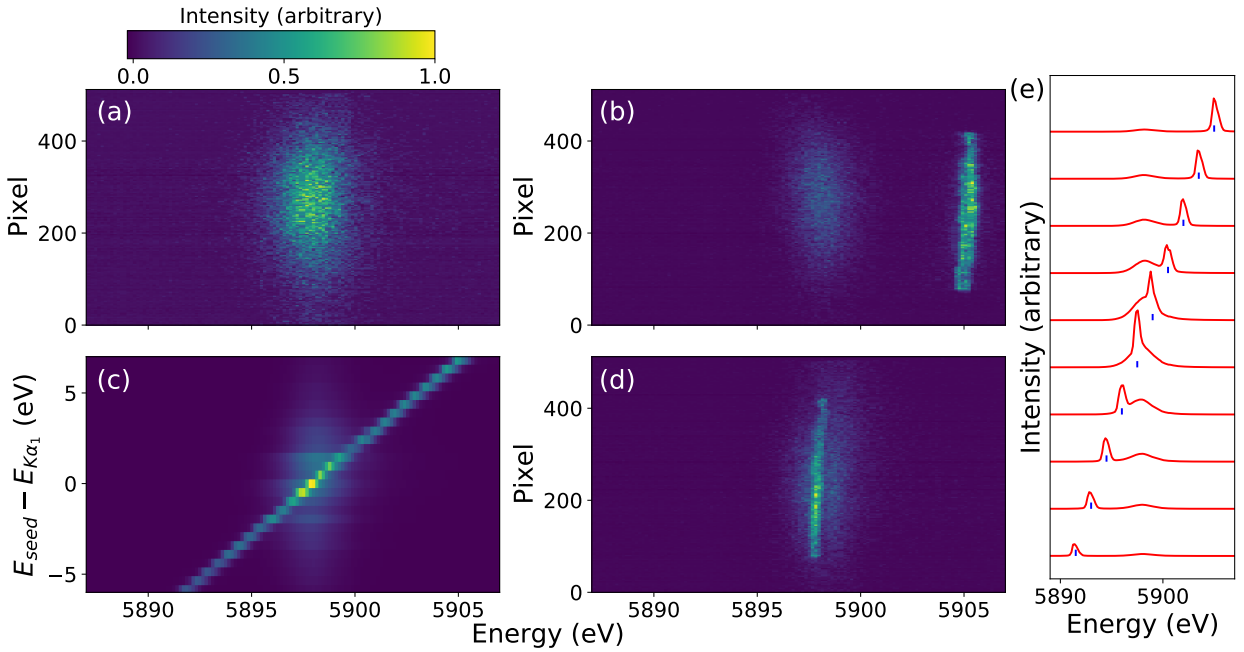


Figure 5.4: (a) Selected single-shot ASE spectrum collected at a pump intensity of 1.7×10^{19} W/cm^2; single-shot SSE spectrum (center) and seed pulse (right vertical line) at (b) δE = 7 eV above $K\alpha_1$ and (d) $\delta E = 0$ eV at a pump intensity of 1.1×10^{18} W/cm^2; (c) detuning scan of $K\alpha_1$ SSE at a pump intensity of 1.1×10^{18} W/cm^2, in which each horizontal row corresponds to the average spectrum for a given seed pulse energy collected across 1924 shots and each seed is indicated by a square box on the diagonal; (e) 1D version of (c), showing $K\alpha_1$ SSE spectra (solid red lines), in which the center of the seed pulse energy is indicated by a blue vertical line. Each 2D spectrum shown was collected at the $K\alpha$ spectrometer, and is peak normalized. Each 1D spectrum in (e) is normalized to the peak pixel value from (c).

We now discuss varying the seed pulse intensity at a constant pump pulse intensity. This can be accomplished by detuning the seed pulse photon energy from the Kα_1 peak. Figure 5.5 shows an average seed pulse spectrum collected across 1924 single shots, placed on a log plot and corrected for the number of photons at the gain medium. For illustrative purposes, two spectral regions with the same bandwidth as the spontaneous $NaMnO_4$ Kα_1 linewidth, $\Delta E_{K\alpha_1}$ = 2.1 eV FWHM [154], are selected for comparing photon counts when the seed is tuned directly on top of Kα_1 (see Fig. 5.4 D) and detuned by δE = 7 eV from Kα_1 (see Fig. 5.4 B for a sample spectrum). Due to the seed's spectral shape, which contains some Lorentzian character, the total photon count in $\Delta E_{K\alpha_1}$ remains above 10^4 photons. It is even true that in cases where the seed is detuned by δE=7 eV, the reduced number of photons (by more than a factor of 400) are still sufficient to stimulate emission.

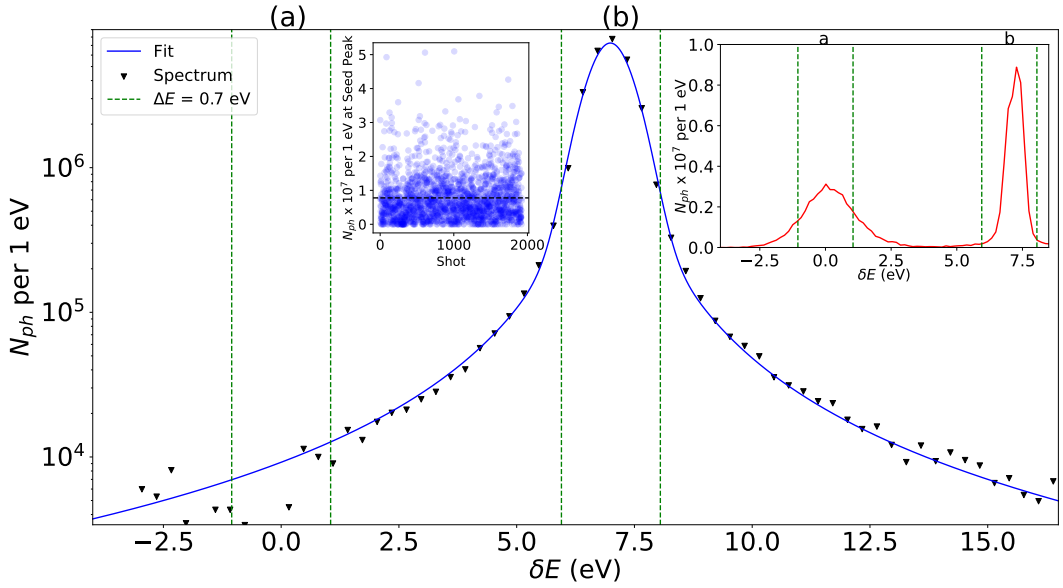


Figure 5.5: Seed pulse spectrum obtained from averaging 1924 single shots, and the corresponding best fit from equation (5.1). In the inset to the left, the shot-by-shot seed intensity, which exhibits 100% fluctuation and has an estimated average value of 7.8×10^6 (black broken line) is shown. The average spectrum and corresponding fit were shifted to $\delta E = 7$ eV from $K\alpha_1$ for demonstration purposes. Next, two regions were selected from the shifted fit: $\delta E = 0$ eV (a) and $\delta E = 7$ eV (b). In region a), 2.0×10^4 photons were observed within $\Delta E_{K\alpha_1} = 2.1$ eV FWHM. In region b), 8.2×10^6 photons were found in a bandwidth of $\Delta E_{K\alpha_1} = 2.1$ eV FWHM. Integrated photon counts were computed from the fit. In the inset to the right, the single shot spectrum corresponding to Fig. 5.4 B is shown to illustrate how, even in the case of a detuned seed, enough photons from the low-energy tail are able to stimulate emission.

5.2.5 Spectral Analysis and Seeding Efficiency

To analyze the spectral characteristics of both the seed pulse and SSE, we projected each detector image onto the energy axis and fit it with a Pseudo-Voigt function

$$I(E, f) = I_0 \left[(1-\eta)G(E, f) + \eta L(E, f)\right] \tag{5.1}$$

where $G(E, f) = \frac{2.355}{f\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{2.355\cdot E}{f})^2}$ is the Gaussian component, $L(E, f) = \frac{1}{2\pi}\frac{f}{((E^2+(f/2)^2))}$ is the Lorentzian component, E is the photon energy, f is the total full width at half maximum (FWHM) and $0 < \eta < 1$ is the Lorentzian parameter[189].

Seed pulse spectra collected at both spectrometers were cross-correlated. Total beamline transmission (from the undulator hall to the target) was found to be 45%. A two-detector setup is necessary for estimating the amount of seed pulse photons in the Kα_1 line spectral bandwidth, especially in case of no detuning. When the seed pulse is detuned, one, in principle, can extract the number of seed pulse photons from the spectral fits using only the Kα Spectrometer. It has an intrinsic spectral resolution of 0.24 eV FWHM corresponding to the Darwin width of the Si 220 analyzer crystal. As this is much narrower than most of our spectral features, we do not attempt to deconvolute, and all reported numbers include this broadening. Note that Figs. 5.6 - 5.7 show results derived from fitting results with an R-squared value > 0.99.

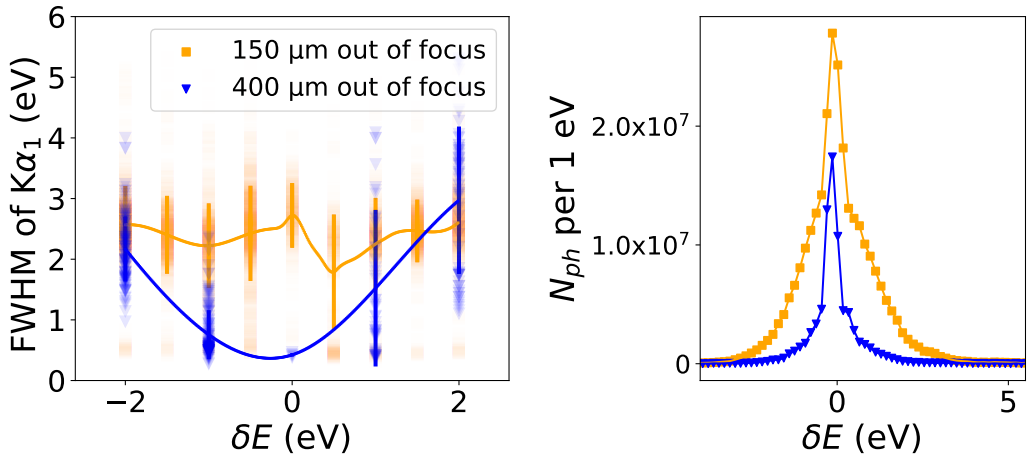


Figure 5.6: Left: The FWHM, f, of the Mn $K\alpha_1$ line as a function of the seed pulse detuning at different pump intensities. Lines are drawn for reference. Right: Example single-shot spectra, in which the seed is $\delta E = 0$ eV from the Mn $K\alpha_1$ line, to further illustrate spectral narrowing.

At high pump pulse intensity (150 μm out of focus), we observed a broad spectral width of $>$ 2.5 eV FWHM with little variation (see Fig. 5.6). The spectral broadening occurs once SSE reaches the saturation regime[154]. In contrast, at reduced pump pulse intensity (400 μm out of focus) the spectral width of the Mn Kα_1 line becomes narrower when the seed pulse is tuned directly to the resonance, with a minimum width $<$1 eV. We assign the observed narrowing to what in the optical regime is known as gain narrowing [190]. When fitting the signal in the low gain regime with a fixed seed pulse width, the two-component fit overestimates the Kα_1 SSE width due to its weak signal. It also overestimates the seed contribution, because in the gain narrowing regime it becomes impossible to differentiate between the spectral contributions from the seed and the $K\alpha_1$ line. Therefore, the fits represent an upper limit to the Kα_1 SSE width.

Figure 5.7 A shows the experimentally determined seeding efficiency ρ as a function of number of seed photons, measured at two different pump intensities (150 μm and 400 μm out of focus). At high pump pulse intensities (yellow squares, Fig. 5.7 A) the SSE signal is saturated, resulting in a seeding efficiency ρ which is greatest at low seed pulse intensities. At a lower pump pulse intensity (blue triangles, Fig. 5.7 A), we observe no obvious trend for ρ as a function of seed photon count.

Because losses in an XLO cavity are expected to be high [191], the seed pulse intensity will be low, hence conditions where ρ is high are required. If an external seed will be used for the initial pass of the XLO, seeding can occur even if the spectral overlap is not yet optimized (i.e. the effective seed pulse intensity is low).

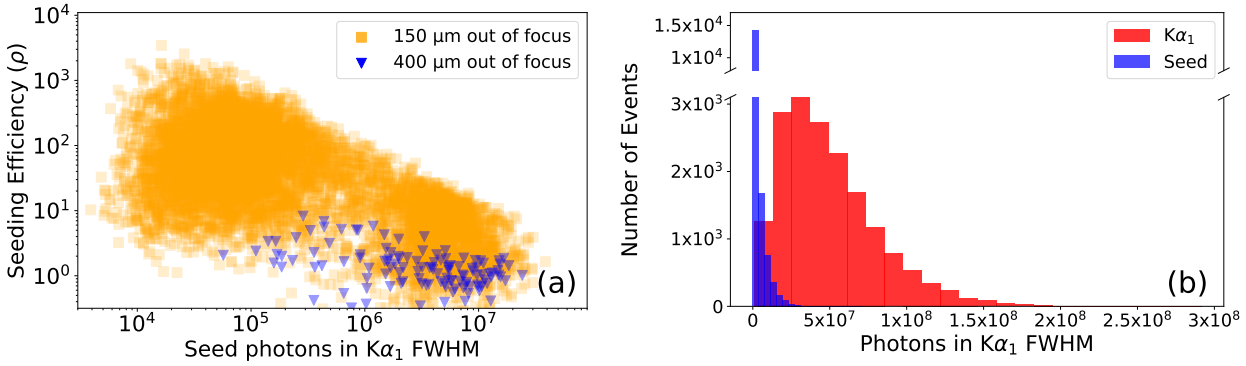


Figure 5.7: (a) Seeding efficiency as a function of seed photon counts that fall within the spectral range of $K\alpha_1$ FWHM for two different pump intensities. Shots are displayed for cases where the seed is in the range of $\delta E = [0, +7]$ eV for the case of 150 μm out of focus, and $\delta E = [0, +2]$ eV for the case of 400 μm out of focus. (b) Photon count distribution of $K\alpha_1$ and seed pulse photons from a seeded scan in focus, at a pump pulse intensity of 1.7×10^{19} W/cm^2.

Fig. 5.7 B indicates that the Kα_1 photon number peaks at $3{\times}10^7$ photons with lower probability of reaching $\leq 1 \times 10^7$ photons. The latter fact is likely attributed to contribution from the ASE signal. The seed pulse, in turn, exhibits close to negative exponential statistics attributed to the SASE pulse filtered through a monochromator.

5.2.6 Conclusion

To summarize our findings, the lowest pump pulse intensity at which we observed a single shot of SSE at the Mn Kα_1 line from a 3.9 molar $NaMnO_4$ solution was $1.1{\times}10^{17}$ W/cm^2, which is three orders of magnitude lower than previously reported ASE results from using the same sample [154]. For this case, a seed pulse with intensity $7.8{\times}10^{12}$ W/cm^2 was used and the seeding efficiency was $\rho = 3.9$. Using our approximation that 50% of the

photons fall within the FWHM of a two-dimensional Gaussian focus, a pump pulse intensity of 10^{17} W/cm^2 can be achieved by focusing an XFEL pulse with 1 mJ pulse energy and 7 fs pulse length to a diameter of 9 μm FWHM. This and comparable parameters are routinely achieved at all existing XFELs, and the intensity requirement can be further relaxed when using more concentrated gain media or samples. The highest pump pulse intensities currently available at XFELs are $>10^{20}$ W/cm^2 suggesting that SSE could be applied to transition metal solutions in the several millimolar range.

This shows the feasibility of carrying out seeded stimulated X-ray emission - an emerging field with potential for developing new nonlinear X-ray spectroscopy methods and generating transform-limited femtosecond X-ray pulses.

Our highest single-shot gain resulting in a seeding efficiency of (ρ= 4.0 $\times 10^3$ was achieved using a pump pulse intensity of 1.1×10^{18} W/cm^2 and a seed pulse intensity of 4.3 $\times 10^{11}$ W/cm^2. This is sufficient to overcome the estimated loss in the first round trip of the proposed XLO [78]. The gain almost disappears below a pump pulse intensity of 2.4×10^{17} W/cm^2 (400 μm out of focus) corresponding to an estimated pump spot size of 1.2 μm^2 FWHM. The largest number of Kα_1 photons observed in a single shot was 2.6 $\times 10^8$, resulting from a pump pulse intensity of 1.1×10^{18} W/cm^2 and a seed pulse intensity of 7.6 $\times 10^{14}$ W/cm^2. As this number scales with the total number of population inverted atoms, it can be further increased by increasing the beam diameter of pump and seed pulses, while providing the same intensity. This requires a larger XFEL pulse energy than we employed at SACLA. Such pulse energies in the mJ range are available at the LCLS and European XFEL, for example.

When removing the seed pulse, ASE was observed at a pump pulse intensity of 1.7×10^{19} W/cm^2. We estimate that the maximally concentrated Mn foil, ASE/SSE can be observed from pump and seed pulse intensity reduced by a factor of 8. The average FWHM from the measured SSE spectra collected at a given seed pulse energy range from $\sim$ 0.4 to 2.7 eV, which compares to the spontaneous $NaMnO_4$ Kα_1 linewidth of 2.1 eV FWHM [154].

5.3 Two-Color Mode

In this section, I will expand on a few additional experimental details, beginning with a deeper dive into SACLA's two-colour operation mode, which was a crucial component of the experiments described in this chapter.

This mode consists of two sections of variable-gap in-vacuum undulators. Each undulator is set at a different K value (see **Chapter 2**) so that two pulses at different wavelengths (not related by harmonics) can emerge. The relative wavelength tunability at SACLA is quite high, covering 5–15 keV, and the allowed wavelength separation between the two colors is over 30% [61]. This surpasses other efforts in the vacuum ultraviolet and soft X-ray regions, where the separation between the two wavelengths was on the order of just a few percent [192].

The undulators are separated by a magnetic chicane, composed of four dipole magnets in a serpentine arrangement, that detours and retards the electron beam. Further adjustment of the chicane parameters can tune the time interval between the two pulses at attosecond resolution (maximum delay is 40 fs), thus allowing for precise control over the time delay in pump-probe experiments [61]. For the experiments discussed in this chapter, controlling the delay between pump and seed pulses is very important because, in order to out-compete the dominant emission channel and selectively seed the transition of choice, the seed pulse must arrive within the core-hole lifetime($\sim$ 1 fs). When adjusting the time delay, there are a number of considerations. Electrons travel almost at the speed of light through the undulator, but photons are even faster, and outsprint the electrons by one radiation wavelength for every undulator period traveled.

After the first set of undulators, the photons (which have slipped ahead) must pass through a channel−cut monochromator and are reflected two times [193]. This adds additional path length and significantly retards the photons on a scale much larger than their initial advancement. As a result, the electrons (which are traveling through the chicane) must be delayed in order for the photons to catch up. By adjusting chicane strength [194], one can match the timing between the electrons and photons to account both for the initial delay caused by photons slipping ahead inside the undulator, along with the second delay experienced by the photons as they pass through the channel cut monochromator.

One of the first victories enabled by this two colour mode at SACLA was the generation of atomic X-ray lasing on the Cu Kα_1, Kα_2 doublet in 2015 [159].

LCLS also has two-colour capabilities, and in previous seeded stimulated work carried out at LCLS, the split undulator method [195] paired with strong pulse compression created the required pump/seed power and spatio-temporal overlap for the two colors. In addition to this, two colour mode has also been carried out at the FERMI light source in Italy in the extreme ultraviolet regime [196].

For the experiments presented in this chapter, it is important to achieve sufficient spatial overlap between the two colors. This is quite challenging when one considers how small both pump and seed pulses are, which is a consequence of SACLA's ability to maintain strict nanofocusing conditions. This is done through Kirkpatrick Baez focusing optics which consist of a set of two curved mirrors placed at right angles to one another. In this configuration, the first mirror focuses the radiation to a line, and the second mirror focuses the radiation line to a point, resulting in a spot size which has been demonstrated to reach down to 7 nm in width[1, 197].

In the previous section, I discussed how pump and seed pulse intensity are reduced by moving the sample further downstream. In theory, changing the pump and seed pulse intensities could have been achieved by detuning the upstream undulators. However, this was not possible during our experiment because detuning affects the focus size. Since maintaining strict nanofocus conditions were so important during our experiment, we chose the more straightforward path of moving the target out of focus as opposed to playing with the upstream undulators.

When considering the seed pulse intensity specifically, it is important to take into account

the fact that the undulators at SACLA are uncoupled from the monochromator. This means that the intensity of the incoming pulse will not remain constant (see Fig. 5.8) as it is scanned in energy (unless the undulators are manually changed in parallel). Moreover, this is further complicated by the fact that the undulators at SACLA take roughly two minutes for every movement. This is because their gap must first open up completely before every change - no matter the final spacing. In addition, the narrow bandwidth of the beam (24 eV FWHM) likely contributes to the sharp fall off in intensity we see at high/low energies for a given undulator position. This issue might become more relevant for an X-ray absorption study, and is worth keeping in mind when conducting experiments at SACLA.

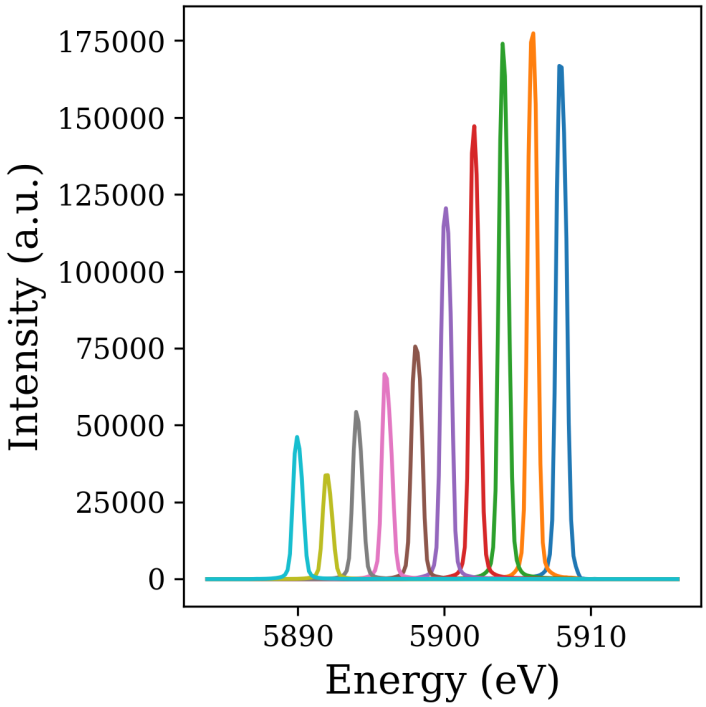


Figure 5.8: Array of seed pulse scans collected without sample for a given undulator position. Intensity drops off towards lower energies and it appears to take on a Gaussian envelope (see equation (5.2)). Shot-to-shot fluctuations, as one may expect at an XFEL, do not appear to be prominent enough to compensate for going off of the monochromator - thus demonstrating the relative stability available at SACLA.

$$G(E, f) = \frac{2.355}{f\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{2.355 \cdot E}{f})^2} \tag{5.2}$$

5.4 Observation of Multiplet Lines in Seeded Stimulated Mn Kα X-ray Emission

In order to determine the transitions that build an XES spectrum, either multi-peak fits are performed or the spectrum is simulated using a theoretical code. The first option usually underestimates the number of transitions, especially when some of them carry only weak intensity or are close in energy, while latter depends on the model that is used to describe the data. In this section, we aim at determining the energies of the XES transitions experimentally by taking advantage of SSE [159, 198].

SSE experiments at the Mn Kα transitions were performed for two complexes with a very different set of transitions due to their multiplet structures: NaMn(VII)O_4 which is formally a $3d^0$ complex, and MnCl(II)$_2$, which is nominally a $3d^5$ high spin S = $\frac{5}{2}$ complex [199]. The Kα_1 spectrum of $NaMnO_4$ consists of only one transition, while the $MnCl_2$ spectrum is built up by a large number of transitions as calculated by charge transfer multiplet simulations [199]. We take advantage of this difference by studying SSE by scanning the monochromatic seed pulse through the spectrum. The results are presented in **Section 5.4.1** and **Section 5.4.2**.

5.4.1 Sodium Permanganate

First, a seeded S-XES measurement was performed on a $NaMnO_4$ solution using the same experimental setup as presented in **Section 5.2**. $NaMnO_4$ has a nominal oxidation state of +7 and hence a formal $3d^0$ configuration. This leads to only one Kα_1 transition (Fig. 5.1 a) [199].

To remove subjectivity in the selection process, a peak finder algorithm (provided by Kirk Larsen) was used to identify peaks and their corresponding energies in the resulting seeded scan. It is loosely based on the concept of a constant fraction discriminator (CFD). For this, our raw signal was split into three components: One reference signal and two offset signals which were shifted in equal and opposite directions from the reference. The indices corresponding to energies at which the reference signal was greater than either offset were retrieved. Peaks were then assigned after selecting the local maximum from each identified cluster of points. A schematic of how this works is shown in Fig. 5.9.

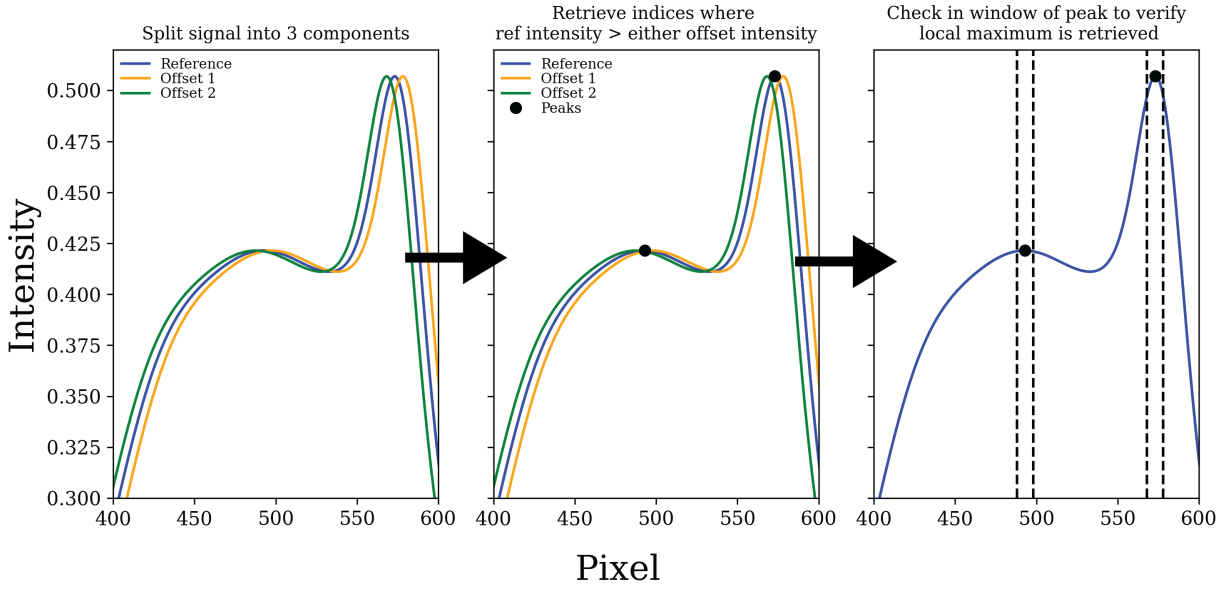


Figure 5.9: Left: Break down of how peakfinder algorithm works. First the raw signal is split into three offset components. Middle: Next, indices are retrieved where the reference intensity (blue signal) is greater than the intensity of either of the two offset signals(green and orange). Right: Finally, in the scenario where a cluster of peaks is found, the window of the peak is checked in order to verify the local maximum has been retrieved.

When deploying the peak finder on our seeded stimulated emission data, peaks are identified both in the signal region, and also for the strong seed pulse. When the seed is significantly detuned from the signal (Fig. 5.10), it is quite easy to distinguish between peaks associated with the seed (blue triangles) and peaks associated with the signal (black dots). The difficulty arises when the seed is underneath the Kα_1 spectrum, making it impossible in some instances to accurately distinguish between the two. While we have information regarding the nominal seed pulse energy (set by the machine), interactions between seed and signal have the potential to affect the central seed pulse photon energy.

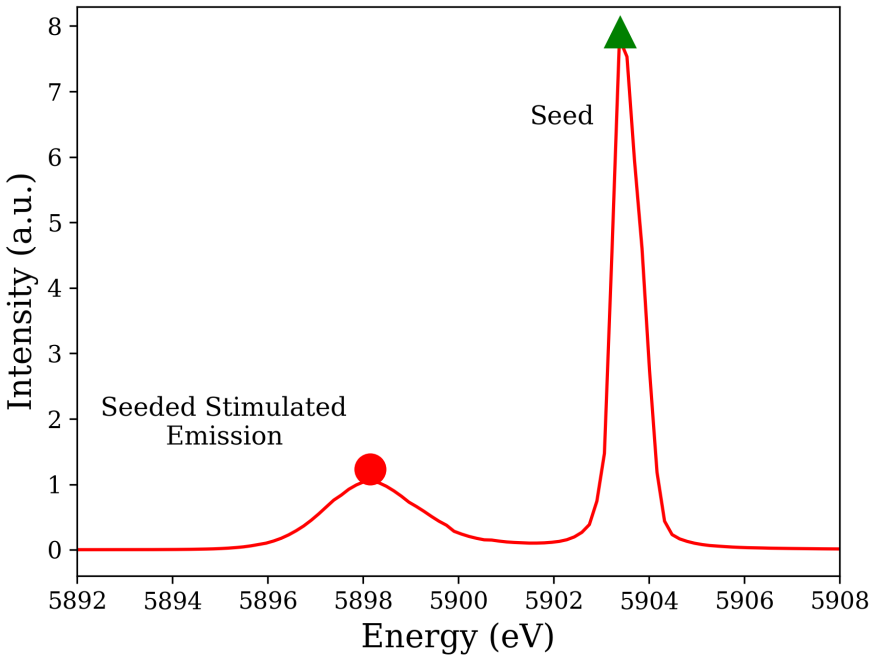


Figure 5.10: Illustration of the peak finder in action on a seeded stimulated emission spectrum, averaged over 1924 shots.

The results of the peak finder algorithm for the seeded stimulated XES experiment on $NaMnO_4$ is shown in Fig. 5.11. The green triangles correspond to the nominal seed pulse energy, and the red dots correspond to peaks identified from spectra averaged across 1924 shots (Fig. 5.11). Besides the seed pulse which is found on the diagonal, only one additional peak is found around ∼5898 eV, resulting in a single vertical line. This result is in agreement with the theoretical finding of only one transition in the $K\alpha_1$ region (Fig. 5.1).

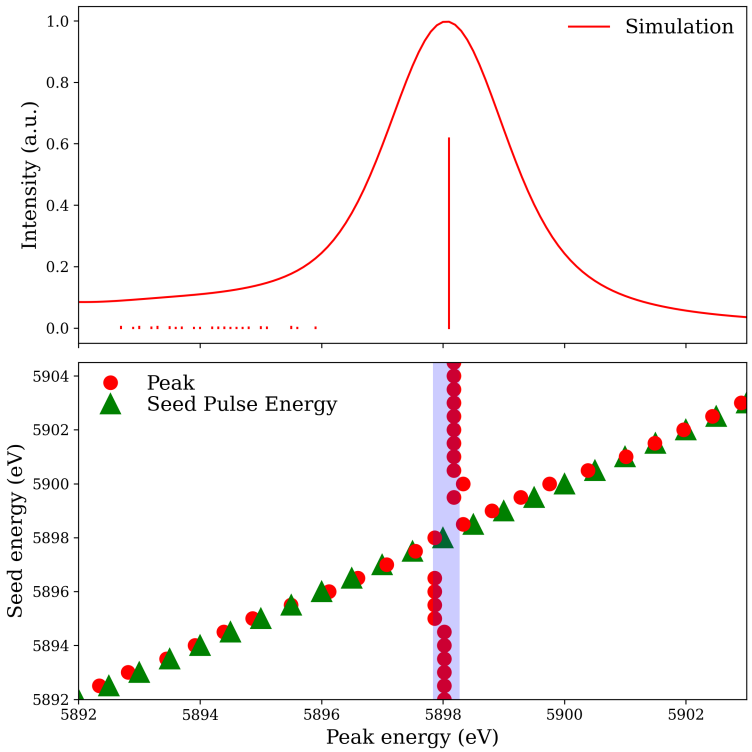


Figure 5.11: Top: Theoretical $NaMnO_4$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Results of the peakfinder algorithm deployed on average stimulated emission spectra.

The peak found corresponding to the transitions shifts to lower energies immediately after the seed pulse crosses the transition energy. The explanation of this behavior, which appears to involve a pulling of the stimulated emission signal towards the seed pulse photon energy, is beyond the current scope of this study.

5.4.2 Manganese Chloride

Next, we drove the seed pulse through the $MnCl_2$ envelope (Fig. 5.12). Similar to the $NaMnO_4$ case, there is an always present base signal that appears even with the seed photon energy being above the transition energies. This was also shown in **Section 5.2** and occurs because the photons in the long Lorentzian tails (equation (5.3)) of the seed pulse have the potential to seed stimulated emission.

$$L(E, f) = \frac{1}{2\pi} \frac{f}{((E^2 + (f/2)^2))} \tag{5.3}$$

While this is useful in suggesting that SSE is not a rare event, it also hinders easy detection of enhanced transitions through seeded stimulated emission.

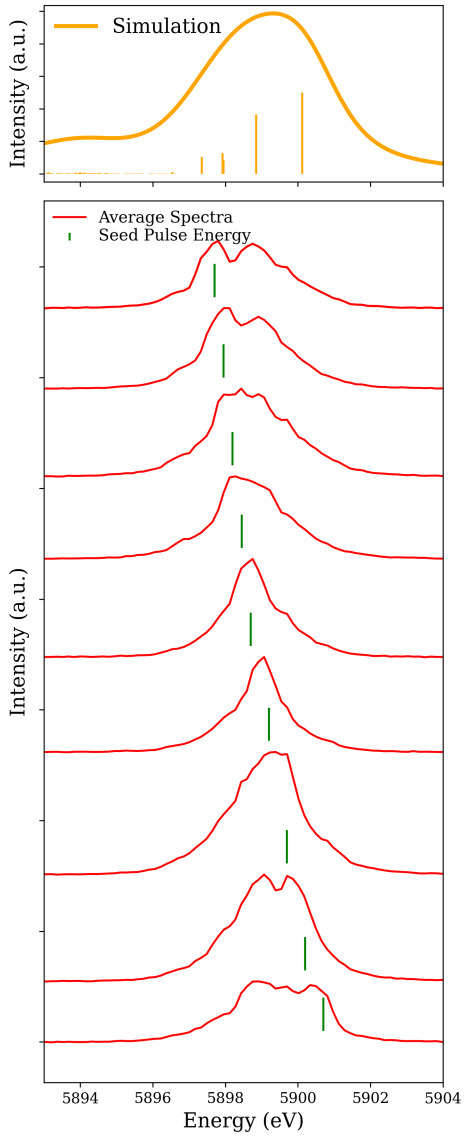


Figure 5.12: Top: Theoretical $MnCl_2$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Waterfall plot showing averaged $MnCl_2$ spectra from a seeded scan through the Kα_1 envelope. Seed pulse photon energy is shown as the green vertical line.

However, there are several small features at higher and lower energies present in the spectra (Fig. 5.12). These features are weak and have the potential to be nothing more than

noise. However, as they appear at multiple consecutive curves they more likely represent the energies of the individual Kα transitions.

The result from the peak finder on $MnCl_2$ is shown in Figure 5.13. The green triangles show the results of the seed pulse photon energy and therefore appear on the diagonal. The red circles show the results of the peak finder algorithm.

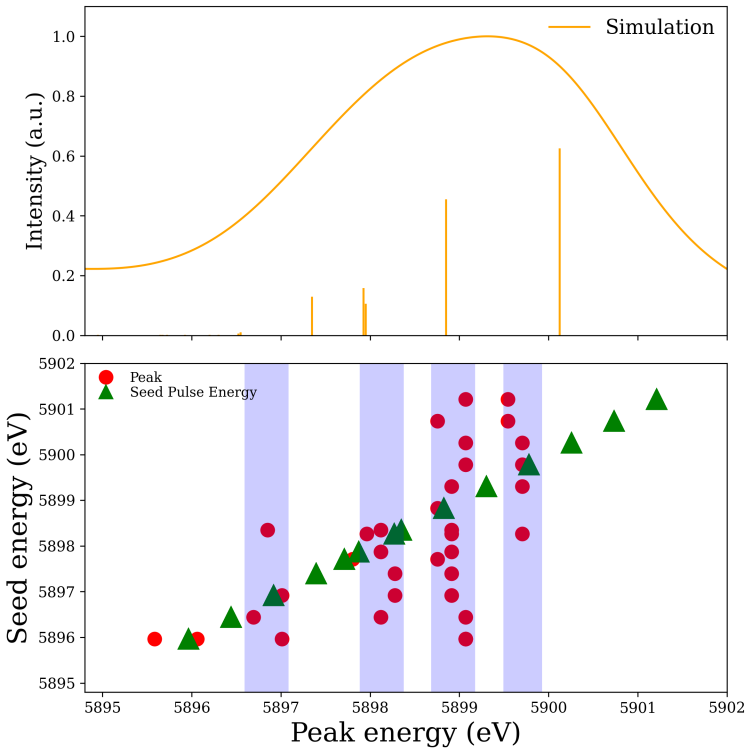


Figure 5.13: Top: Theoretical $MnCl_2$ Kα_1 spectra, the vertical lines underneath represent the calculated individual transition intensities. Bottom: Results of the peak finder algorithm from the top 2% strongest shots. Light blue vertical bars are shown to guide the eye to partially reproducible peak positions.

Peaks were found that appear partially at the same emission energy. From this analysis method, four vertical energy regions are found that are highlighted in Figure 5.13 by light blue vertical bars as guide to the eyes. These distinguishable transitions agree with the theoretical result.

The first transition, calculated as the strongest transition, does not lead to the strongest intensity. This suggests that either the multiplet calculations do not find the correct intensities, or that in seeded stimulated emission, the main intensity arises from the main intensity of the spontaneous spectrum due to an overlap of multiple stimulated transitions.

Together with the multiple energy peaks found for $MnCl_2$, this offers confirmation that the Kα features observed in $MnCl_2$ appear from a more complex multiplet structure.

In subsequent iterations of this experiment, this $MnCl_2$ result was not reproduced completely. While the existence of vertical lines is reproduced in other data sets, their number and energy spacing is only roughly reproduced.

The multi variable parameter space for these experiments is quite large and, due to the nature of the XFEL instrument, has the capacity to fluctuate significantly depending on beamline parameters. Because hours are limited during a beamtime, and instrument adjustment carried out by accelerator physicists upstream can cut into data collection hours, users must plan strategically how they use their time during an allotted experimental period. Future experiments dedicated to adjusting variables such as pump and seed pulse intensity, pulse duration, pulse shape and delay between pulses have the potential to be quite fruitful in identifying the precise set of parameters required for optimal, reproducible experimental conditions.

5.5 Outlook

In the final section of this chapter, I will discuss a few potential applications of SSE to other forms of XFEL-based science.

5.5.1 Seeded Stimulated Emission on Low Concentration Metalloenzymes

The application of SSE for spectroscopic studies at XFELs offers exciting and fruitful opportunities across a range of research areas, and some may be near at hand. In the experiments described in this chapter so far, highly concentrated samples (4M) were used, and SSE was observed at pump intensities down to 10^{17} W/cm^2. The highest pump intensity currently available at XFELS are greater than 10^{20} W/cm^2. If stimulated emission occurs when the concentration times pump pulse intensity reaches a certain value, one can imagine leaving the pump intensity the same, but dropping the actual sample concentration from 4 M to 4 mM (closer to PS II levels, for example) and still getting some enhancement of the signal. Future studies dedicated to confirming the concentration limit for SSE are required to verify this assumption, which does not take into account differences related to metal-metal distances within the sample.

5.5.2 Exploring Manganese Complexes with Different Multiplet Structures

In the future, one can also consider measuring seeded stimulated emission on chemically important mixed-valent model compounds, such as Mn_2(III,IV)Terpy, which has been used in studies related to characterizing the O_2 evolution reaction [200]. This compound has many

transitions into multiplet final states and observing SSE from individual multiplet lines from this compound would help establish the general application of this approach and provide an important benchmark for multiplet calculations.

5.5.3 Phase Stable Pulse Pairs

Two stimulated X-ray emission pulses at the 5.9 keV Mn Kα line, separated by a few femtoseconds each, can be generated from a single SASE pump pulse. The interference between these two pulses can then lead to an additional fringe pattern in the resulting spectrum [163].

Such interference fringes have been observed in our Mn Kα S-XES datasets collected at SACLA. This phenomenon is thought to be a result from the random occurrence of two strong temporal spikes in the SASE pump pulse.

Much effort has been made to thoroughly sort our fringe-contaminated datasets into two distinct categories: One for SSE analysis, and the remainder for fringe-dedicated analysis. Up until this point, a modified version of the fringe-finding algorithm presented by Zhang et al. [163], is used as the starting point. Efforts are underway to expand on this and generate a fringe-only dataset, which has proven to be quite challenging based on the fact that defining a fringe is non-trivial.

When we are able to study the generation of fringes in better detail, and are finally able to understand the precise set of parameters that are involved in "turning on" and "turning off" the fringes, we can then begin to think about controlling and deploying them with potential applications in precision measurements, advanced nonlinear spectroscopies and imaging.

5.5.4 An X-ray Laser Oscillator

XFELs offer high brightness, but they exhibit broad bandwidth, poor shot-to-shot stability and limited temporal coherence. By deploying similar techniques as those used with conventional optical lasers, these limitations can be improved upon [201]. In theory, this would simply involve utilizing a gain medium that is pumped to achieve a population inversion [202], as well as an oscillator to provide stable, transform limited pulses. In practice, however, achieving coherent hard X-ray pulses is quite challenging for a number of reasons. Notably, it's difficult to find the right optics that can withstand the heat loads imposed by high energy X-rays. For this reason, materials like diamond are typically used due to their high thermal conductivity (up to 2,200 $\frac{W}{m \cdot K}$) [203].

Additionally, the lifetimes of inner shell excited states are short, so large pumping power is needed in order to maintain a population inversion. As has been demonstrated throughout this chapter, high peak power can be provided by XFELs, making them useful pumps for this purpose.

This section will discuss some of the efforts surrounding a proposed X-ray Laser Oscillator (XLO), which involves seeded Kα S-XES at its core, and is based on the concept of population inversion using an XFEL as a periodic pump. The general scheme of the XLO is as follows.

First, an XFEL delivers a train of pump pulses, spaced about 30 ns apart (equivalent to the round trip through the cavity). The first pulse is focused into the sample, a transition metal compound. For these initial experiments, we focused on copper-based samples so a 9 keV pump tuned above the K edge was used, generating stimulated X-ray emission. As the resulting radiation circulates through the cavity, which consists of four Bragg crystal optics composed of either silicon or diamond, it can only propagate within a small divergence angle. This creates a cavity-like gain narrowing.

After its first pass, the newly monochromatic pulse arrives simultaneously with the next pulse, at which point it serves as a seed for stimulated emission. The cycle repeats such that the initially large bandwidth, non-transform limited ASE pulse gets progressively narrower in every pass, eventually generating longitudinally coherent pulses anticipated to be around 50 meV FWHM bandwidth [78]. Theoretical studies have shown that the oscillator saturates and reaches full coherence in 4 to 6 cavity transits [78].

The initial experiments for the XLO took place at the CXI instrument at LCLS [204]. These experiments have been carried out in two modes, one which involves ASE and seeded photon measurements, the other involves cavity tests. This section will focus on the former. The setup for these measurements is illustrated in Fig. 5.14.

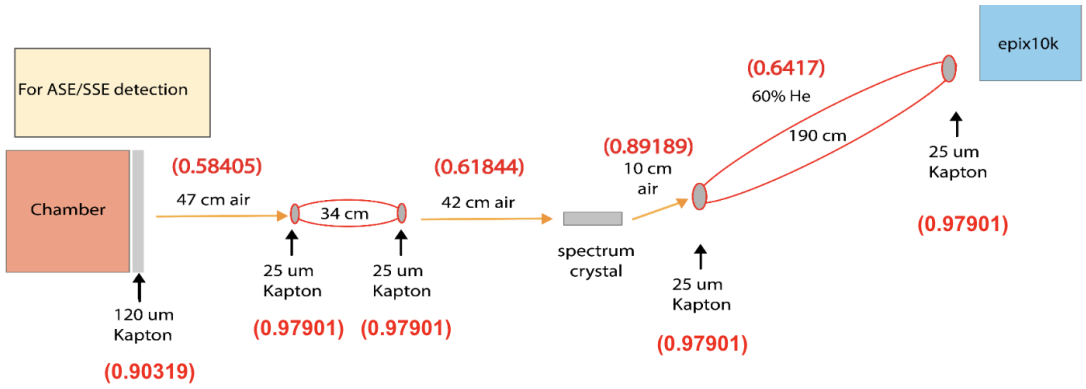


Figure 5.14: Experimental setup (including estimated losses) for copper ASE and SSE measurements at CXI at LCLS.

In initial experiments, ASE was collected from a jet of $Cu(NO_3)_2$, which consists of roughly 2 $\frac{\text{atoms}}{nm^3}$, and up to $\sim$ 1 mJ pump. We then switched over to collecting ASE from a solid spinning copper foil disk, which contains around 86 $\frac{\text{atoms}}{nm^3}$, with a $\sim$ 100 μJ pump. A solid target offers higher density of the target atomic species while containing no other elements which absorb the pump pulse, thus resulting in a similar number of Kα photons generated using a lower pump pulse intensity (Fig. 5.15). Seeded data, using around 10^4 to 10^5 seed photons, was also collected from the foil. This resulted in an increased number of copper Kα photons by about one order of magnitude while using the same pump pulse energy (Fig. 5.15)

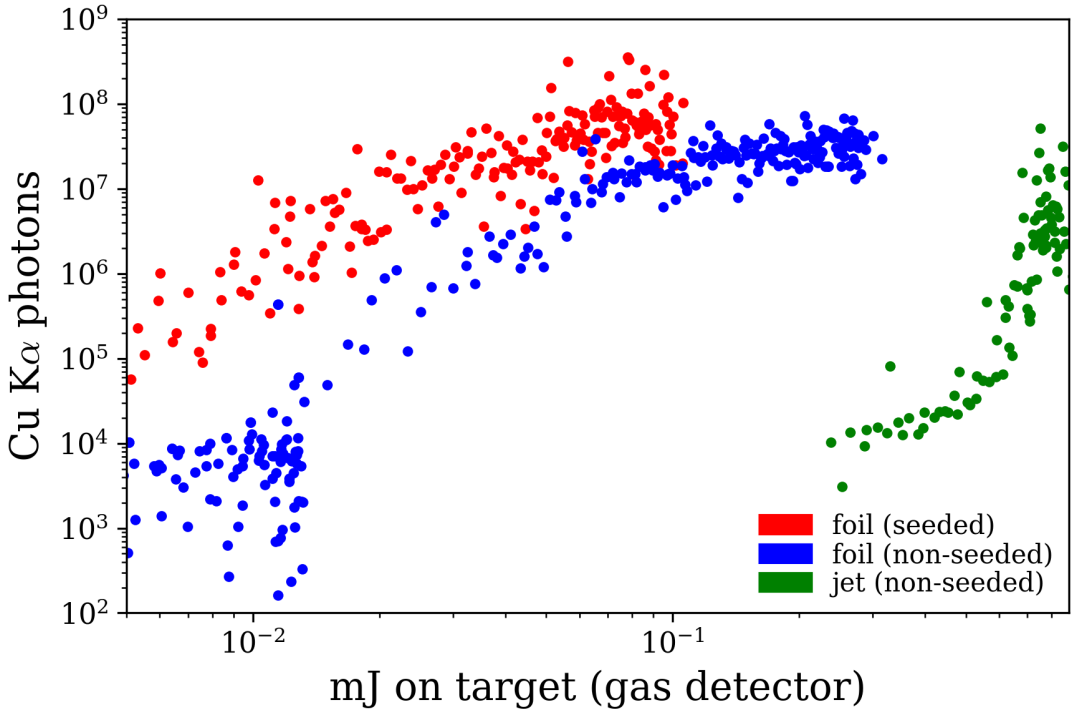


Figure 5.15: ASE and seeded photon flux measurements for a jet solution and solid copper gain medium. Reproduced from [205].

Lowering the pump pulse requirements from mJ to tens of μJ is advantageous because it mitigates thermal problems that could occur in the crystals. It also can increase the likelihood that fresh gain medium is provided for each incoming pulse. Any ultra high intensity X-ray pulse will transform a target material into a hot and rapidly expanding plasma - so it is absolutely crucial that the gain medium is replenished before the next pulse arrives.

The number of Kα photons generated is an important metric to consider because on the first pass through the cavity, the losses will be quite significant (99.7% [78])and must be overcompensated for by large initial gain. Net amplification must exceed net reflection loss.

When it comes to the spinning copper foil, there must also be a careful balance between pump pulse energy and disk speed, as the former will influence the size of the affected area in the foil. As a result, this will requiring faster rotation in order to move to a fresh spot before the next pulse. The relationship between damage radius and mJ on target can be shown in Fig. 5.16, with crater size analyzed by means of scanning electron microscopy (SEM). From this analysis, it appears that it is possible to fully clear $\sim$ 10 μm with a velocity of 300 $\frac{m}{s}$.

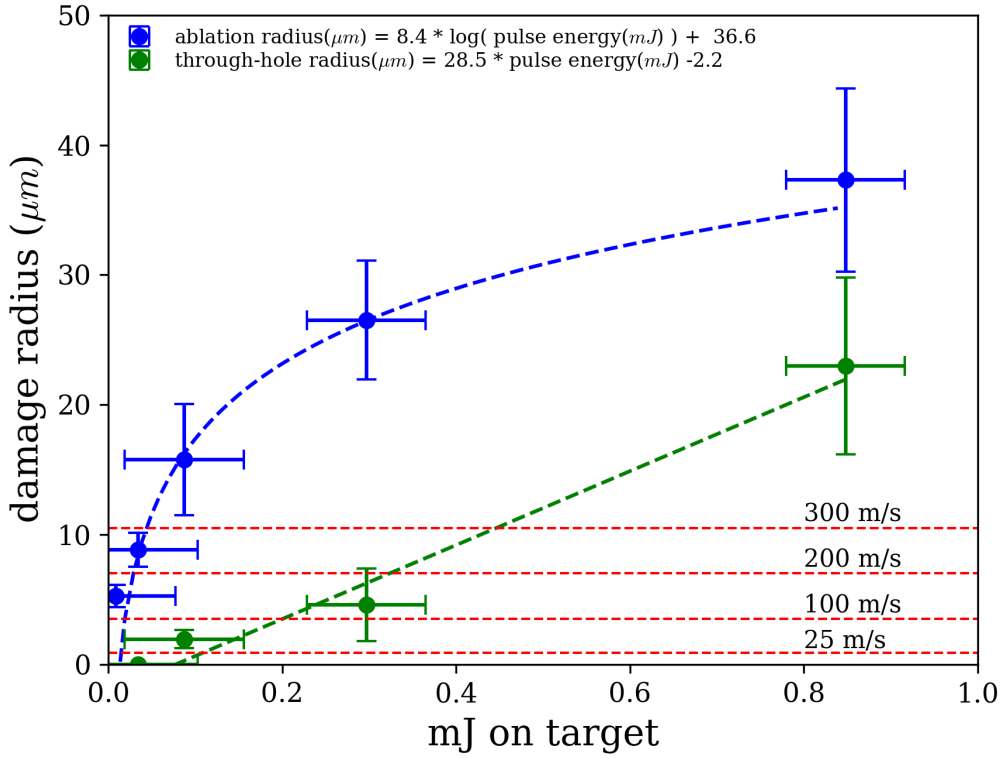


Figure 5.16: Crater size scaling laws resulting from SEM analysis. Reproduced from [205].

I will end this section by making the point that the XLO should not be confused with the previously proposed X-ray Free Electron Laser Oscillator (XFELO) [80]. The first major difference between the two is that the XLO configuration is much more compact than the XFELO. The XLO also produces a better-defined photon energy due to the use of atomic transitions, and is not sensitive to electron beam energy fluctuations or jitter.

Chapter 6

Molecular Dynamics Simulations on Photosystem II

6.1 Introduction

The previous two chapters have focused on X-ray based spectroscopic techniques in order to probe the electronic structure of 3d transition metal complexes. Generally speaking, to form a more complete picture of the reaction mechanism or system under study, electronic changes must be placed in the context of structural changes occurring at the active site as well as throughout the surrounding protein environment. Using the DOT system described in **Chapter 4**, our group has collected and published high resolution room temperature crystal structures of the meta stable intermediate states of PS II (S_1, S_2, S_3, S_0) [69], as well as data at several time points in between those transitions [206, 118, 131].

The structural data at current high resolution (2 Å and better) allows modeling highly occupied waters into the structures, and investigate their functional roles. This allows for higher accessibility when it comes to studying waters and the networks they form. This is especially interesting in the case of PS II, since the enzyme has a mechanism to transport water (substrate) to the active site from the bulk and also transport protons out from the active site for efficient catalysis.

Three water channels, which connect the manganese cluster to the bulk, have been hypothesized to play an important role in proton and water transport throughout the catalytic cycle. They are called the O1, O4 and Cl channels - each named after the atom in the cluster which they terminate at. Each of these channels were predicted using computational approaches based on crystallographic studies [207, 208, 209, 210] and has been proposed to serve a specific purpose during at least one portion of the catalytic cycle.

The O4 channel, for example, has been proposed as a proton transfer channel during the $S_0 \rightarrow S_1$ transition [211, 128]. The Cl1 channel, on the other hand, has been proposed as the main proton release pathway in the $S_2 \rightarrow S_3$ and $S_3 \rightarrow S_0$ transitions. Finally, the O1 channel has been proposed to be the main water intake pathway based on the high mobility

of waters in this channel [206, 118].

Our hypothesis has been that proton transfer channels have highly structured waters with less mobility, and the water intake pathway has high mobility. To investigate the mobility of waters, we extracted the B-factors obtained through the SFX data. The B factor, otherwise known as the temperature factor or Debye-Waller factor, can be described by equation (6.1) [212]. Since $\langle u^2 \rangle$ is the mean squared isotropic displacement of the atom, a higher B factor should correlate with more mobile waters.

$$B = 8\pi^2 \langle u^2 \rangle$$

Extracting water-related insight beyond this from the experimental data can be quite challenging. After all, the crystallographic data provides only a static, averaged picture. While the electron density maps we obtain at an XFEL are packed with information, they are often difficult to interpret; and this is especially the case in the region of the water networks where dynamics are prevalent. Moreover, gathering hydrogen position information from the experimental data is not possible due to the current resolution (one would need to approach a resolution around 1 Å), rendering hydrogen-bonding related questions difficult to answer.

Computational tools such as Molecular Dynamics (MD) simulations can help expand upon what is currently possible from our room-temperature crystallographic data while also aiding in further interpretation of our electron density maps. In this chapter, I will describe crystalline MD simulations, starting from the room temperature crystal structures of PS II collected at an XFEL and simulating a crystalline environment which accounts for the four copies of the PS II dimer in a P1 unit cell. I will explain the reasoning behind various preparation criteria used for the simulation, which differ from previous MD simulations of PS II (**Section 6.2**) [213, 214, 215, 216, 217]. The explicit choices in the configuration of the simulation allow for element specific electron density maps to be computed from the resulting trajectory, which can then be directly compared to our experimental maps. Then, analysis of PS II's water and hydrogen bonding networks can be carried out (**Section 6.3**). Finally, I will discuss the future directions of this work (**Section 6.4**).

6.2 Crystalline Molecular Dynamics Simulation Preparation

6.2.1 The Case for Crystalline MD

Previously reported simulations on PS II have involved placing a single protein, collected under cryogenic conditions, into a box of solvent [213, 214, 215, 216, 217]. Pioneering work on endogluconase by Wall et al. demonstrated a new crystalline MD approach that was designed specifically for comparisons to X-ray diffraction data [218]. This has the advantage

of replicating the crystalline conditions, and subsequent crystal packing effects, during data collection.

This crystalline approach is similar but not identical to the traditional solution approach, and it provides the foundation for the work recounted in this chapter. In brief, this involves propagating the asymmetric unit to the unit cell using the system's symmetry operations. In the case of PS II, this corresponds to $P2_12_12_1$. In previous works, a periodic 2 x 2 x 2 supercell was constructed [218, 219]. This has the advantage of improving the modeling across unit cell boundaries; however, for systems as large as PS II, this is quite computationally expensive. For now, just a single unit cell is modeled.

6.2.2 Ensemble Selection

The conditions in which an MD simulation is carried out can be referred to as an ensemble. The chosen ensemble for the simulation will depend on the experimental conditions that one wishes to simulate. Three popular ensembles used in MD simulations consist of the micro canonical ensemble, the canonical ensemble and the isobaric-isothermal ensemble [220].

The first of which, the microcanonical NVE ensemble, describes the most basic case. In this ensemble, the total number of particles (N), the total volume (V) and the total energy (E) remain constant. Even though the total energy of the system is conserved, fluctuations in potential V and kinetic energy K are allowed. Since the temperature T can fluctuate, selecting this ensemble can result in complications for molecular simulations because at high temperatures, the protein can unfold [221, 222].

By contrast, with the canonical NVT ensemble, the energy E of the system is *not* conserved, while the number of particle N, the volume V, and temperature of the ensemble T remain fixed. The canonical NVT ensemble is desired for crystalline simulations because it maintains the integrity of the unit cell dimensions. Additionally, since this ensemble allows for heat to be exchanged with the outside environment (via heat bath), the temperature remains fixed, which can avoid the aforementioned protein unfolding issues.

In the isobaric-isothermal NPT ensemble, the total number of particles N, the pressure P and the temperature T remain constant. Since many experiments are carried out under constant pressure conditions, this ensemble is used often [221] .

It is not necessary to select a single ensemble when carrying out an MD simulation, and in fact one may mix ensembles depending on experimental imperatives. However, the work presented in this chapter concerns simulations that can be directly compared to the crystalline structure using the unit cell and space group information. As a consequence, simulation box volume is quite important to keep in mind; therefore, a NVT ensemble is chosen.

6.2.3 Restraints in Production Simulations

In the simulation results presented in this chapter, all heavy atoms from the protein (does not include solvent) have been tethered to their initial crystal structure positions via

harmonic restraints. This is required because, although the force fields employed in MD have improved substantially in recent years, they are inherently approximate. Quite strong restraints ($1000\frac{\text{kJ}}{\text{mol}\,\text{nm}^2}$, the default of the simulation software) were applied in this work. However, the actual strength of restraints chosen is an active area of research. Previously, Wych et al.[223] performed a titration of restraints and discovered that lower restraints actually result in better reproduction of the crystallographic B factors. Experimenting with the restraints is an interesting area for future explorations, but is beyond the current scope of this work.

6.2.4 Computing Densities from MD trajectories

This chapter compares results from MD simulations with experimental crystallographic data. I am able to do this by first computing average structure factors over the trajectory (equation (6.1)).

$$F_{h,k,l} = \sum_{j}^{atoms} f_j e^{2\pi i(hx_j+ky_j+lz_j)} \tag{6.1}$$

In equation (6.1), h,k,l correspond to the Miller indices, f_j corresponds to the atomic form factor for atom j, and x,y,z correspond to fractional coordinates of the atom within the unit cell. This computation is carried out using the using xtraj.py software, which is available in the following repository (*https://www.github.com/mewall/lunus*). Initially, a trajectory in .xtc format is used as the input, and amplitudes and phases of structure factors from the simulation are output into .mtz file format using cctbx tools [224]. The lunus software enables atom selectivity so one has the option of selecting only solvent moleculues, or only oxygen or hydrogen atoms, and completing the calculation over that sub-structure as opposed to the full structure. This feature is not available in conventional crystallography.

Once structure factors from every frame are computed, they can be averaged across multiple frames. The calculation can be run on multiple nodes of a high performance computing cluster, thus allowing for efficient calculation of densities for large simulation. By setting the space group information, one also has the option to fold results computed across the entire unit cell back into the asymmetric unit using symmetry operations. Doing so helps take better advantage of the fact that the simulation is being run across multiple copies of the PS II dimer and improve statistics. Then, through a Fourier transform, electron density maps can be obtained for direct comparison to experiment.

6.3 Water Networks in Photosystem II Using Crystalline Molecular Dynamics Simulations and Room Temperature XFEL Serial Crystallography

The content and figures of this section are adapted from:

Margaret D. Doyle, Asmit Bhowmick, David C. Wych, Louise Lassalle, Philipp Simon, James Holton, Nicholas K. Sauter, Vittal K. Yachandra, Jan F. Kern, Junko Yano and Michael E. Wall, "Water Networks in Photosystem II Using Crystalline Molecular Dynamics Simulations and Room-Temperature XFEL Serial Crystallography," Journal of the American Chemical Society 145.27 (2023): 14621-14635.

It can be accessed at the following link:
https://doi.org/10.1021/jacs.3c01412

6.3.1 Abstract

Structural dynamics of water and its hydrogen bonding networks play an important role in enzyme function via the transport of protons, ions, and substrates. To gain insights into these mechanisms in the water oxidation reaction in Photosystem II (PS II), we have performed crystalline molecular dynamics (MD) simulations of the dark stable S1 state. Our MD model consists of a full unit cell with 8 PS II monomers in explicit solvent (861,894 atoms), enabling us to compute the simulated crystalline electron density and to compare it directly with the experimental density from serial femtosecond X-ray crystallography under physiological temperature collected at X-ray free electron lasers (XFELs). The MD density reproduced the experimental density and water positions with high fidelity. The detailed dynamics in the simulations provided insights into the mobility of water molecules in the channels beyond what can be interpreted from experimental B factors and electron densities alone. In particular, the simulations revealed fast, coordinated exchange of waters at sites where the density is strong, and water transport across the bottleneck region of the channels where the density is weak. By computing MD hydrogen and oxygen maps separately, we developed a novel Map based Acceptor Donor Identification (MADI) technique that yields information which helps to infer hydrogen bond directionality and strength. The MADI analysis revealed a series of hydrogen bond wires emanating from the Mn cluster through the Cl1 and O4 channels; such wires might provide pathways for proton transfer during the reaction cycle of PS II. Our simulations provide an atomistic picture of the dynamics of water and hydrogen bonding networks in PS II, with implications for the specific role of each channel in the water oxidation reaction.

6.3.2 Introduction

Water enables the chemistry of life. It is the medium in which all biological reactions take place, it controls the kinetics and energetics of molecular encounters, and it also serves as a substrate for proton transport. The mobility and flexibility of water in forming hydrogen bond networks are central to these roles; however, our understanding of how water dynamics enable and control biological function is far from complete. Attempts to visualize the chemical and structural changes of water molecules in real time have yielded key insights, while also revealing technical and methodological challenges that must be overcome to increase our understanding [225, 226].

X-ray crystallography has been the main experimental tool for understanding protein structure along with the location of waters and the networks they form. Neutron crystallography also has been used, in particular to identify the location of hydrogen nuclei. Recently, cryo electron microscopy has become a mainstream approach for visualizing single molecules. After deriving a starting model from the structures collected through these techniques, computational methods like molecular dynamics (MD) simulations have provided insight into the atomic scale interactions in proteins, which has led to better modeling of water in biomolecules.

For macromolecular crystallography, the introduction of X-ray free electron lasers (XFELs), which generate ultrabright X-ray pulses by accelerating relativistic electrons through a periodic arrangement of undulator magnets, has enabled us to look at time resolved structural changes of enzymes. The data can be collected at physiological temperature prior to the onset of X-ray induced changes, by taking snapshots after triggering reactions with a wide array of methods [162, 69, 227, 228, 118, 206, 229]. The effects of temperature on protein structure and dynamics have been investigated extensively, with the implication that cryocooling could modify the conformational state of enzymes and alter the water network [230, 231, 11, 13, 232, 12, 66, 233, 234]. Thus, by determining the 3D structure of a protein under physiological temperatures and at timepoints along the reaction pathway, XFEL crystallography allows us to capture the most biologically and chemically relevant conformational states of a protein.

Concurrently, advances in MD methods and high performance computing have enabled >100 ns MD simulations (such as the one discussed in this paper) for protein systems with 10^5-10^6 atoms to be performed routinely. Recently, crystalline MD simulations have been utilized in X-ray diffuse scattering studies as well as for closing the "R factor gap" in proteins (the gap in agreement between observed and predicted structure factors, i.e., R_{work} and the experimental errors i.e., R_{merge}) by modeling the protein solvent interaction [235, 236, 237, 238, 239]. In these studies, the crystalline conditions were simulated by using the known crystal packing and experimental sample conditions to generate a periodic simulation box that is consistent with the unit cell dimensions. Using this scheme enables the simulations to be compared directly to crystallographic data using standard crystallography methods available, e.g., in Phenix [240] and CCP4 [241]. In addition to using standard methods, it is also possible to use simulated diffraction images, produced from crystalline MD trajectories,

to make comparisons [242].

Here, we combine XFEL crystallography data with crystalline MD simulations to investigate water structure, dynamics, and hydrogen bonding networks in Photosystem II (PS II), the enzyme responsible for catalyzing the light induced oxidation of water to molecular oxygen [122, 123]. Water serves a dual role as both substrate and solvent in PS II. The active site for PS II is a metal cluster called the oxygen evolving complex (OEC; Mn_4CaO_5), which is located on the lumenal side of the protein complex. The OEC is shielded from the direct exposure to bulk solvent by large extrinsic loops of the D1, D2, CP43, and CP47 subunits as well as by several membrane extrinsic subunits PsbO, PsbU, and PsbV in the case of the thermophilic cyanobacterial complex.

However, the cluster is connected to the bulk solvent via hydrophilic channels (Figure 6.1a) [207, 208, 209, 210]. The formation of molecular oxygen from two water molecules at the cluster takes place in a sequence of four intermediate steps called the Kok cycle (Figure 6.1b), starting from the dark resting S_1 state [243, 124]. Each photon absorption by the chlorophylls in the PS II reaction center (P680; not shown in the figure) leads to oxidation of the cluster and further advancement along the cycle ($S_i \in 0,1,2,3,4$) until the transient S_4 state is reached at which point molecular oxygen is formed and the cluster is reset to the S_0 state for the process to begin again.

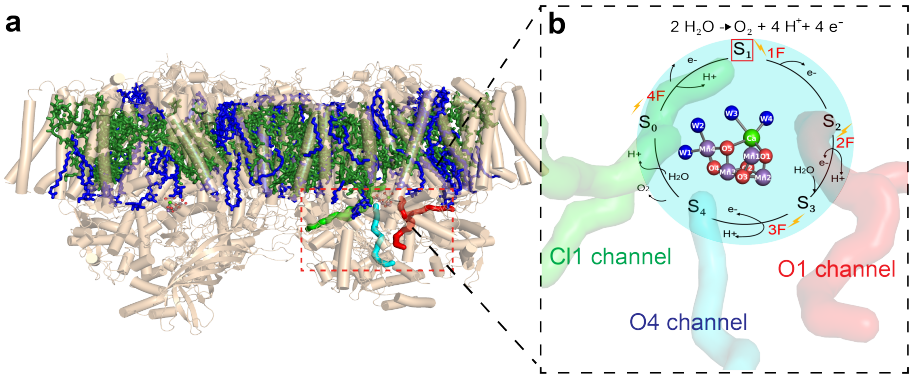


Figure 6.1: Overview of PS II and its electron donor site where water is oxidized to molecular oxygen. (a) The protein structure of PS II from Thermosynechococcus vestitus serves as the starting point for our MD model. The lipids (blue) are highlighted along with cofactors (green) and the three water channels (O1, pink; O4, cyan; Cl1, green). (b) OEC (only the $Mn_4CaO_5(H_2O)_4$ cluster) is shown in the context of the water channels, along with the Kok S state cycle, advancing through the S states (S_i, i = 1-4) by illumination with 1 through 4 flashes (1-4F). The S_1 state (solid red box) is the focus of the current work.

Both water transport and the proton release, in a well controlled stepwise manner, are critical for the water oxidation reaction, as it advances through the S state intermediates in the Kok cycle. Bulk water proximal to the catalytic center is detrimental as the intermediates

that are generated with the high oxidative potentials present at the catalytic center can lead to very reactive species that can damage the protein. Thus, it is important that the water is delivered to the catalytic center in a stepwise and controlled manner, and following this process is critical for understanding the mechanism of water oxidation by PS II. Several channels in PS II have been identified using many experimental studies and simulations [207, 208, 209, 210, 244, 216, 217, 245, 23, 215]. The three which are thought to be important for water and proton transport are the O1 channel ("large channel"), the O4 channel ("narrow channel"), and the Cl1 channel (Figure 6.1). All three channels start near the OEC and extend outward toward the lumenal side of the complex to the bulk water. The O1 channel (proposed water channel, near the O1 bridging ligand of the Mn_4CaO_5 complex) [206, 210] originates near Ca in the OEC before splitting into two branches, A and B. The O4 channel (proposed proton channel in the S_0 to S_1 transition, near the O4 bridging ligand of the Mn_4CaO_5 complex) [215] starts by the O4 side of the OEC and extends toward the subunits PsbO and PsbU without branching. Finally, the Cl1 channel (proposed proton channel at least in the S_2 to S_3 transition, Cl binding site is located near this) [206, 246, 247, 213] originates near Mn4 and contains both a short branch (A, sometimes known as the broad channel) [207] and a long branch (B, extends toward PsbO). Further details on these channels have been discussed in Hussein et al. [206].

To better understand the involvement of the water networks of PS II in the transport of water and the transfer of protons through the channels, we prepared a crystalline MD model of the PS II unit cell in explicit water, using a physiologically relevant room temperature XFEL crystal structure in the dark stable S_1 state. While simulating proton transfer itself requires a quantum mechanical simulation, here we only look to infer the likely pathways of proton transfer from the hydrogen bond networks derived from classical MD simulation. Doing a crystalline MD simulation with our protocol also allows us to reproduce the experimental water structure and cover a much longer timescale and larger system size than would be possible using a quantum simulation.

We carried out the simulation using two different ionic solvent conditions. In the first case, we added 479 Na^+ions to neutralize the negative charge of the protein.

In the second case, we added Na^+ and Cl^- ions in equal proportion (489 ions each) to mimic experimental buffer conditions of the XFEL crystallography, and then an additional 479 Na^+ ions to neutralize the negative charge [206]. We refer to the former in the text as the "simulation" and the latter as the "Cl^- simulation". The comparison of these conditions enabled us to investigate the transport of ions which cannot be resolved experimentally as well as electrostatic effects of elevated salt levels.

In contrast to previous solution simulations of PS II, [216, 217, 245, 215, 214] performing MD simulations in the crystalline state enables the construction of MD electron density maps that can be directly compared to the room temperature XFEL diffraction data [118, 206]. Such a comparison has recently been carried out by Wych et al. on the catalytic subunit of mouse protein kinase A (PKA C) leading to a new solvent model and an improved protein structure including alternative conformations [248]. The construction of electron density maps follows the procedure outlined in Wall et al. [218]. In addition to calculating total

density maps for the waters in the system, we decomposed the water density into hydrogen and oxygen components. This helps us infer information about hydrogen bonding and water dynamics in the channels that have not been accessible previously. We developed a new technique called MADI (Map based Acceptor Donor Identification) for this purpose. The MADI analysis yields information about both the strength and directionality of the inferred hydrogen bonds leading to a clear picture of the network connecting the Mn cluster to the bulk. Using our methods, we are able to make several key observations pertaining to the water transport and hydrogen bond network and to connect them to the functioning of the PS II enzyme.

6.3.3 Methods

6.3.4 Molecular Model and Force Field Parameters

The protein crystal MD model was prepared from the X-ray diffraction structure of PS II from Thermosynechococcus vestitus (PDB ID Code 7RF2), collected at the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory and SACLA in Japan. Since our S1 state data has a resolution of 2.08 Å, water positions could be modeled in the electron density with high accuracy. Details of the refinement have been described elsewhere. [118] Starting from the crystal structure, we used the P212121 space group to expand the asymmetric unit to the P1 unit cell with UCSF Chimera, resulting in four copies of the PS II dimer in the unit cell (with unit cell dimensions a = 116.92, b = 221.63, c = 307.83 Å, $\alpha = \beta = \gamma = 90°$). The coordinates for 13 types of nonstandard lipids/cofactors/ions were extracted from the expanded PDB file, and positions for their hydrogen atoms were generated using the LEaP program. [249] Molecular model information (.mol2) and force field modification files (.frcmod) from Sakashita et al. [214] were used for a majority of the lipids/cofactors/ions (OEX, PHO, HEM, LHG, LMG, PL9, BCR, CLA, DGD, W1, W4) to generate the topology (.itp) and structure (.gro) files. For the remaining lipids/cofactors/ions (SQD, STE, FE, CL), .mol2 files were generated using UCSF Chimera and parameterized with parmchk tools [249, 250] (see Tables 6.1 - 6.3). The topology file for the OEC was based on the one used by Sakashita, [214] however, a few modifications were made in order to preserve the geometry of the Mn cluster. Specifically, the bond lengths were adjusted to match those in our S1 state structure [69] and the angle force constants were increased by a factor of 4.17 to prevent the Mn4 atom from drifting.

No nonstandard protonation states were used in this work. For a table specifying net charges on all lipids/cofactors/ions, see Table 6.4. Additionally, relevant topology files can be found at https://github.com/mdoyle17/Molecular Dynamics Analysis in Python. Protein residues were parameterized with the AMBER14SB force field. Hydrogens for the protein atoms were added through GROMACS by pdb2gmx. The parameterized system was then solvated with TIP3P waters using gmx solvate in GROMACS. Note that this solvent is in addition to the crystallographic waters which were already present in the model. The four waters ligated to the Mn_4CaO_5 cluster (W1, W2, W3, W4) were also modeled as

H2O; however, they were treated as distinct ligands, and their residues were renamed to WON, WTW, WTH, WFO. They were not grouped in with the rest of the crystal waters so that the option of varying their protonation states remained open in the future. It has been reported previously that crystalline simulations with protein and ligand restraints to the crystal structure allow for a good recall of crystallographic waters [218]. With this information in mind, our protocol ensures that all heavy protein and ligand atoms are biased toward the crystal structure using harmonic restraints. The restraints are applied using springs that tether each atom to its position in the crystal structure. While the atoms are biased toward the crystal structure, they are still allowed to move including adopting alternative conformations of side chains if sufficiently favored by the MD force field. All ligand atoms have springs with a strength of 209.2 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ associated with them which is relatively moderate, and all protein heavy atoms (anything but hydrogen) have springs with a strength of 1000 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ applied to all heavy atoms to bias them toward the crystal structure. With the exception of the four waters ligated to the OEC (W1, W2, W3, W4) which experienced the same moderate restraint strength as other ligand atoms, all water atoms were unrestrained, enabling insights into water exchange dynamics from the results, as discussed in detail in the later sections.

6.3.5 MD Simulations

GROMACS version 2018 was used for the MD simulations. [251] To avoid errors in computing mean structure factors due to changes in the unit cell, the simulations were performed using a constant volume (NVT ensemble), with fixed periodic box size. NVT simulation was performed using the Cori Supercomputer at NERSC using 20 nodes, 32 tasks per node at a rate of 11.668 ns simulation time per day of wall clock time. Our simulation was confined to a box with dimensions a = 117.56, b = 222.88, c = 309.90 Å, and $\alpha = \beta = \gamma = 90°$, with the purpose of obtaining close packing conditions like those observed in crystals, resulting in a total volume of 8120 $\mathrm{nm^3}$. The simulation box was chosen to be slightly larger than the unit cell of the crystal to provide a small padding at the boundaries. Our buffer for this simulation consisted of just Na^+ ions. We began by solvating the void volume of the crystal with 135,577 solvent molecules followed by 479 Na+ ions with the purpose of neutralizing the net charge on the protein. This was necessary as the particle mesh Ewald (PME) method, which was employed for computing long range electrostatic interactions, requires charge neutrality for an accurate calculation[252, 253]. The water and ions were grouped separately from the protein and ligands for temperature coupling, as the former tends to generate more heating by force and integration errors [251].

Next, an iterative solvent addition procedure for achieving near atmospheric pressure in crystalline NVT simulations was applied. This involves rounds of energy minimization via the steepest descent algorithm followed by a 250 ps NVT simulation with simulated annealing, and then more solvation, until a mean pressure between -100 and 100 bar was achieved [218]. The final model contained 861,894 atoms (81,471 of which belonged to lipids/ions/cofactors).

A detailed breakdown of these atoms and their preparation may be found in Tables 6.1 − 6.3, along with the copies per lipid/cofactor/ion.

To carry out the simulation, leap frog integration with a time step of 1 fs was implemented. The output frequency for coordinates, velocities, and forces was set to 2 ps. Velocities were assigned from a Maxwell Boltzmann distribution at 300 K. The Lincs constraint algorithm, which resets bonds to their correct lengths after an unconstrained update, was used along with a Verlet cutoff scheme [251]. Van der Waals and electrostatic cutoff distances were set to 1.4 nm. Periodic boundary conditions (PBC) were also employed to simulate an infinite crystalline system. Artificial correlations across the unit cell boundary exist in such a simulation, but here we are mainly interested in using the simulations to compute the mean electron density, which we do not expect to be very sensitive to these correlations. Accurate studies of the large scale dynamics would need to involve simulations of a 2 × 2 × 2 supercell or larger system, [239] which is beyond the scope of the present study. In total, the simulation was run for 100 ns: The first 45 ns of which were used to ensure proper equilibration of the protein/lipids/cofactors with the solvent and thus were not used for further downstream processing. The trajectory from the final 55 ns of the simulation was extracted for statistical analysis.

A second simulation was carried out using a NaCl buffer instead of simply neutralizing the simulation volume using Na^+ ions. For this, 489 waters were first replaced by Cl ions in order to achieve a 0.1M Cl concentration, just like in the experimental buffer. Next, 968 additional waters were replaced by Na^+ ions in order to balance the charge of the system (note that this now includes balancing the original net negative charge on the protein as well as the negative charge introduced through the additional Cl^- ions). Everything else remained the same and the new simulation was carried out for 100 ns. For identification purposes, this is referred to as "the Cl simulation."

6.3.6 Mean Structure Factors

Mean structure factors FMD were computed to 2.0 Å resolution for the last 55 ns of the MD trajectory and output as .mtz files using methods described previously, [239] using xtraj.py in the Lunus software suite available at http://github.com/lanl/lunus. To summarize, cctbx methods are employed [224] to enumerate over N snapshots. At each snapshot i, the $xray.structure.select()$ method generates a new structure from only the selected atoms and then the $xray.structure.structure_{\text{factor()}}$ method is used to calculate its structure factor F. Electron density maps were then calculated by averaging these structure factors and computing the fast Fourier transform (FFT). This process was done individually for all waters, water oxygens, and water hydrogens. Absolute scale maps were generated from each MD.mtz file by scaling the maps using cell volume and F000 of the selection when using the fft function from CCP4's program suite [241]. The result is an average electron density map for specific atom or molecule types.

6.3.7 Experimental XFEL Crystallography Data

The crystallography data for the dark stable S_1 state (PDB ID 7RF2) was collected at the MFX instrument of LCLS at the SLAC National Laboratory, Stanford, CA. The data was collected using a 40 fs X-ray pulse length at 9.5 keV with a pulse energy ranging between 2 and 4 mJ. The data was collected on the Rayonix 340HS detector with a 3 $\times$ 3 binning mode and with a 3 μm X-ray spot size. The Drop on Tape method was used for sample delivery [110].

The XFEL data was processed with the DIALS software using the programs *dials.stills.process* for initial indexing/integration (using a target unit cell of a = 117.0 Å, b = 221.0 Å, c = 309.0 Å, $\alpha = \beta = \gamma = 90°$, space group P212121), *cctbx.xfel.stripe.experiment* for ensemble refinement of the indexed lattices including unit cell and orientation refinement and finally cxi.merge for merging of the integrated reflections. Lattices that did not reflect beyond 3 Å were rejected. Images were integrated to the edge of the detector and a per image resolution cutoff was used in the merging step. In total, 11734 lattices were merged to 2.08 Å (Final unit cell of a = 116.9 Å, b = 221.6 Å, c = 307.8 Å, $\alpha = \beta = \gamma = 90°$, space group P212121).

Using the obtained merged intensities, initial model building was done using the high resolution isomorphous 7RF1 structure as a template (1.89 Å) with rigid body refinement using phenix.refine. Subsequently, several alternate rounds of coordinate and B factor refinement were carried out. Certain side chain/backbone regions were also fixed in Coot using real space refinement. Custom restraints for the OEC were used that were based on available spectroscopic and chemical data. Custom restraints were also used for the Chlorophyll a (CLA) and lipid molecules (STE). Waters were added in the later stages of refinement using the default criterion in phenix.refine as well as manually adding waters in certain regions that were identified using the 7RF1 structure. The reliability of water positions was checked by calculating polder omit maps (phenix.polder) and evaluating the peak height positions in those maps. Waters with peak height below 3σ were removed from the model. Final R work/R free values were 18.52/23.85 %. Full details of the data collection and refinement for PDB ID 7RF2 are available in Table S2 of Hussein et al. [206]. We also refer the reader to Table S3 of the same paper in order to match the water numbering used here with that in the deposited structure.

6.3.8 Results

We developed an MD model for crystalline PS II in the dark resting S_1 state. The model consists of a single periodic P1 unit cell with four copies of the PS II dimer (one of the four dimers is shown in Figure 6.1a). Our protocol, which was developed initially by Wall et al., [218] allows for a direct comparison of the electron density between the simulation and the experimental room temperature XFEL crystallography data. The procedure that was followed for the analysis of the water and proton channels in PS II is summarized in Figure 6.2 and is briefly described here. The scattering contribution of each atom (including hydrogens)

in the simulation box (i.e., unit cell) was superimposed to calculate an electron density map in the symmetry setting for the crystal structure (Figure 6.10a, and see the Supporting Information for details). The electron density calculation was done using multiple snapshots of atomic coordinates extracted from a range of simulation time thus yielding an average electron density map that can be compared directly with the electron density map of the XFEL room temperature crystallography data. To yield a satisfactory visual comparison between MD and experimental maps, we show all MD maps in this study at a level of 2.08 e/Å^3, unless explicitly stated otherwise (see Figure 6.9).

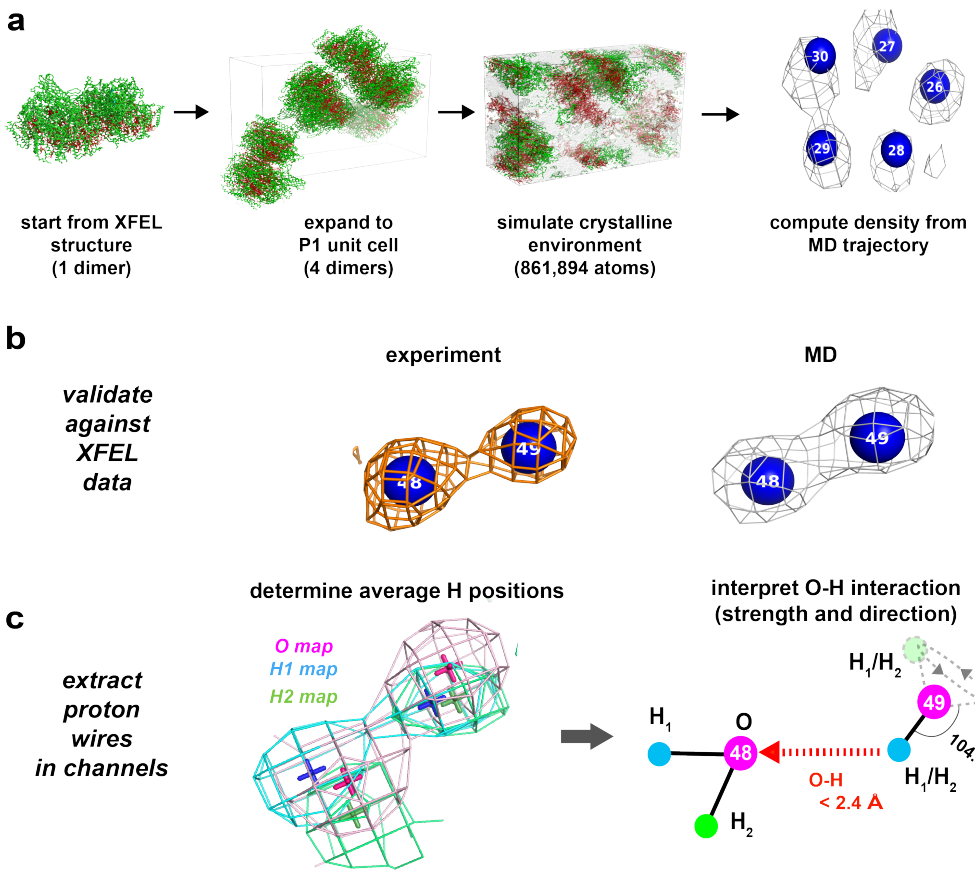


Figure 6.2: Methodological approaches used in this work. (a) The protein structure of PS II, which crystallizes in the P212121 space group, is first collected at an XFEL. The asymmetric unit is then expanded to the full P1 unit cell. Periodic boundary conditions are applied, wrapping the protein coordinates around the simulation box. The simulation is then carried out in a box approximately the size of the unit cell, with waters shown in light gray, protein atoms shown in green, and lipid/cofactor atoms shown in red. After carrying out the simulation, electron density maps are computed from the trajectory. (b) Comparison of experimental density (1σ / 0.173 e/Å^3) to the simulated water density (2.08 e/Å^3) is first required to make sure the XFEL electron density is reproduced with high fidelity. (c) Once the validation step is complete, simulated oxygen (purple, shown at 2.08 e/Å^3) and hydrogen (H1, cyan; H2, green, both shown at 0.26 e/Å^3 see the Supporting Information) densities are computed for the hydrogen bonding analysis. Average simulated oxygen and hydrogen peak positions, extracted from the densities, are marked as crosses. An example schematic, demonstrating our interpretation of the hydrogen and oxygen maps around W48 and W49, is shown on the far right. W48 and the H1 and H2 associated with it are likely fairly rigid over the course of the simulation, while the H1 and H2 associated with W49 are likely more mobile and are swapping into and out of the hydrogen bonding position. The nonbonded hydrogen can sit anywhere on the dotted cone.

After an initial MD simulation that included a 45 ns equilibration run, a 55 ns production run was carried out for analysis. Mean structure factors and electron densities were computed from this trajectory and analyzed to investigate the water structure and hydrogen bonding networks in the channels surrounding the OEC. For simplicity, we focus our discussion of the analysis on one of the two monomers in the PS II dimer unit, the monomer referred to as "unlocked" (so called because the subunits of this monomer have less contact with the neighboring PS II dimer in the crystal packing) [254]. This monomer corresponds to chains annotated as uppercase in the deposited PDB 7RF2.

To gain insight into the simulated water structure around the active site, an average water MD density map was computed over the last 55 ns of the trajectory and compared to the crystallographic derived water electron densities. To obtain a detailed picture of the independent contribution of each water to the MD density, the isolated electron density maps for specific waters were also computed (Figure 6.2a,b). These maps allowed us to understand the origins of specific patterns in the simulated MD density and to gain insight into the mobility of the waters (see Figure 6.10). We start by validating the MD derived water density (which we will refer to as the MD density) against the crystallographic waters from the crystal structure obtained from XFEL data (Figure 6.2b).

To analyze the strength and orientation of the hydrogen bonding network from the simulations, we developed a novel analysis technique, called MADI: Map based Acceptor Donor Identification (see the Supporting Information). MADI analyzes MD hydrogen and oxygen maps to obtain information about mean simulated water positions and orientations. We note here that fixed point charge force fields like those used in this work lack higher order electrostatics terms needed to precisely describe hydrogen bonding interactions. Nonetheless, the mean field approach in these force fields has been successful, on average, in reproducing basic geometric properties of hydrogen bonding that are relevant to our study [255, 256]. The information from MADI is then used to identify putative hydrogen bonds and to produce a measure of their strengths (an example is shown in Figure 6.2c). Uniquely, MADI also provides a means for explicit donor and acceptor identification. This information is interesting to examine in the context of the previously hypothesized changes in the directions of donor acceptor pairs before and after proton transfer [215].

See Table 6.5 for a complete list of significant water wires identified by MADI.

6.3.9 MD Density and Comparison with XFEL Experimental Data

After calculating the MD density, we compared it to the crystallographic waters around the active site. In Figure 6.3a, the crystal waters surrounding the OEC are displayed along with the channels of interest. Out of the 222 crystallographic waters within a 25 Å radius from the OEC (centered on O1), 189 had a strong ($\geq$ 2.08 e/Å^3) MD water density peak within 1 Å of the crystallographic water positions (see Figure 6.11 for a visual of this ROI in the context of the full PS II dimer as well as the monomer/monomer interfacial region). The agreement in all three channels was generally better toward the interior of the protein

than toward the bulk. The O1 channel contains the most waters whose positions were not recovered by the simulation.

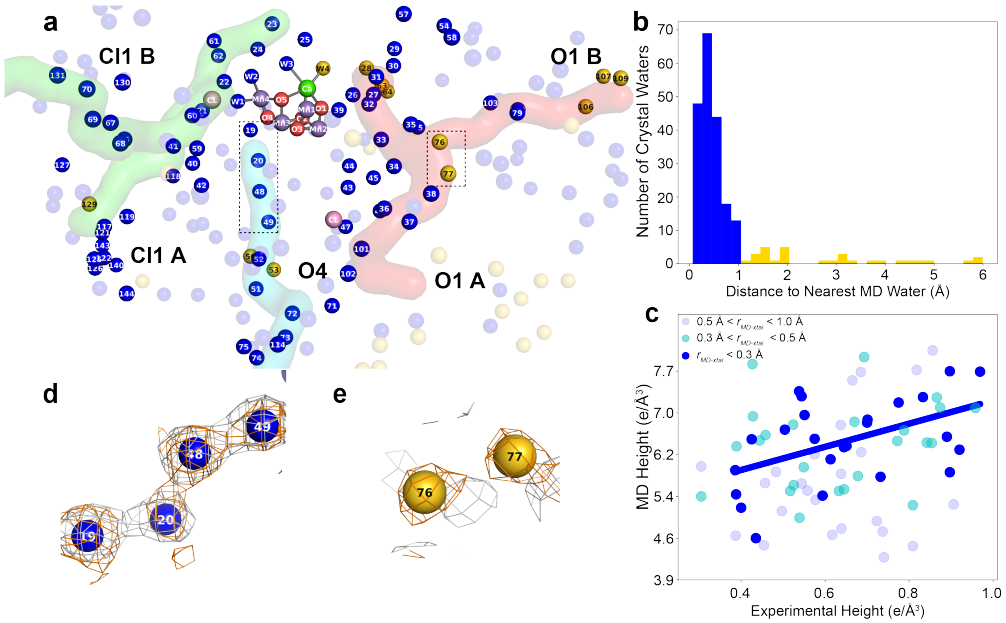


Figure 6.3: Comparing the MD derived electron density for waters with the crystallographic electron density. (a) Crystallographically modeled waters around the OEC, color coded by their relative agreement to the MD density. Blue colored waters are less than 1.0 Å from a peak in the simulated electron density map, while gold colored waters are more than 1.0 Å away from the nearest simulated peak. The three water channels of interest: The O1 channel (red), Cl channel (green), and O4 channel (blue) are also overlaid. (b) Distribution of distances between crystallographic water positions and MD water density peaks within a 25 Å radius from the OEC (centered on O1). (c) Scatterplot showing the correlation between the MD and experimental electron density peak heights for different groups of MD waters in the inner (blue), middle (cyan), or outer (light blue) shell. (d) Example of a region with close MD water peaks and corresponding crystallographic waters. The simulated MD water density (gray contour at 2.08 e/Å^3) is shown in the context of the crystallographic waters (numbered blue spheres). The experimental 2Fo Fc map is overlaid (orange) and contoured at 1σ. (e) Example of a region with distant MD water peaks and corresponding crystallographic waters. The simulated MD water density (gray and contoured at 2.08 e/Å^3) is shown in the context of the crystallographic waters (numbered yellow spheres). Experimental 2Fo−Fc map is overlaid (orange) and contoured at 1σ. The water numbering convention for all waters in the vicinity of the OEC has been described previously. [206] The numbers of waters are based on the 0F (S_1 state) experimental crystal structure used by Hussein et al. [206] (PDB ID Code 7RF2). The channels were mapped by Caver 3.0 Pymol plugin using this room temperature crystal structure [257, 258]

.

To compare the MD water density with the XFEL crystallographic data, we define $r_{\text{MD-xtal}}$

as the distance between the crystallographic waters and the nearest MD water density peak. The histogram in Figure 6.3b shows the distribution of $r_{\text{MD-xtal}}$. The average displacement is around 0.77 Å. Figure 6.3d,e illustrates how densities for both "close" ($r_{\text{MD-xtal}} < 1$ Å) and "distant" ($r_{\text{MD-xtal}} > 1$ Å) MD waters look in the context of the crystallographic waters and density; the visual agreement is consistent with the close vs distant classification.

We next analyzed the height of the peaks extracted from the simulated MD density. In Figure 6.3c, simulated peak heights are compared to the corresponding experimental peak heights. Their relationship was quantitatively analyzed by computing the Pearson correlation coefficient for three sets of MD water density peaks: those within the outer (0.5 Å $< r_{\text{MD-xtal}} <$ 1 Å), middle (0.3 $< r_{\text{MD-xtal}} <$ 0.5 Å) and inner ($r_{\text{MD-xtal}} <$ 0.3 Å) shell about the corresponding crystallographic water. The Pearson correlation coefficient is -0.014 for the outer peaks, 0.210 for the middle peaks, and 0.496 for the inner peaks. Thus, on average, the closer an MD water peak is to its corresponding crystallographic water, the stronger the agreement between experimental and simulated peak heights.

We compared simulated peaks located in each of the channels to ∼1320 randomly selected peaks found at least 28 Å from the OEC. We differentiated between inside and outside of the channels by using amino acid residue pairs to define channel boundaries (see Tables 6.6 - 6.8). Intrachannel waters exhibited greater overall peak height (mean peak height of 5.75 e/Å^3) than those which were closer to the bulk (mean peak height of 3.85 e/Å^3). Assuming a higher peak height corresponds to a lower mobility, this difference suggests that individual water positions within the channels are less variable in the simulation, allowing for density to build up to more elevated levels (see Figure 6.12).

As in previous crystalline MD studies where the protein is restrained to the crystal structure, [248, 218] the number of strong peaks in the MD water density map exceeds the number of crystallographic waters in our simulation. For example, there are 249 strong peaks in the MD density within a 20 Å radius of the OEC, compared to 161 waters in this region of the crystal structure. A recent unrestrained solution state MD study by Sirohiwal and Pantazis [259] of a single PS II monomer found 182 MD waters in this same region, using a different analysis method based on counting waters close to the protein every 105 ps. The increased number of peaks in our simulation compared to that of Sirohiwal and Pantazis is consistent with our previous finding that protein restraints can increase the number of strong peaks in the water density [218]. It is important to note that, although crystallographic waters can be matched to equivalent waters in unrestrained MD simulations by identifying the sites on the protein where they interact, [260] at present restraints are necessary for accurately recovering the positions of ordered waters using density maps, and that increasing the agreement when restraints are relaxed is a key strategy for improving MD force fields in the crystalline phase [218].

Using the simulated solvent map as a guide, we assessed the potential of using the MD water peaks to improve the water model in the crystal structure, similar to what was done in a previous MD study [248]. After remodeling six new waters into the deposited 0F (S1 state) structure (PDB ID 7RF2) based on peaks found in the MD solvent map, we observed visible improvement in the experimental 2Fo−Fc map. Two examples are shown, involving

newly solved waters W202 (occupancy of 0.59 and B factor of 29.01) and W205 (occupancy of 0.72 and B factor of 36.94) that only register subsequent to MD simulations (Figure 6.13). These results also show that there is scope for using the MD simulations to model additional waters with much weaker electron density within a 20 Å radius of the OEC. The six waters mentioned above were not added to our deposited XFEL model (PDB ID 7RF2) by the automatic water picking algorithm in the structural refinement software based on default threshold parameters. However, when they are manually modeled in and refined, there is enough electron density for these waters to be retained after filtering during automated refinement. Thus, our crystallographic data support the addition of six waters within a 20 Å radius of the OEC that were purely suggested by the MD simulations. This finding is consistent with a recent study reporting that MD simulations can be used to enhance crystal structures by locating additional ordered waters that are supported by the crystallographic data [248].

Next, we transition from a broad description of the MD density in the region surrounding the active site toward a more in depth, channel specific analysis. By using results from our simulated MD based electron density maps, we can explore water mobility and gain insight into hydrogen bonding networks within the O4, O1, and Cl1 channels.

6.3.10 O4 Channel

The O4 channel, a proposed proton channel, contains one main donor acceptor chain, which splits off at several points, as revealed by the MADI analysis (Figure 6.4a). The chain originates at W49 and proceeds toward the active site through W48−W20−W19, branching toward O4 (O4 is a bridging ligand in the Mn_4CaO_5 cluster, and the reason for the O4 channel nomenclature) at both W20 and W19, and also branching off toward ASP 61 at W48 and W19. The simulated waters in the neighborhood of the Mn_4CaO_5 active site more closely reproduce the crystallographic water positions, and have higher peak heights, relative to the waters farther up the channel.

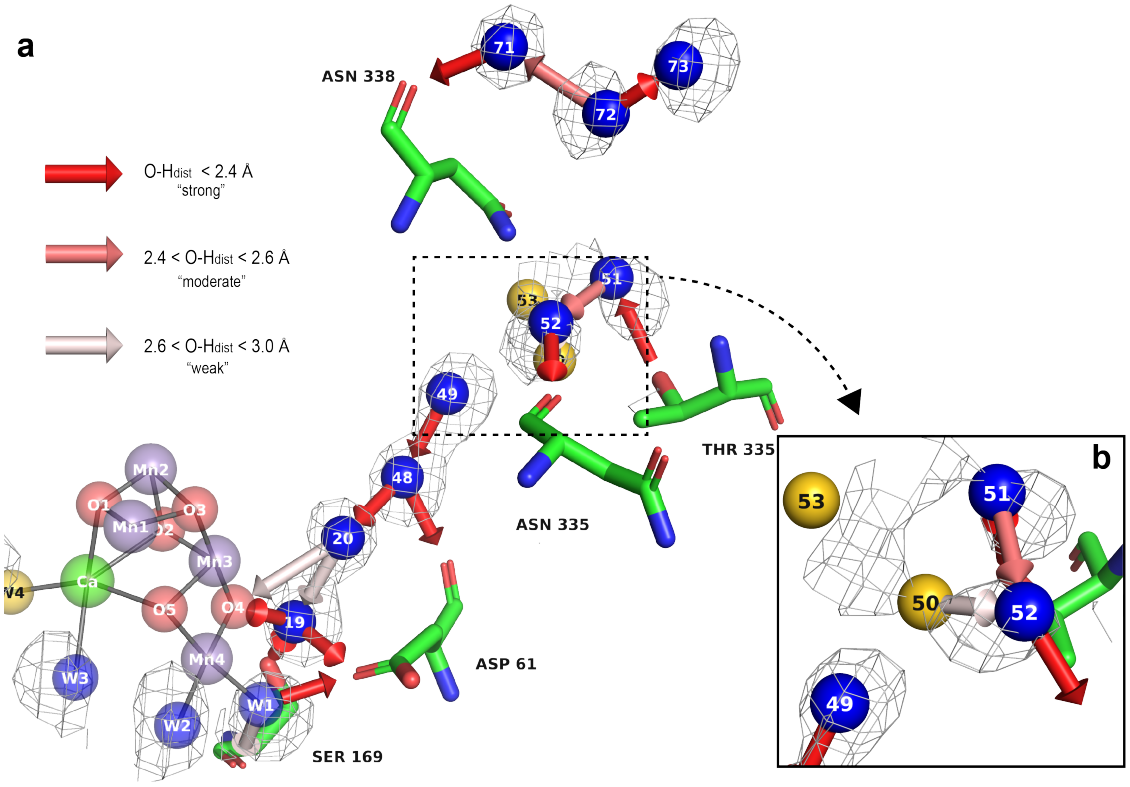


Figure 6.4: Hydrogen bond and water dynamics in the O4 channel. (a) Simulated MD solvent map in the O4 channel (gray, contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure (numbered, in color). The results from the hydrogen bonding analysis are depicted as arrows, pointing from the donor toward the acceptor. The strength of the bond identified by our algorithm is represented by a color gradient as defined in the legend. (b) Rotated view of the W50−W51−W52−W53 region, showing that crystallographic waters W50 and W53 lie outside of the MD map wireframe. All simulated maps were carved at a 2 Å radius about each atom. Oxygen and hydrogen peaks from the MADI analysis, demonstrating the average orientation of waters showing favorable hydrogen bonding geometries in this channel, can be found in Figure 6.16.

A dense cluster of seven strong bonds exists close to the OEC. This has the potential to create kinetic barriers to water mobility, leading to a buildup of density. This is demonstrated by the high average simulated water peak height of 6.79 e/Å^3 in the W19−W20−W48−W49 region (see Figure 6.14). This high value (relative to the average peak height value of 6.01 e/Å^3 associated with the 222 waters within a 25 Å radius of the active site, centered on O1), paired with the excellent agreement between the MD water density peaks and crystallographic water positions for this quartet of waters (mean $r_{\text{MD-xtal}}$ of 0.22 Å), suggests that the structure of these waters is quite rigid.

Moving toward the W50−W51−W52−W53 region (Figure 6.4b), the simulated water

density weakens (mean MD water density peak height of 4.62 e/Å^3) and a lower overlap between MD and crystallographic densities is observed (average $r_{\text{MD-xtal}}$ value of 1.31 Å). This supports the idea that there is (1) higher mobility and (2) greater conformational diversity sampled by the simulated waters in this region. The lack of an H bond connection to the rest of the main chain suggests that MD waters are moving more freely in this zone (see Figure 6.15).

Snapshots from the MD trajectory reveal rare instances when the gap between W51 and W72 may be transiently bridged (see Figure 6.17). Immediately before/after a molecule passes through this gap, however, we observed waters located in a commonly sampled position nearby W51 and ASN 338 (Figure 6.5). We hypothesize that this position is frequently occupied by waters transiting this crevice, as the buildup of additional simulated water density here is visible at 2.08 e/Å^3.

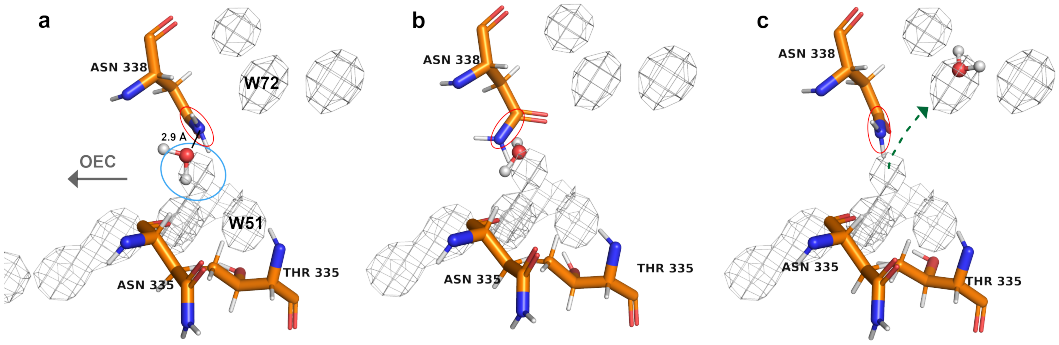


Figure 6.5: Drawbridge mechanism of ASN 338. Snapshots of the ASN structure from a single copy at (a) 90.94 ns, (b) 92.42 ns, and (c) 92.88 ns suggest that the ASN side chain movement is coupled with water transfer away from the OEC. The average solvent map from the entire trajectory across all four copies is overlaid and contoured at 2.08 e/Å^3. As the water molecule approaches ASN 338, it sits in the excess density, not found in the experimental map (blue circle). The side chain momentarily flips, potentially providing more room for the water molecule to hop over. After the water molecule crosses over and fills the W72 position, the drawbridge raises. It is observed to continue to raise and lower for the rest of the simulation.

The movement of waters across this gap appears to be correlated with the rotation of the nearby ASN 338 side chain. A solvent molecule was observed moving from the aforementioned excess density into the W72 position during the last 10 ns of the simulation after the side chain of ASN 338 rotates ~90°, resembling the "lowering of a drawbridge" (Figure 6.5a c). We postulate that this provides more room for a molecule to hop over. A similar phenomenon, involving a Na^+ ion, was also observed (see Figure 6.18). However, in our Cl

simulation, where Cl ions likely screened charged areas within the channels and reduce the long range electrostatic effects, no Na^+ ions were observed approaching the OEC via the O4 channel.

6.3.11 O1 Channel

Several consecutive, short, donor acceptor chains were identified by the MADI analysis in the O1 channel, a proposed water channel (O1 is a bridging ligand in the Mn_4CaO_5 cluster, and the reason for the O1 channel nomenclature) (Figure 6.6a). They were found in both the penta cluster of waters (also referred to as the water wheel [118]) by the Mn4CaO5 active site (W26−W27−W28−W29−W30) (Figure 6.6c), and around the main channel trunk (W31−W32−W39−W33−W34), although no chain was found directly connecting this channel to the OEC. Toward the ends of branches A and B, no chains were identified. The simulated waters in the trunk of the channel were found to have strong peak heights and showed overall better agreement with crystallographic waters relative to those closest to the active site and in the channel branches.

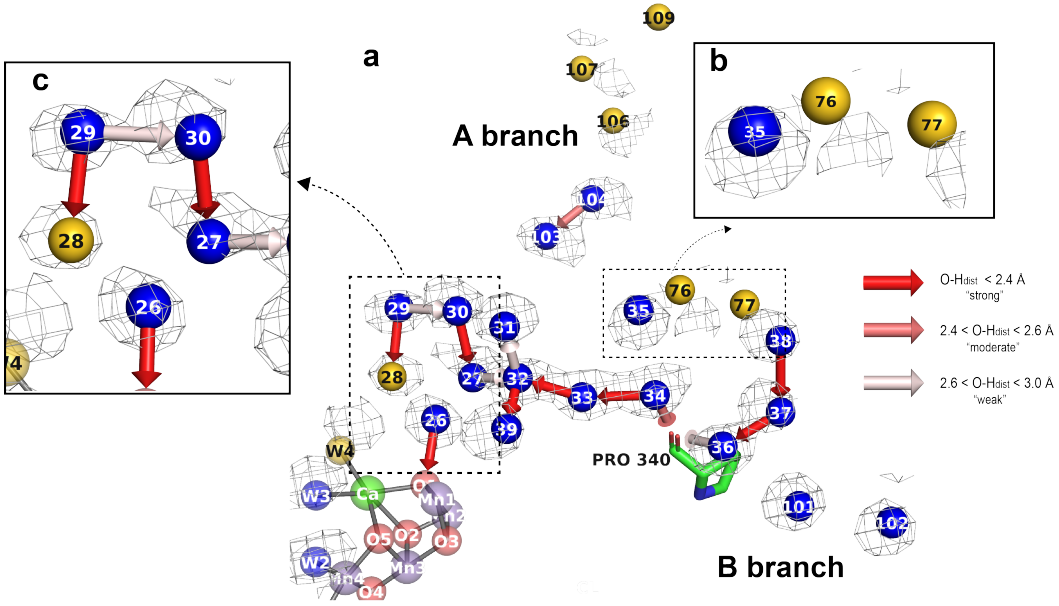


Figure 6.6: Hydrogen bond and water dynamics in the O1 channel. (a) Simulated MD solvent map in the O1 channel (contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure. The results from the hydrogen bonding analysis are depicted as arrows, pointing from donor toward acceptor. The strength of the bond identified by our algorithm is represented by a color gradient. (b) W76 and W77, where there is significant disagreement between the map and crystal structure. (c) Crystal water and MD map agreement/H bond network in the region of the penta cluster waters (W26−W27−W28−W29−W30). All simulated maps were carved in a 2 Å radius about each atom. Oxygen and hydrogen peaks from the MADI analysis, demonstrating the average orientation of waters showing favorable hydrogen bonding geometries in this channel, can be found in Figure 6.19.

The closest chain to the active site was found to originate in the penta cluster at W29, passing through W30−W27−W32 before splitting off between W31 and W39. The average $r_{MD-xtal}$ value of 0.61 Å in the penta cluster is nearly 3 times larger than what was found for the waters closest to the active site in the O4 channel. Moreover, the O1 waters close to the active site have, on average, comparatively smaller peak heights (5.83 e/Å^3) than those closest to the active site in the O4 channel (6.79 e/Å^3), suggesting that the O1 channel waters are less rigid over the course of the simulation.

Within the penta cluster, the individually calculated densities for W27 and W28 were the weakest, suggesting that they correspond to more readily exchangeable positions (see Figure 6.20). Additionally, both W28 and the nearby W4 are, on average, offset from the crystallographic positions by at least 1 Å. This is especially notable as W4 was moderately

restrained to its initial position like the other three waters ligated to the Mn4CaO5 cluster. Therefore, its movement, despite these applied restraints, indicates that the model in this region should be a focus of future optimization of the MD simulation.

As one moves into the trunk of the channel, the average $r_{\text{MD-xtal}}$ value improves to 0.40 Å, and the mean MD water peak height increases to 6.40 e/Å^3. However, individual water density analysis in this region indicates that these waters undergo fast and well coordinated movements (with water exchange occurring on <15 ns timescale), while simultaneously preserving well defined average positions (see Figure 6.10).

Toward the bulk water region, the average $r_{\text{MD-xtal}}$ value decreases and the simulated water density weakens in most areas. The largest deviation is for W76 and W77 (Figure 6.6b) which are on average 1.88 Å away from the crystallographically modeled positions, and exhibit a weak average peak height of 4.84 e/Å^3. Individually calculated water maps (Figure 6.21) indicate that mobility is relatively high in this zone, which might contribute to the decreased agreement with the crystal structure. MD waters toward the end of branch A also demonstrate greater average offset from the corresponding crystal positions (average $r_{\text{MD-xtal}}$ value is 1.0 Å) relative to the trunk region. They also correspond to a weak average peak height (5.19 e/Å^3), with no significant nearby donor acceptor chains.

6.3.12 Cl1 Channel

A long, unidirectional hydrogen bonded chain identified via MADI analysis was found in the Cl1 channel (Figure 6.7a), a proposed proton channel in the S_2 to S_3 transition, [206, 213, 261] originating from W3 and pointing away from the OEC via W23−W24−W2−W22 before forking toward ASN 181 and W21. From W21, it splits toward W1 and ASP 61 (Figure 6.7b). Overall, the Cl1 channel shows the best agreement between MD water density peaks and crystallographic waters (see Figure 6.22). MD waters in branch B demonstrated especially high agreement with the crystal structure and contained strong peaks.

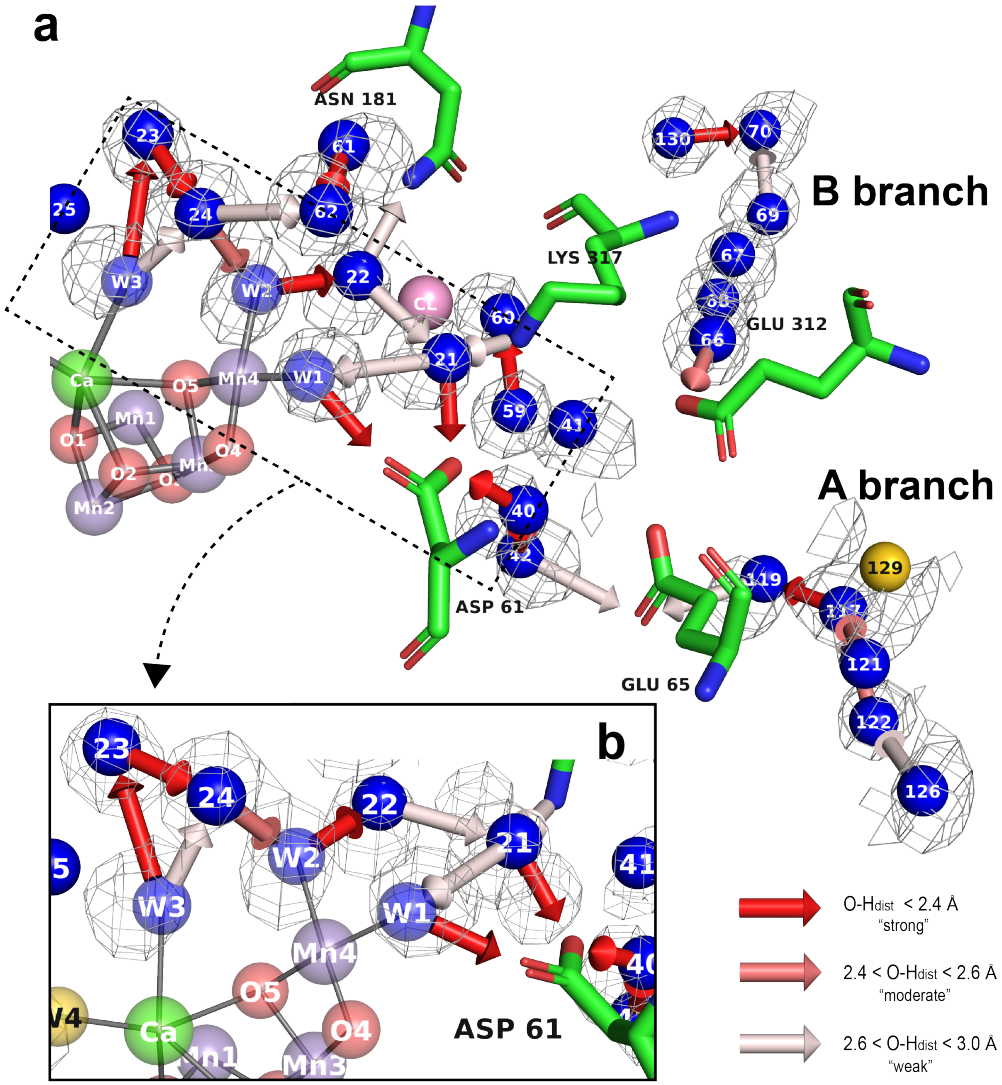


Figure 6.7: Hydrogen bond and water dynamics in the Cl1 channel. (a) Simulated MD solvent map in the Cl1 channel (contoured at 2.08 e/Å^3) along with identified hydrogen bonds, overlaid with the crystal structure. The results from the hydrogen bonding analysis are depicted as arrows, pointing from donor toward acceptor. The strength of the bond identified by our algorithm is represented by a color gradient. (b) Continuous hydrogen bond wire from W3 to ASP 61, identified through our MADI analysis. Only the core participating waters are shown. Peaks found from MADI analysis in this channel can be seen in Figure S15.

We postulate that the large number of nearby side chains, as well as anchoring effects provided by the participating ligand waters (W1, W2, and W3) in the OEC, play a role

in directing the long hydrogen bonded chain described above. As one moves further along the chain through W40 W41 W42, the average $r_{\text{MD-xtal}}$ value remains close (0.65 Å) and the average peak height remains high (6.24 e/Å^3). The MADI H bond directions do not form a unidirectional chain in this region, therefore no clear picture is provided regarding proton transfer in the S1 state structure we used in this study. We note, however, that a connection is still forged through ASP 61 and GLU 65 out toward the end of branch A. In our recent XFEL crystallography study of the S_2 to S_3 transition, it has been observed that this network appears to be poised to realign, via a rearrangement of the GLU 65/GLU 312 bottleneck region (proposed as a proton gate), enabling transfer of the proton out to the bulk via GLU 65, likely connected to a water deprotonation step upon a substrate water insertion at the OEC [206].

In the current MD model, the lack of a shared proton modeled between GLU 65 and GLU 312 relaxed the bottleneck so that its width at the final snapshot of the trajectory was >3 Å across all four PS II dimer copies. Since the protonation state of residues is expected to influence water movement, studying the effects of charge sharing between GLU 65/GLU 312 on hydration in the Cl1 channel will be a focus of future work. With our current model, additional locations with elevated water density in the MD density were observed both inside and outside this GLU 65 GLU 312 "proton gate" (see Figure 6.24). Other species such as Na+ ions were also observed moving toward this bottleneck region, likely drawn in by the negatively charged environment. However, no Na + ions were found moving into this negatively charged pocket from the Cl simulation. This difference might be attributed to alterations in the charge distribution surrounding the channel through adding additional Na + and Cl ions.

As one moves toward the bulk water region, a characteristic difference in mobility between branches is observed. Specifically, waters toward the end of branch B have very high agreement with the MD density (average rMD xtal value of 0.30 Å) and a strong average peak height of 6.61 e/Å^3. In branch A, however, the average rMD xtal value is 0.73 Åand the average peak height is only 5.54 e/Å^3. The trend between these two branches also aligns with the previous experimental B factor analysis. [206]

6.3.13 Discussion

Our study used calculated electron density maps computed from room temperature crystalline MD simulations to interpret the water dynamics of PS II and infer hydrogen bonding networks across three channels of interest in the dark resting S1 state, obtained from the room temperature serial femtosecond crystallography data measured at an XFEL. Building on a previous MD study of water structure in endoglucanase, [218]we applied the crystalline MD method to PS II, which is a substantially larger system (nearly 6 times as many atoms as the cited study).

We achieved a high level of reproducibility of crystallographic waters; within a 25 Å radius of the OEC (centered on O1), ∼85% of the crystallographic waters were recovered within 1 Å of their crystallographic position. There is a trend for the MD waters to have

higher electron density in the channel regions compared to waters in the bulk, which indicates that the water structure is more ordered within the channels (see Figure 6.12). We discuss below the relevance of our findings in the context of the function of the enzyme. Figure 6.8 summarizes our primary findings from this work to accompany the discussion.

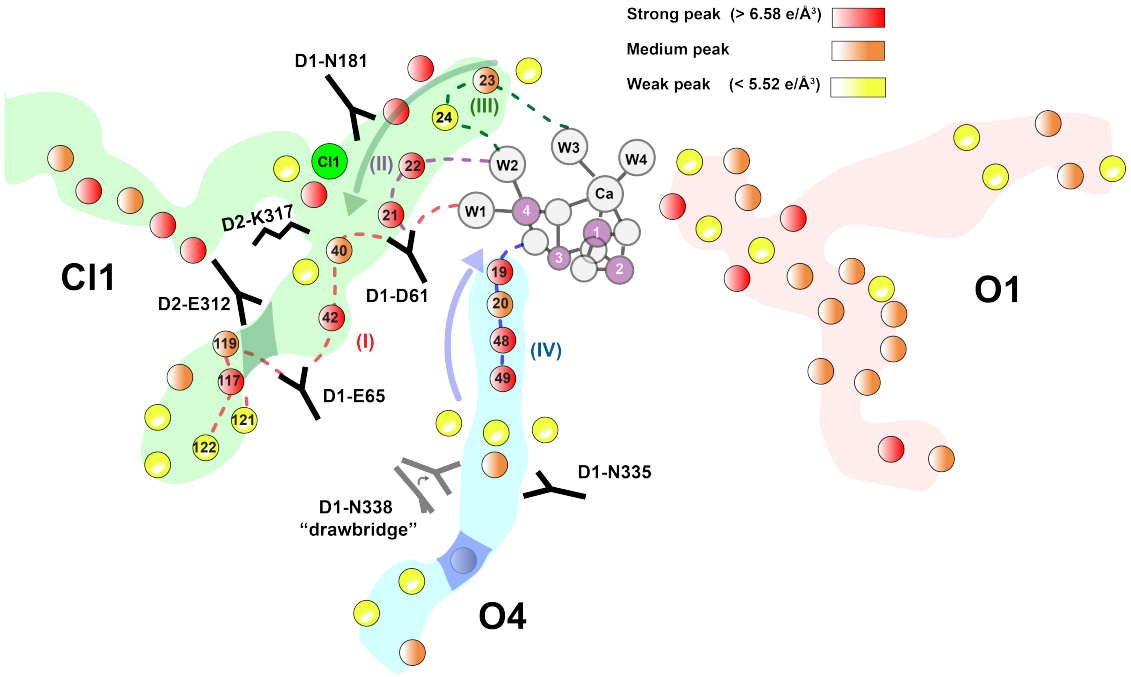


Figure 6.8: Overview schematic of potential proton pathways, with arrows showing the direction of the hydrogen bond networks inferred for the S1 state, along with highly dynamic regions in each of the three channels. We are interpreting the low density peaks (in the 33rd percentile for waters within a 25 Å radius of the active site, centered on O1) as corresponding to more mobile waters, and the high density peaks (above the 66th percentile for waters within a 25 Å radius of the active site, centered on O1) as corresponding to more rigid waters. Two previously discussed bottleneck regions in the Cl1 and O4 channels are shaded. Possible proton pathways from the OEC to the bulk are indicated by distinctly colored dashed lines, labeled I−IV. Selected waters which are thought to participate in each pathway are labeled based on the numbering convention described previously [206].

6.3.14 Networks for Proton Transfer

Given the fact that we are reporting on results derived from a classical MD simulation, we are unable to capture the quantum mechanical aspects of the proton transfer process. The intricacies involved in analyzing proton transport in biological systems, and specifically the tunneling processes that can govern such transport, have been described elsewhere [262, 263, 264]. However, this current work is simply focused on the dark resting S1 state of PS II with

no reaction activated. We thus conjecture about the proton transfer pathways used during the reaction cycle based on hydrogen bond network orientation and connectivity inferred in the S1 state.

Our MADI analysis provides us with information regarding the strength and directionality of the inferred hydrogen bond network. This information provides a way to consider the potential proton release pathways through the channels. We note that our results do not provide supporting evidence for a proton transfer pathway in the O1 channel. Unlike in the O4 and Cl1 channels, MADI did not reveal a significant hydrogen bond wire in the O1 channel which is connected to the cluster. Furthermore, as one moves out toward the ends of branches A and B of the O1 channel, very few hydrogen bonds are detected, making the potential path for proton transfer significantly unlikely. Disruption of the network in the O1 channel was also seen by Kaur et al. in a study which combined MD with Multi Conformation Continuum Electrostatics (MCCE) and Network Analysis. In their work, such disruption was observed 10 Å away from the OEC, and due to the high level of interconnectedness found around the active site, a released proton had potential to access all three channels. However, the O1 channel was not as likely to be the proton release path [265].

In the case of the O4 channel, we observe that the donor to acceptor direction is from the bulk to the OEC, similar to a post proton transfer orientation discussed by Sakashita et al [215] . It has been proposed from theoretical studies that the O4 channel is likely used for a proton release during the $S_0 \rightarrow S_1$ transition, [211, 128] which is supported by the data presented here as it is from the S1 state. For the Cl1 channel, the donor to acceptor orientations are mostly from the OEC to the bulk, a pre proton transfer orientation, which indicates the Cl1 channel is primed to release a proton upon a deprotonation event in a subsequent transition beyond the S1 state. This channel has been proposed previously as a proton release pathway during the $S_2 \rightarrow S_3$ transition based on pKa calculations of residues, FTIR spectroscopic studies, and more recently in MCCE simulations, [213, 261, 265] as well as in the snapshots of room temperature crystallography data taken during the $S_2 \rightarrow S_3$ transition [206]. In contrast to previous computational studies of the hydrogen bond network primarily in the O4 channel, our MADI analysis is able to determine the strength and direction of the inferred network in all of the channels.

For the Cl1 channel, we catalog multiple plausible pathways depending on the origin of the proton (W1, W2, or W3). A proton originating at W1 would be transferred to ASP 61 and donated to W21 or W40 after a flip in the H bond orientation of ASP 61. We note that W21 LYS 317 is a dead end, so the proton is most likely to travel through W40−W42−GLU 65 (path I in Figure 6.8). At this point, it can continue out toward the bulk via branch A.

A proton originating at W2 could move along W22−W21−ASP 61 (path II in Figure 6.8) before carrying on the same route as a W1 proton (we note that the W2 W22 ASN 181 pathway is a dead end). In the case of a proton starting at W3, the pathway would primarily use the W2 route after getting shuttled through W23 and W24 (path III in Figure 6.8). A recent paper by Allgöwer et al. suggests path III to be a strong candidate for proton transfer in the $S_2 \rightarrow S_3$ transition based on multiscale quantum and classical simulations [266].

Based on the MADI analysis (see Figure6.4), a proton released from the O4 channel would

likely originate from the O4 atom of the OEC before traveling up the quartet of rigidly fixed waters W19−W20−W48−W49 (see Figure 6.14) via a Grotthuss mechanism, which involves a free proton hopping from one oxygen to the next. The next steps are less clear and require the formation of multiple transient bonds in the mobile W50−W51−W52−W53 region, as well as between W51 and W72, before a proton can advance farther up the channel. The latter likely involves movement of ASN 338 as seen in our simulations (path IV in Figure 6.8).

6.3.15 Water Transport

Previous steered MD simulations carried out to study the energetics of water permeation in PS II have revealed that no channel permits absolutely unrestricted access of water to the OEC. [217] The idea that water transport is a controlled process because of the very oxidative potentials present at the catalytic Mn site, and which involves regulatory mechanisms to prevent bulk water from being present at the active site, makes studying water dynamics within the channels important.

In the present simulation, we found that waters in the Cl1 channel exhibited the best all around agreement with the crystallographic waters. We postulate that the Cl1 channel is relatively rigid (see Figure 6.22). For water transport to occur in a given channel, we hypothesize that it is necessary, at the very least, for the region directly adjacent to the active site to be favorable for water mobility. According to this hypothesis, the likelihood of water transport in the O4 channel would be low due to the four rigid waters (W19−W20−W48−W49, see Figure 6.14) close to the active site.

Close to the OEC, the simulated water peaks in the O1 channel were, on average, weaker than those in the O4 and Cl channels. The O1 channel is also unique in the sense that it is the only channel that contains a crystallographic water near the OEC that is distant from the nearest MD water density peak (W28). Dynamic pockets in this channel are also observed toward the end of branch A with W106, W107, and W109, and in the middle of the channel near W76 and W77.

By analyzing individually calculated water maps, we also observed that quick exchange can occur in the O1 channel trunk (see Figure 6.10) even though average MD water positions in this region are well preserved. From this observation, we conjecture that some waters are capable of simultaneously preserving a well defined average structure while exchanging through what are likely fast and well coordinated movements. It will be interesting to further study the connection between the strength of the MD water density and the frequency with which waters occupying a site are exchanged during the simulation.

The aforementioned observations support the idea that the O1 channel is dynamic in multiple regions throughout, making this channel a top candidate for the delivery of substrate water. This is consistent with our recent XFEL crystallography data which, based on time point snapshots during the S2 to S3 transition, suggested that the O1 channel waters are highly mobile, making it a strong candidate for facilitating water transport [206, 118].

Additionally, the O1 channel has been found to be highly accessible to bulk waters through a simulation based study carried out by Sakashita et al. [214]

The number of strong peaks in the MD water density exceeds the number of crystallographic waters as described earlier. This is expected, as in the experimental crystal structure weakly interacting and highly mobile waters may not be resolved due to the reduced occupancy of atoms in a given position. This is more so at room temperature than at cryogenic temperature because of the difference in mobility. The additional MD water density peaks can potentially be used to increase the number of waters modeled in the crystal structure, as was found in a study of cAMP dependent protein kinase. [248] Indeed, we found some isolated cases where the MD suggested reasonable additions of water molecules that were supported by the experimental data (see Figure 6.13). Sirohiwal and Pantazis also recently noted that the number of water molecules associated with PS II appeared to be higher in their MD simulation than in many PS II crystal structures. [69] While the authors concluded that the decreased number of waters is due to the dehydration of the crystals, the absence of water in a crystal structure does not imply that a region is dry since larger voids are to be interpreted as being filled with bulk solvent. In addition, the number of waters observed in the previous crystal structures [69, 118, 206, 210, 267, 268, 269] also differs due to the resolution and measurement temperature (cryogenic vs room temperature). Moreover, decreasing the detergent and water content in PS II crystals is a specific strategy that has been used to increase the resolution of the diffraction data, [270, 271] which tends to increase, not decrease, the number of ordered waters in the crystal structure, as shown in Table 1 by Sirohiwal and Pantazis. [259] We also note that the S state advancement shown in the XFEL data by Kern et al. proves there is no sample dehydration that would inactivate PS II and release Mn(II) [69]. Our findings therefore do not support the claim that the crystal structures are dehydrated compared to the MD simulations.

6.3.16 Conclusions

The results presented here have demonstrated how a crystalline MD simulation approach can be used to understand the role of water networks in the PS II enzyme. By calculating average electron density maps from the trajectory, we can compare experiment to simulation and identify potential mechanisms of water transport and networks for proton transfer. Our study also introduced the MADI approach for analysis of MD simulated electron density maps to identify potential water hydrogen bonding networks. The result indicates that the O4/Cl1 channel has multiple hydrogen bond wires emanating from the OEC toward the bulk but which need to be bridged in certain sections by the motion of waters/amino acids. We also found in the O1 channel that the inferred hydrogen bond network is disconnected from the bulk and instead contains several mobile regions, including regions where fast exchange occurs beneath a well structured average density. Thus, we postulate that the O4/Cl1 channels are most likely involved in proton transfer, with the Cl1 channel demonstrating the clearest path for a proton to travel out toward the bulk, while the O1 channel is likely involved in substrate water delivery.

While not a focus of the current study, crystalline MD also has the potential to resolve inadequacies in crystallographic modeling [248]. Better modeling of the solvent, accounting for differences in symmetry related copies and sampling of multiple conformations of mobile waters/amino acids are some of the improvements we can potentially make with crystalline MD.

Future MD investigations beyond the dark stable S1 state presented in this work and into the meta stable S states as well as time points in between would shed light on the functioning of these channels during specific transitions. This will require reparameterizations of the Mn cluster, using quantum mechanical calculations of electronic structures that differ from the S1 state studied here. Obtaining protonation states of critical amino acids from neutron diffraction studies is another approach to further improve the MD model [218].

6.3.17 Supporting Information

6.3.17.1 Methodology

6.3.17.2 Mean Structure Factors (Continued)

In order to deconvolute the MD density, individual water maps for selected channel waters were also computed. To do this, the expanded crystal structure in the P1 space group was used as a reference for the water positions in each copy. First, to make sure that the placement of each water was aligned for each of the four copies, the water of interest was selected across all copies and an Fcalc was computed just for the starting .pdb file. This map was then visualized to ensure that it only contained one peak, which confirmed that when the copies were folded back in on themselves, they overlapped. Next, in order to solely focus on dynamics that occurred in the 55 ns production run, a snapshot of the system immediately after equilibration (t = 45 ns) was taken. It is expected that during this 45 ns equilibration period, explicit waters will have rearranged themselves while still maintaining the overall water structure around the active site.

The updated water identities post equilibration were then mapped to the reference positions. For example: If a new water "W_X" exchanges with W19and fills the canonical W19 position during the 45 ns equilibration, then the residue id associated with "W_X" will be registered as W19 for the individual calculation. If no water at 45 ns is within 1 Angstrom of the canonical W19 position for a given copy, then no waters were selected for that copy when computing the average W_{19} map. Some of the individual maps which will be discussed are an average over all four copies. Some are an average over only two copies, etc. These computed electron density maps were then used to determine the average positions of water molecules and identify regions of potentially high water mobility. Additionally, a short ion analysis was conducted by also computing a map over all Na^+ ions.

It is important to note that these maps were computed by averaging over all four dimers: the structure factors were computed by feeding in the complete trajectory which occurred in the MD simulations that were performed using a P1 unit cell, and the structure factor calculations were performed by specifying the P212121 space group and folding all atoms

back into the asymmetric unit. The result is a mean density for the P212121 asymmetric unit. Strong peaks (above a threshold of 2.08 e/Å^3 for water molecules and oxygen atoms, and above 0.26e/Å^3 for hydrogen atoms the factor of eight difference comes from the fact that hydrogen has 8x fewer electrons than oxygen)were extracted from the simulated electron density maps using CCP4 peakmax, [241] and their x,y,z coordinates were output into .pdb files.

For comparison between MD and experiment, CCP4MOLREP [241] was used to align the crystal structure to the MD density computed over the entire system. 133 waters were randomly selected outside of the channels for comparison.

6.3.17.3 Comparing MD Map to Experimental Map

In order to compare both the MD and experimental maps at a level which still preserved interesting water details,the following protocol was used. First, an absolute scale (electrons per Å^3) map was generated for the experimental 2Fo Fc electron density. The absolute level of that display in electrons per Å^3 was noted at 1σ (0.173 e/Å^3). However, the full MD map contoured at the same level will not capture the shape and extent of the features we are interested in (aka, most individually resolved water densities). This is likely because the B factors are quite low in the simulation, and the theoretical maxima of the electron distribution partially depends on the B factor of the atoms. So, for visualization purposes, the complete MD map was adjusted by eye in multiples of 0.173 until their contour roughly matched that of the experimental map in several regions which are believed to be rigidly fixed (ARG 320 C, HIS 337 A, PHE 317 C) (see Figure 6.9). All MD maps are shown at a level of 2.08 e/Å^3 unless explicitly stated otherwise for visualization purposes. Any experimental maps are shown at 0.173 e/Å^3 (equivalent to 1σ).

6.3.17.4 MADI: Map Based Donor Acceptor Identification

In order to infer about PS II's hydrogen bonding networks, a new method called MADI (map based donor acceptor identification) was deployed. The method was developed to evaluate average hydrogen bond strength and direction between neighboring waters over the course of the trajectory: First, strong peaks (above a threshold of 2σ)were extracted from the following three MD simulation electron density maps using CCP4 peakmax, and their x,y,z coordinates were output into .pdb files

1. A map computed using all water oxygen atoms
2. A map computed using the first hydrogen atom in all waters (HW1 only)
3. A map computed using the second hydrogen atom in all waters (HW2 only)

Next, the list of peaks from the water oxygen map was looped through. If an oxygen peak was found within 1.5 Å of a crystallographic water, then that peak would be grouped with that water and labeled as that water's "average water oxygen position" over the course of

the simulation. The two hydrogen peak lists were looped through next. For each list, if a hydrogen peak was found to be within 1.4 Å of the average water oxygen position for a crystallographic water, then the hydrogen peak would also be "grouped" into that water position. The end result ideally involved one average oxygen and two average hydrogen peaks per crystallographic water (when looking at the 100 waters of focus, we identified 1.7 hydrogens per water on average) leading to an average water orientation for every crystallographic water over the course of the simulation. The above procedure yielded putative MD water orientations for many, but not all, of the crystallographic waters. Optimal O−H bonds were determined and classified by strength (see Figures 6.25 and 6.26.) To avoid pitfalls in interpretation, it was important to apply several layers of filtering. Simply looping through, checking average oxygen hydrogen distances between nearest neighbors, and considering it as satisfactory when an O−H distance was under a given threshold could bury important orientational information regarding the direction of the bond and the identity of acceptor vs donor. See Figure 6.27 for a flowchart which details the filtering logic employed in the hydrogen bonding analysis.

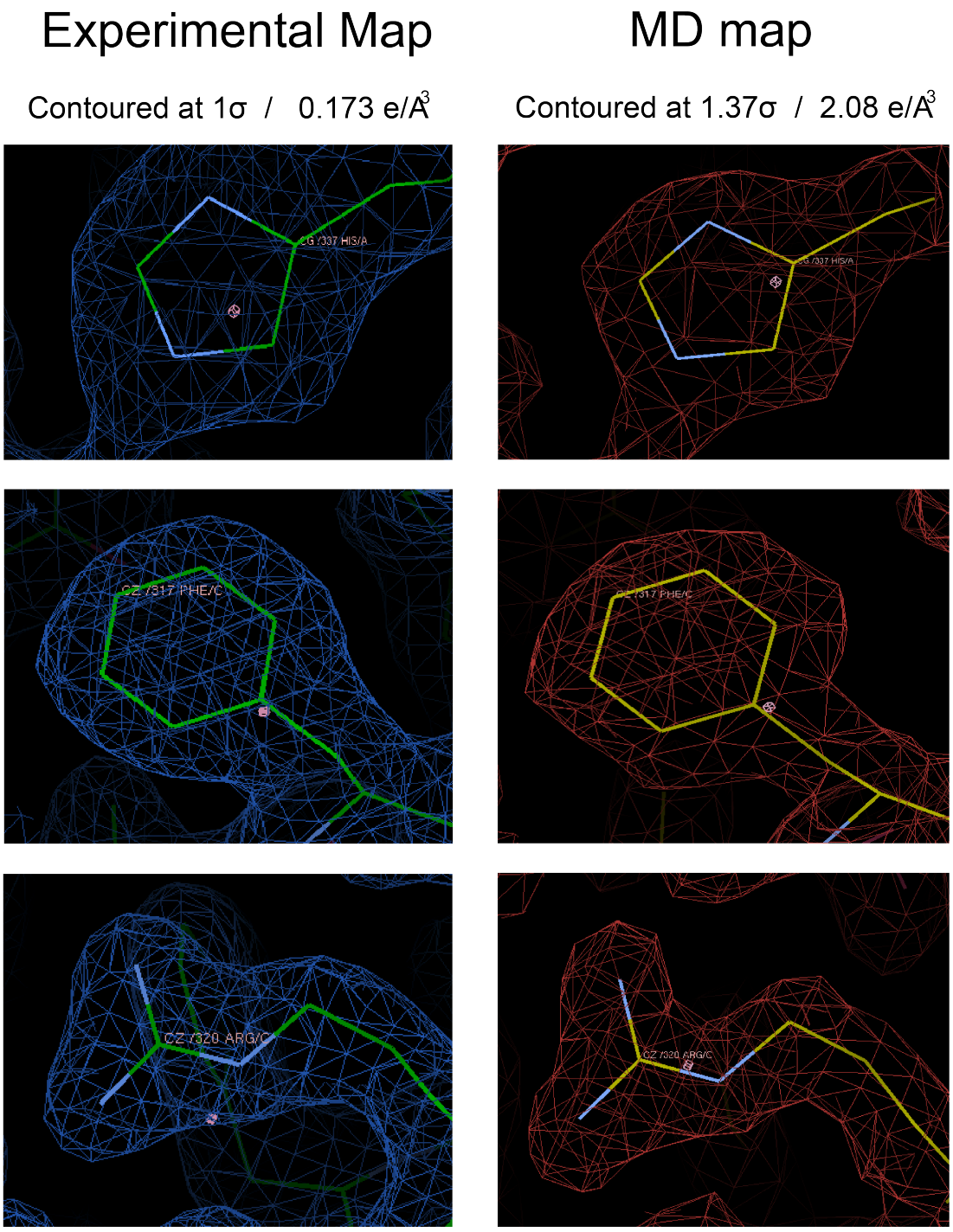


Figure 6.9: Converging to comparable maps between the experimental $2Fo-Fc$ crystallographic densities (left, blue, contoured at 0.173 e/Å^3 which is equivalent to 1σ), and the MD densities (right, contoured at 2.08 e/Å^3). Images shown in Coot[272].

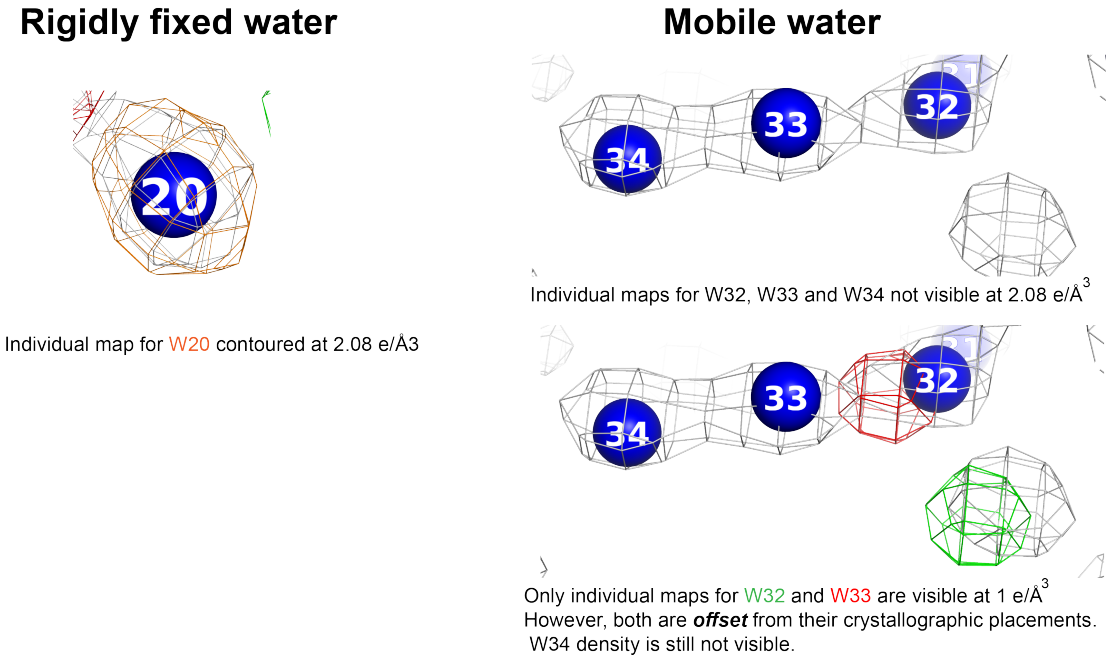


Figure 6.10: Individually calculated water maps (colored) overlaid with the total average water map across all copies (grey) to demonstrate how mobility is classified. (Left) Example of how the individual water map looks for a rigidly fixed water. W20's individual map, which shows up as a visible and strong sphere of density, is comparable to the total water map when both are at the same 2.08 e/Å^3 level. (Right) On the other hand, the individual maps for more mobile waters like W32 W33 W34 are not visible at 2.08 e/Å^3, since they are not in any location long enough over the course of the simulation to build up substantial density. At 1 e/Å^3, individual maps for W32 and W33 become visible (and are offset from their crystallographic placements, with the density for W33 sitting closer to W32 and the density for W32 sitting in the W39 position). The density for W34, on the other hand, remains still too weak to be seen at this level due to extreme mobility. It only becomes visible at levels $<$ 0.7 e/Å^3 .

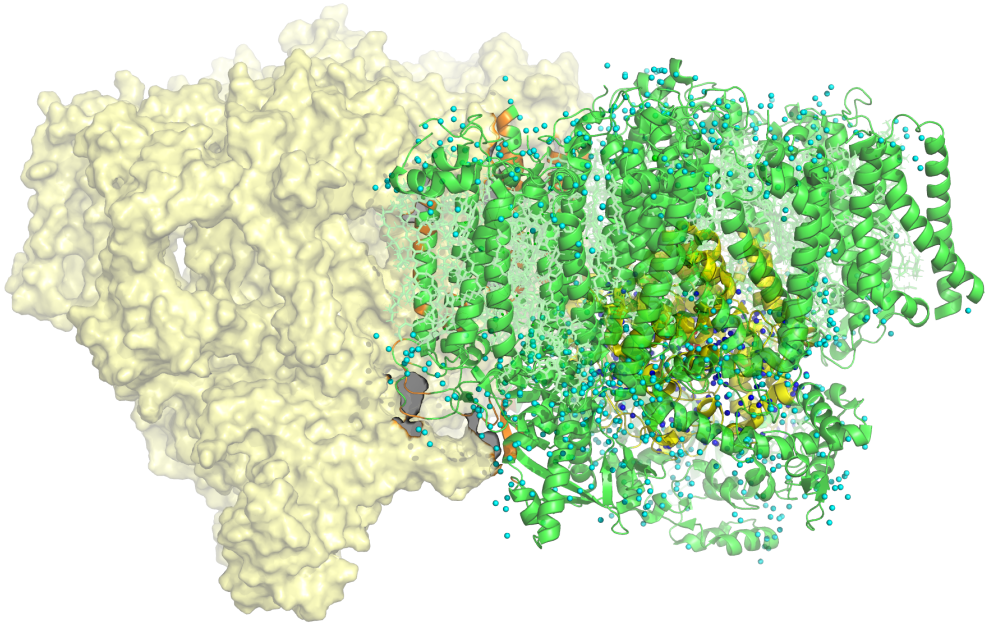


Figure 6.11: The full PS II dimer (PDB ID **7RF2**) which was used as the basis for our MD model. The ROI referred to throughout the text (centered on the O1 atom in the OEC with a 25 Å radius) is shown (yellow) along with its enclosed waters (dark blue). Note that the orientation of the dimer is flipped relative to how it is displayed in **Figure 6.1**. All outside waters are shown in cyan. The 'unlocked' monomer, which is the focus of this study, is shown in green. The 'locked' monomer [254] is shown as a sand colored surface. The orange region corresponds to contacts within 3.5 Å of the monomer monomer interface extended to the entire residue/ligand.

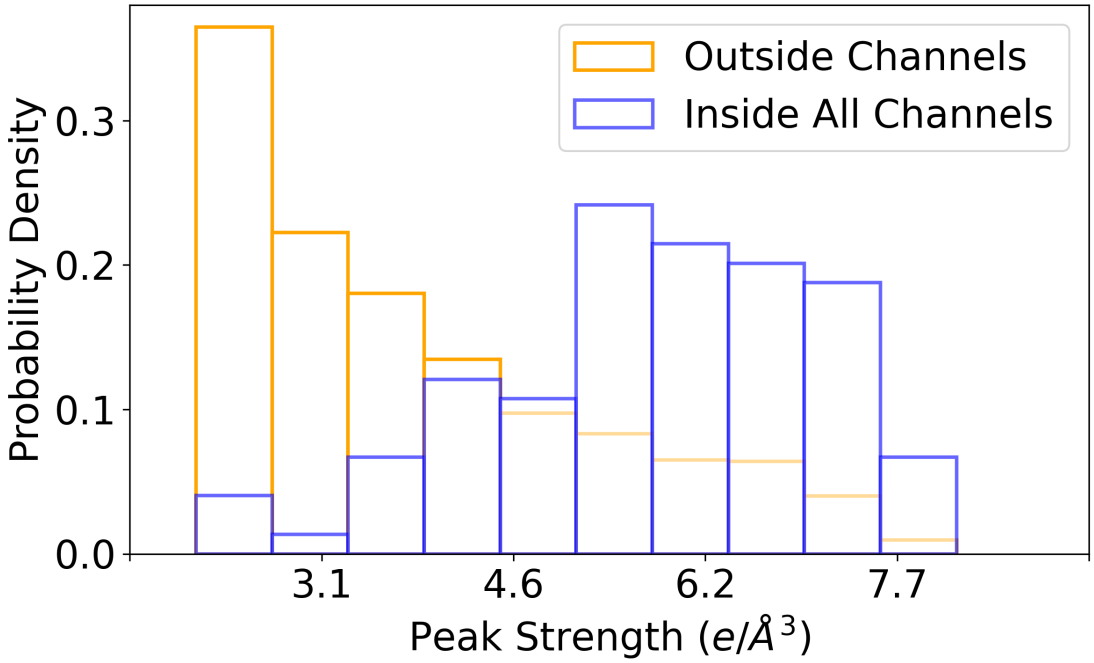


Figure 6.12: Histogram comparing the distribution of peak height of simulated waters inside the three channels of interest (O1, Cl and O4) vs. outside the channels. On average, waters inside the channels demonstrated overall stronger peaks. Such a finding suggests that positioning of waters within the channels is more anchored over the course of the simulation. Each bin corresponds to raw count divided by the total number of counts as well as the bin width. As a result, the area under each histogram is normalized to one thus providing a probability density.

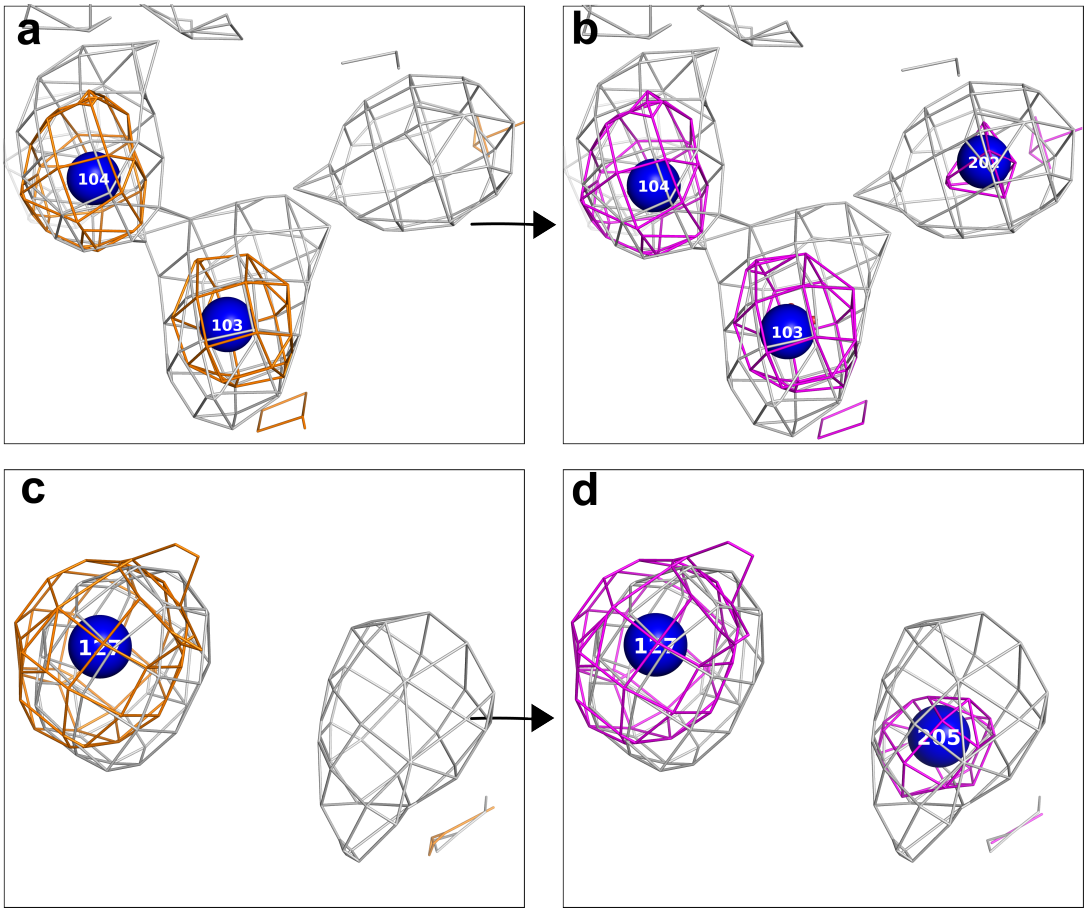


Figure 6.13: Demonstration of how crystallographic refinement can be improved through the combination of MD. (a) MD solvent map (grey and contoured at 2.08 e/Å^3) overlaid with the experimental 2Fo−Fc map (orange and contoured at 1 σ). (b) After an additional round of MD aided refinement, a new water, W202, was registered in this region (updated 2Fo Fc map is colored magenta and contoured at 1σ). (c) An additional MD solvent density, not found in the experimental map, appears near W127. (d) After an additional round of MD aided refinement, yet another new water, W205, was registered in this region.

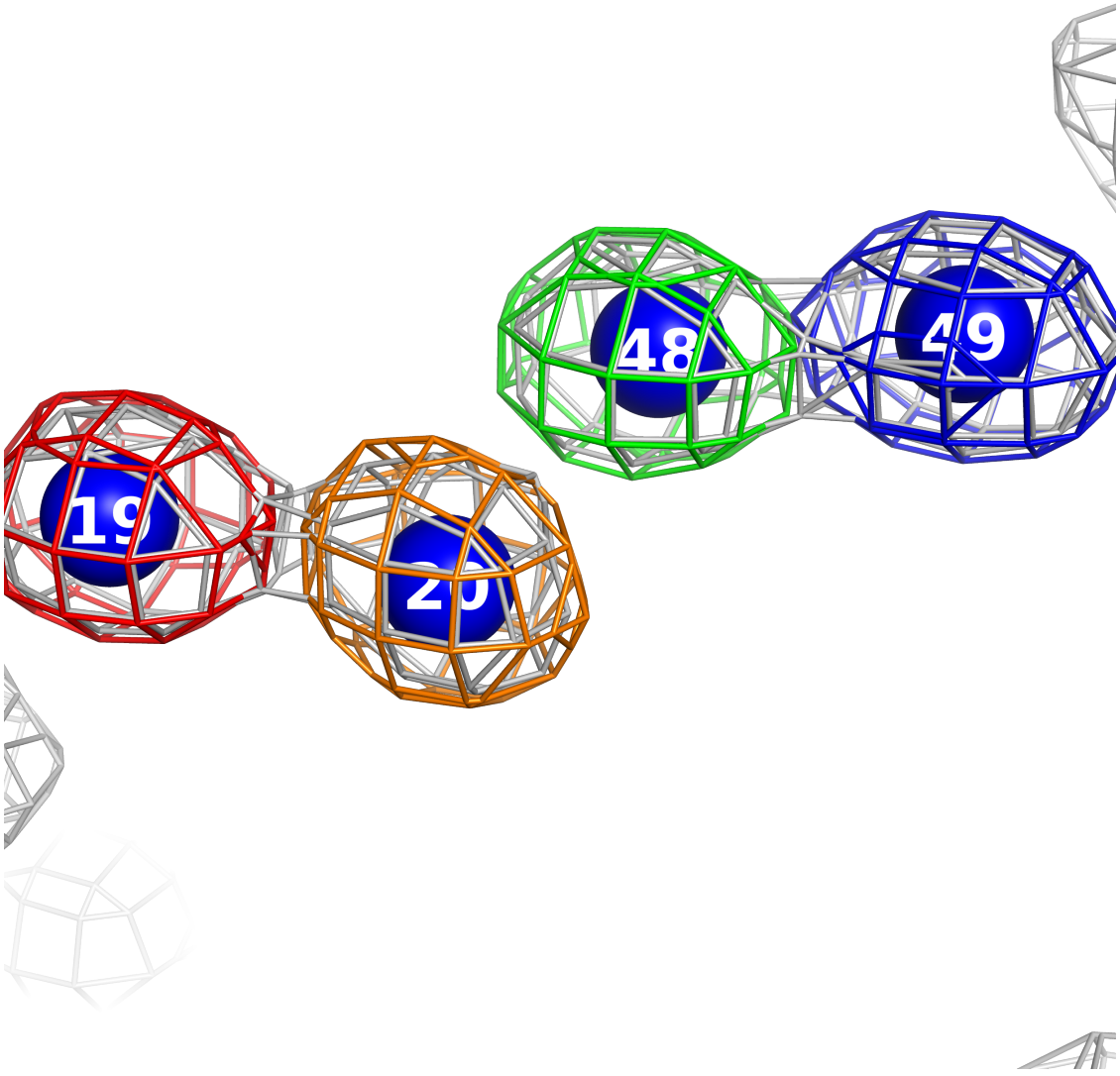


Figure 6.14: Individually calculated water maps (colored) overlaid with the total water map (grey) in the O4 channel. All contoured at 2.08 e/Å^3.

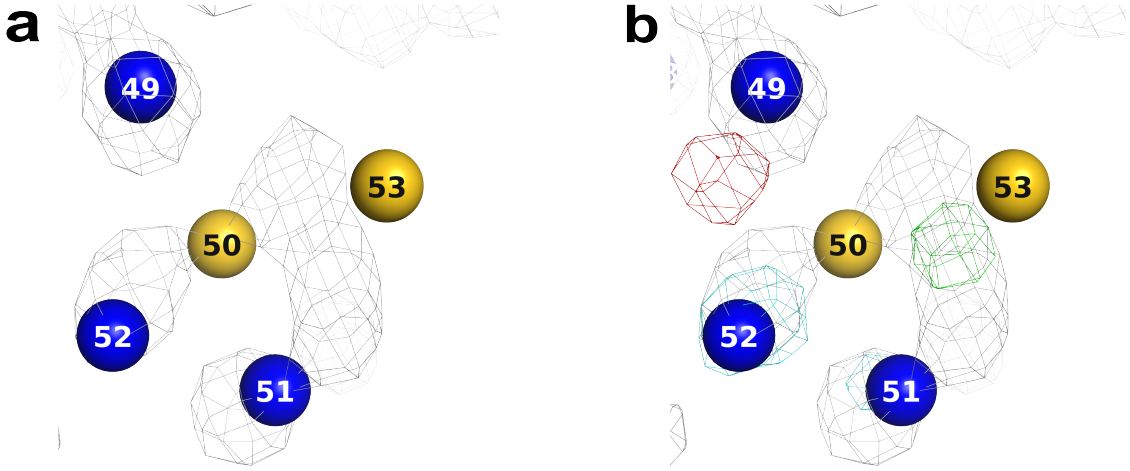


Figure 6.15: Individually calculated water maps (colored) in the region of W50−W51−W52−W53 overlaid with the total water map (grey). (a) All maps contoured at 2.08 e/Å^3, yet only the full water map is visible at this level. (b) Individual maps for W50 (red), W51 (green) and W52 (cyan) contoured at 1 e/Å^3 for better visualization, while the total water map remains contoured at 2.08 e/Å^3 . Dynamic tendencies are observed in this region. Note that no individual map was calculated for W53 as there was no equivalent water near its position after the 45 ns equilibration (see Methods).

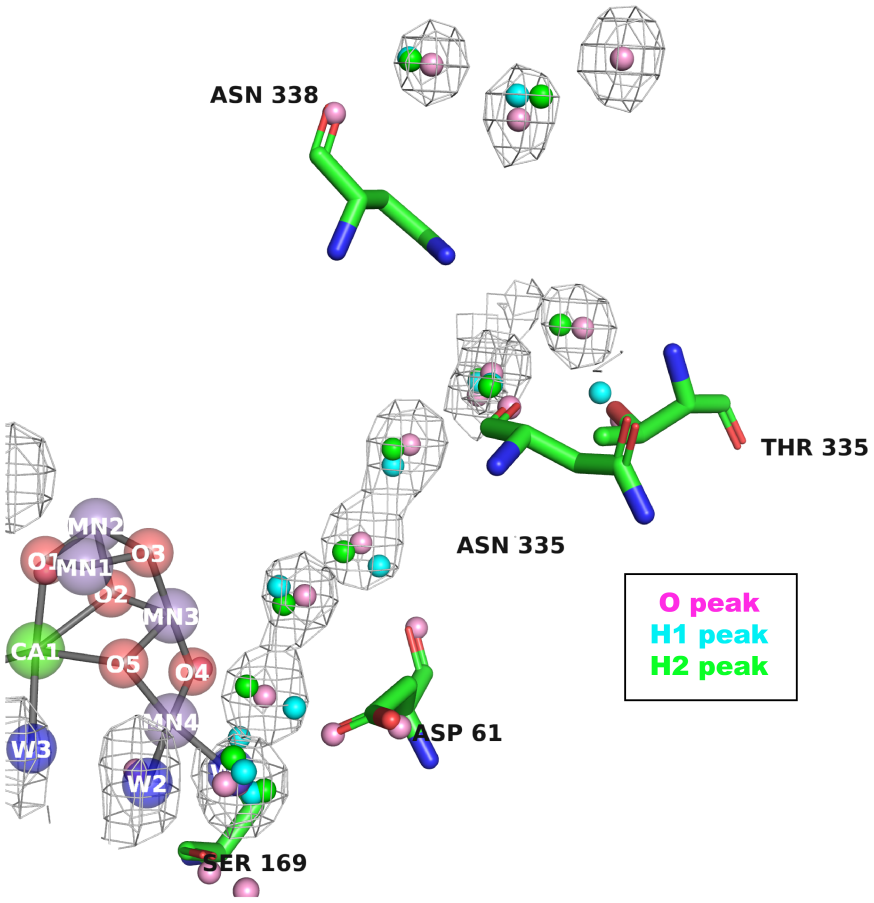


Figure 6.16: Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the O4 channel.

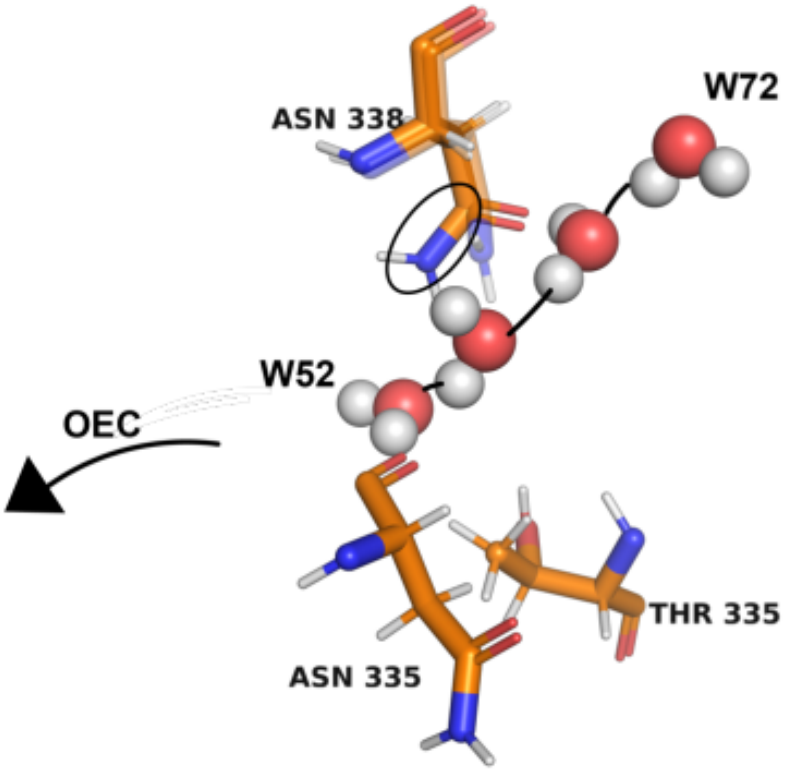


Figure 6.17: Snapshot at 45ns (third dimer, monomer A, for details see Methods), at which point an H bond chain transiently bridges the gap between the W50−W51−W52−W53 region and the W71−W72−W73 triangle. The ASN sidechain appears also to bend into a conformation which maximizes space for the transiting water. The resting state structure for ASN 338 is overlaid at a higher transparency to demonstrate ASN movement.

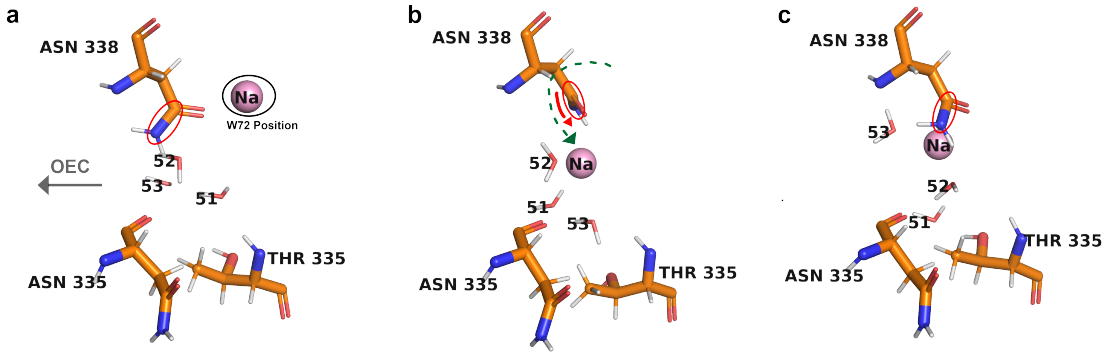


Figure 6.18: Schematic of the ASN drawbridge mechanism observed at 0 ns (a), 0.28 ns (b) and 2.74 ns (c) during the equilibration period (second dimer, monomer A, for details see Methods). In this example, the twisting of the ASN sidechain makes room for a Na+ ion hopping from the W72 position into a region near W51−W52−W53.

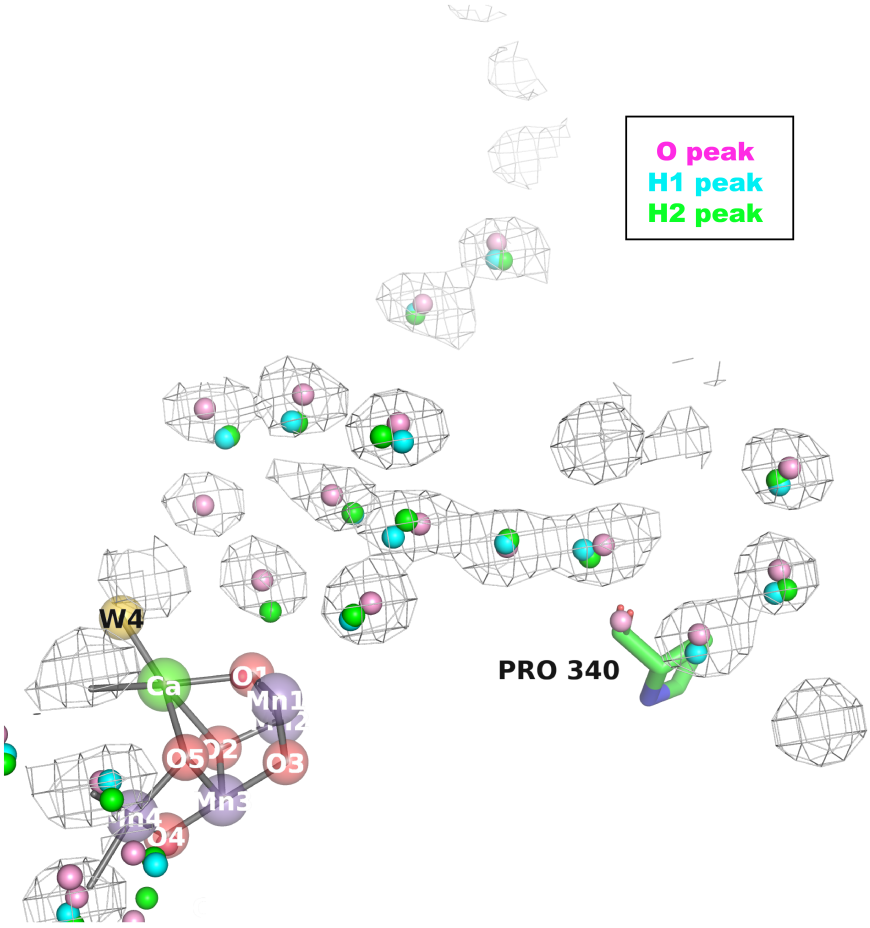


Figure 6.19: Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the O1 channel.

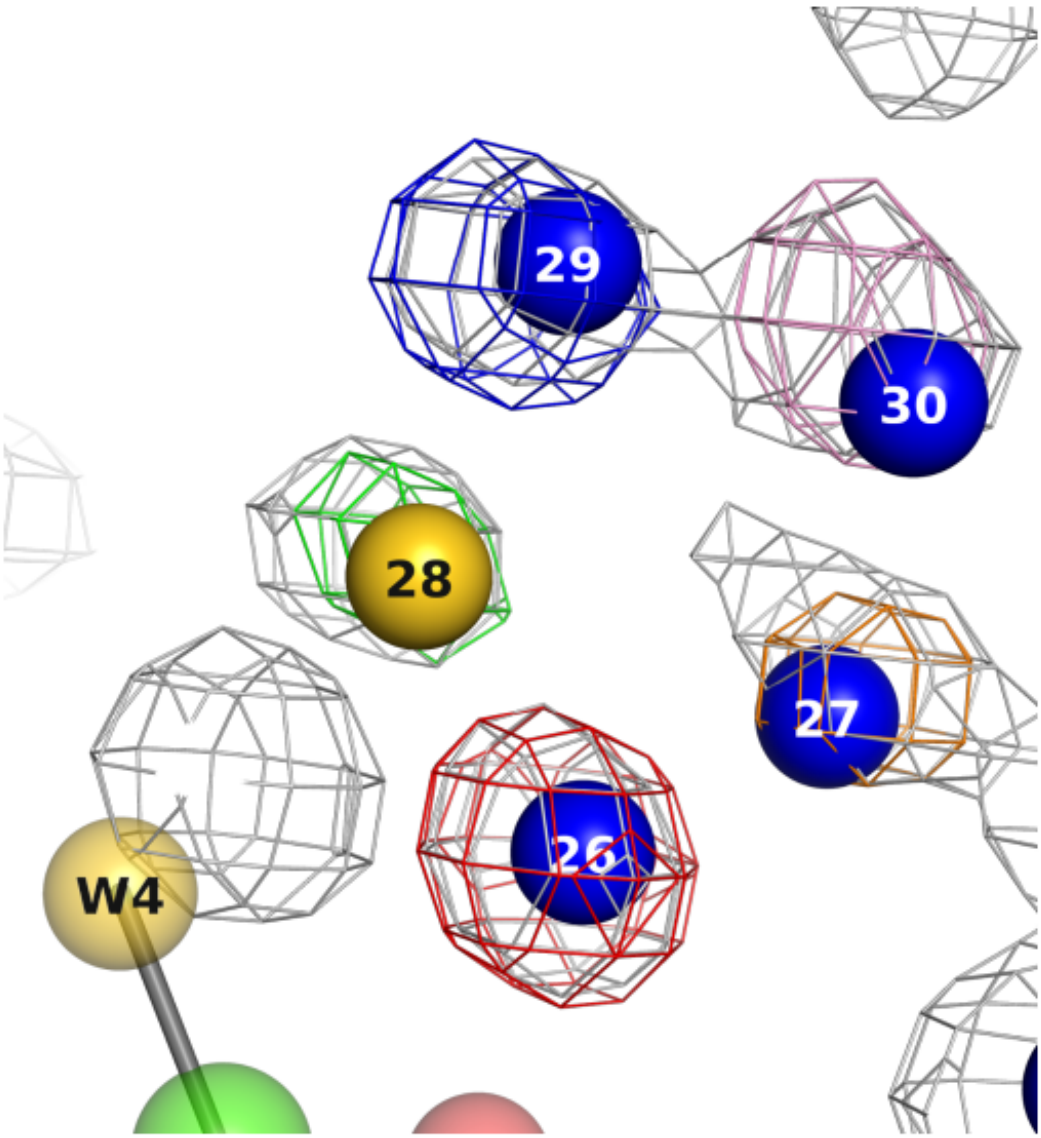


Figure 6.20: Calculated water maps for individual waters (colored) in the pentacluster (or water wheel), overlaid with the total water map (grey). Maps contoured at 2.08 e/Å^3 .Note that all individual water maps are visible at this level. The grey density next to W4 corresponds to an offset W4 position during the simulation.

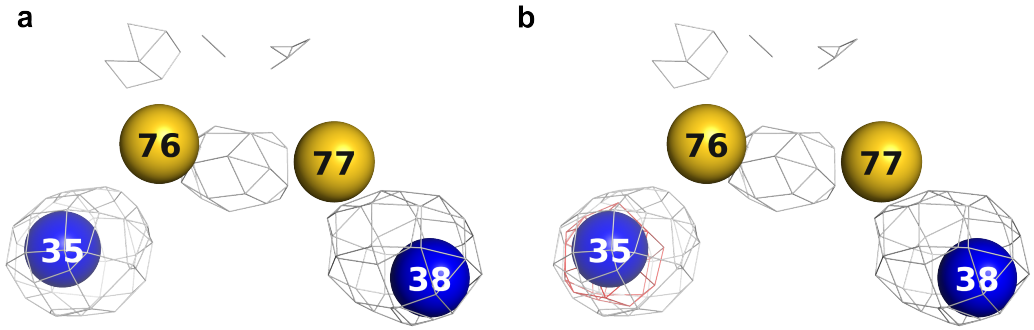


Figure 6.21: Calculated individual water maps for W35−W76−W77 overlaid with the total water map (grey). (a) All maps contoured at 2.08 e/Å^3 , yet only the total water map (grey) is visible at this level. (b) Individual water maps for W35 and W77 contoured at 1 e/Å^3 for better visualization, yet only the map for W35 (red) is visible. Note that no individual map was calculated for W76 as there was no equivalent water near its position after the 45 ns equilibration (see Methods). The density for W77 only becomes visible at levels < 0.55 e/Å^3.

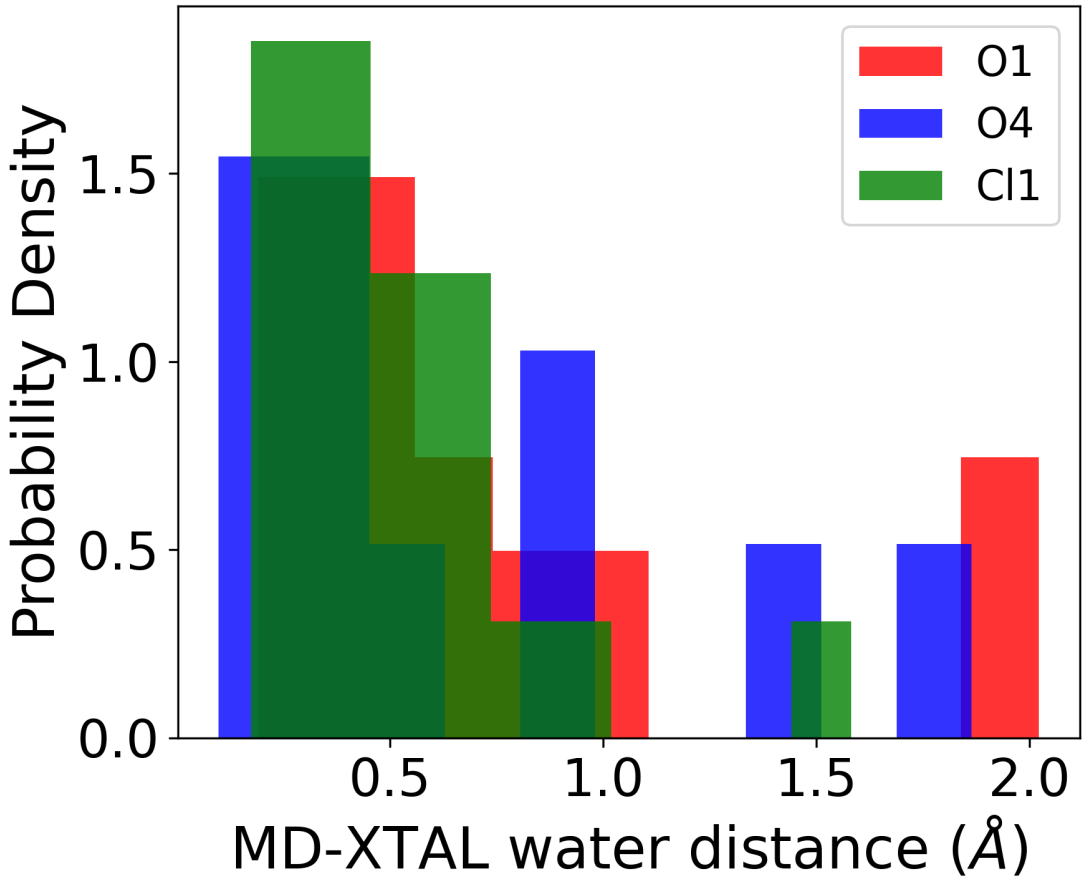


Figure 6.22: Distribution of the distances between crystallographic water positions and the nearest strong MD peaks (>2.08 e/Å^3) within three channels of interest. Each bin corresponds to raw count divided by the total number of counts in each channel as well as the bin width. As a result, the area under each histogram is normalized to one thus providing a probability density.

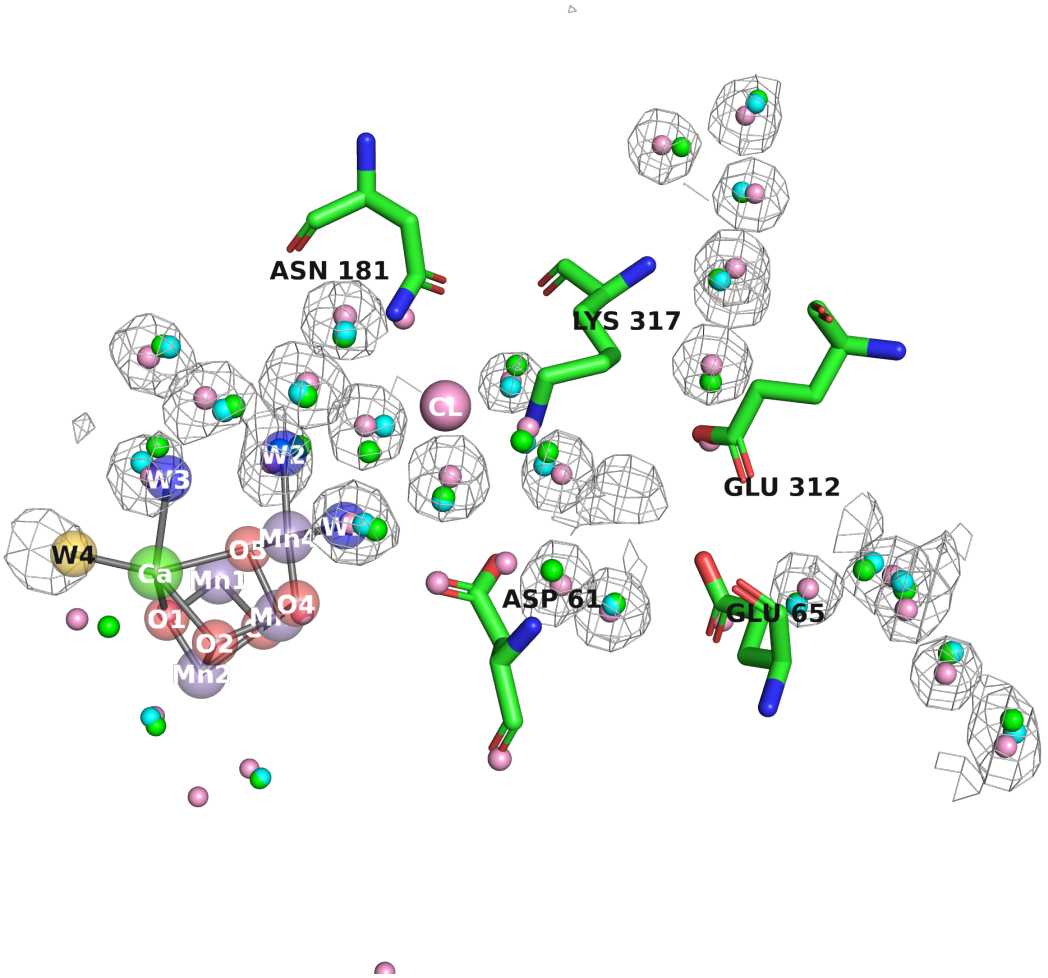


Figure 6.23: Oxygen and hydrogen peaks from the MADI analysis demonstrating the average orientation of waters which are hypothesized to participate in bonding in the Cl1 channel.

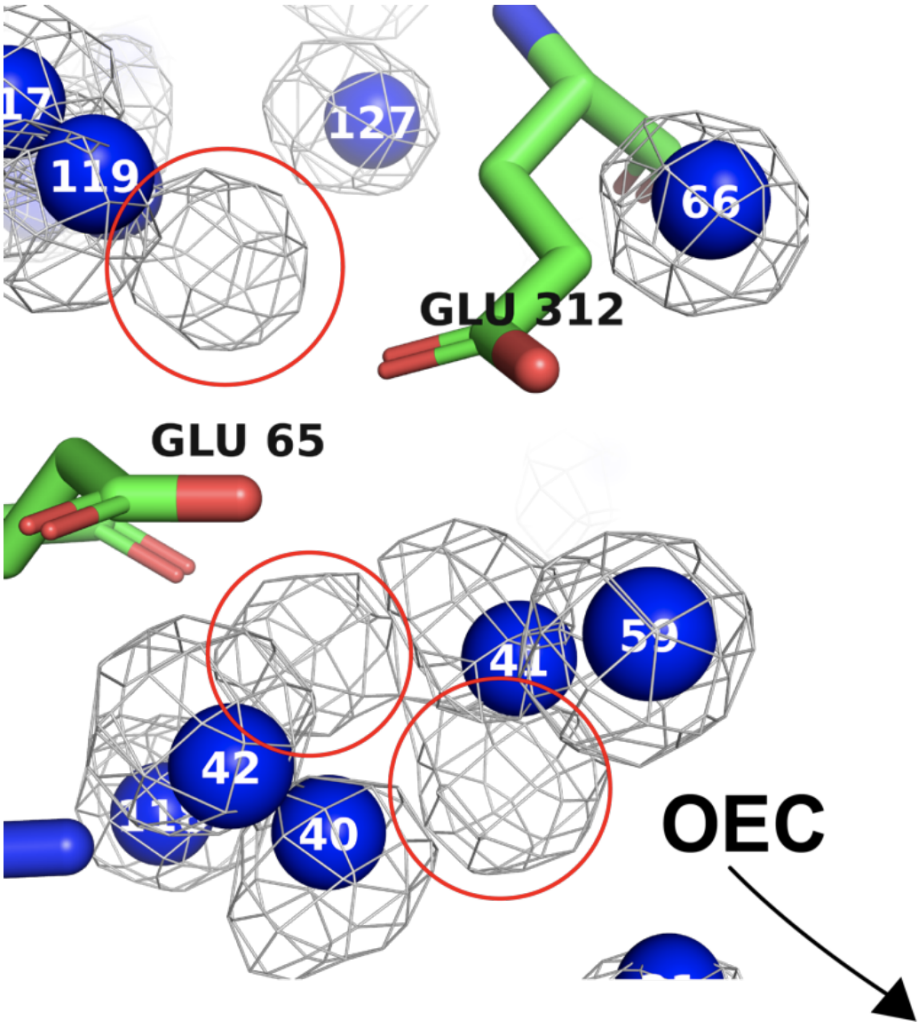


Figure 6.24: Three extra densities observed on both sides of the GLU 65 −GLU 312 'proton gate' in the Cl1 channel. Solvent map contoured at 2.08 e/Å^3 and overlaid with the crystal structure. In the crystal structure, the GLU 65 and GLU 312 distance is 2.7 Å. In the final snapshot from the MD trajectory, the GLU 65 and GLU 312 distance is at least 3 Å across all four copies. We hypothesize that these additional water positions can only be occupied when the bottleneck is wider than in the XRD structure.

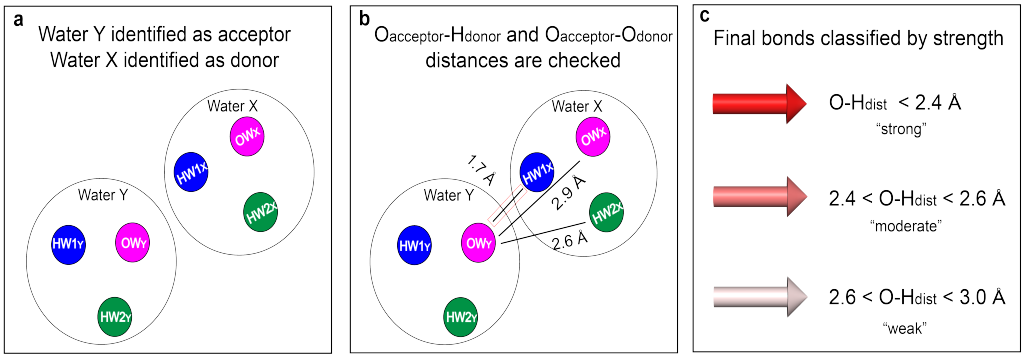


Figure 6.25: Determining optimal O−H bonds. (A) After oxygen and hydrogen peaks are grouped, forming an average water orientation from the simulation, all combinations of O−H peak distances between nearest neighbors were checked in order to converge on the optimal donor acceptor direction for a given pair. From panel A, water Y is clearly in a more optimal configuration for accepting a proton from water X than the other way around. (B) Once acceptor and donor roles were established, a few strict thresholds were set to start filtering the results. OH distances had to be below 3.0 Å, and $O_{donor}-H_{acceptor}$ distance $<$ $O_{donor}-O_{acceptor}$ distance. The latter filter was put in place to ensure that the orientation between donor acceptor pairs was optimized for hydrogen bonding. While both OW_y-HW1_x and OW_y- $HW2_x$ from panel B are feasible bonds, the distance of the former is smaller, so it is selected as the optimal O H bond distance in the pairing. After all optimal bonds are identified between pairs,if a water is registered as participating as a "donor" in three surviving bonds, the weakest of the three is knocked out. (C) The strength of the final bonds were classified based on the distribution of O−H distances retrieved from the MADI analysis (see SI 18). The results are represented by arrows following a color gradient, pointing from donor to acceptor (see Results).

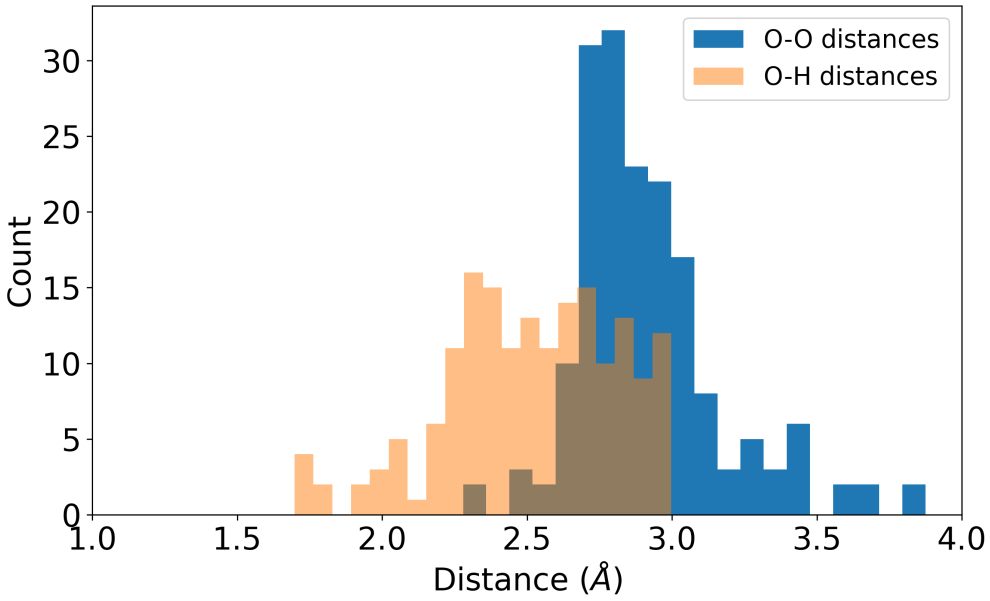


Figure 6.26: Distribution of O−O and O−H distances retrieved from the MADI analysis. For the O−H distribution the 33rd, 66th and 99th percentile are at 2.4 Å, 2.6 Å, 3.0 Å, respectively.

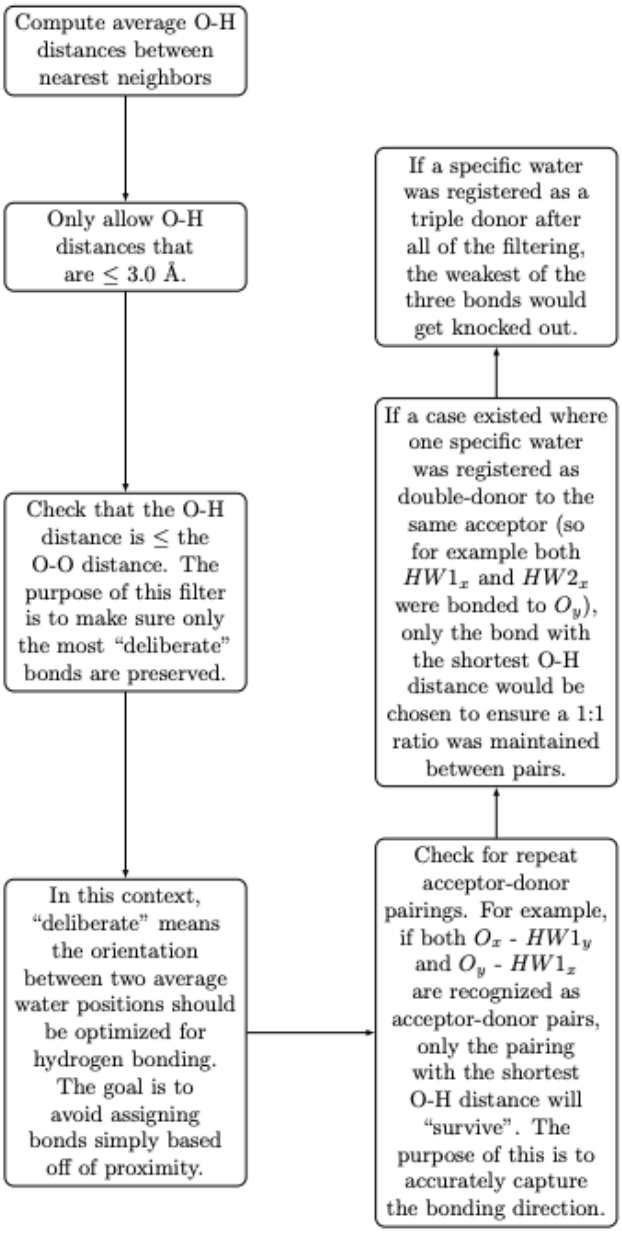


Figure 6.27: Flowchart detailing the filtering logic employed for the hydrogen bonding analysis.

Lipid/Cofactor/Ion	Copies in Simulation
OXYGEN EVOLVING COMPLEX	8
PHEOPHYTIN A	16
1,2 DI O ACYL 3 O [6 DEOXY 6 SULFO ALPHA D GLUCOPYRANOSYL] SN GLYCEROL(SQDG)	32
PROTOPORPHYRIN IX CONTAINING FE(Heme)	16
1,2 DIPALMITOYL PHOSPHATIDYL GLYCEROLE(PG)	40
1,2 DISTEAROYL MONOGALACTOSYL DIGLYCERIDE(MGDG)	52
2,3 DIMETHYL 5 (3,7,11,15,19,23,27,31,35 NONAMETHYL 2,6,10,14,18,22,26,30,34 HEXATRIACONTANONAENYL 2,5 CYCLOHEXADIENE 1,4 DIONE 2,3 DIMETHYL 5 SOLANESYL 1,4 BENZOQUINONE(PQ)	16
BETA CAROTENE(Car)	86
CHLOROPHYLL A	280
DIGALACTOSYL DIACYL GLYCEROL (DGDG)	40
STEARIC ACID	124
FE (II) ION	8
CHLORIDE ION	16
W1, W2, W3, W4	8

Table 6.1: Breakdown of lipids/cofactors/ions used in simulation

Element	# Atoms
Protein	330432
Crystal Waters	25260
Solvent molecules added by GROMACS	424731
Ligands/lipids/cofactors/ions	81471

Table 6.2: Breakdown of atoms in simulation

Ligand	Mol2 Source	Frcmod Source	Hydrogen Source
PHO	Sakashita	Sakashita	Tleap
BCR	Sakashita	Sakashita	Tleap
PL9	Sakashita	Sakashita	Tleap
HEM	Sakashita	Sakashita	Tleap
CLA	Sakashita	Sakashita	Tleap
SQD	This study	This study	Tleap
DGD	Sakashita	This study	Tleap
STE	This study	This study	Tleap
LMG	Sakashita	This study	Tleap
LHG	Sakashita	Sakashita	Tleap
FE2	This study	This study	N/A
CL	This study	This study	N/A

Table 6.3: Documentation of how various lipids/cofactors/ions were prepared for this study.

Residue	Net Charge
PHO	0
BCR	0
PL9	0
HEM	2
CLA	0
SQD	0
DGD	0
STE	1
LMG	0
LHG	1
FE2	+2
CL	1

Table 6.4: Net charges for the lipids/cofactors/ions used in this simulation.

Channel	Constituents	Orientation
O4	W49−W48−W20−O4	
	W49−W48−W20−W19−O4	
	W49−W48−W20−W19−ASP 61	Towards active site
O1	W29−W30−W27−W32−W31	
	W29−W30−W27−W32−W39	Away from active site
O1	W34−W33−W32−W31	
	W34−W33−W32−W39	Towards active site
O1	W38−W37−W36	Towards active site
Cl1	W3−W23−W24−W2−W22−ASN 181	
	W3−W23−W24−W2−W22−W21−ASP 61	
	W3−W23−W24−W2−W22−W21−W1−ASP 61	Away from active site
Cl1	W121−W117−W119	Towards active site
Cl1	W126−W122−W117−W119	Towards active site

Table 6.5: All significant water wires (must contain ≥ 3 waters with a consistent acceptor donor orientation) identified by the MADI analysis

Water	**Amino acid atom name #1**	**Amino acid resn #1**	**Amino acid chain #1**	**Amino acid resid #1**	**Amino acid atom name #2**	**Amino acid resn #2**	**Amino acid chain #2**	**Amino acid resid #2**
103	O	ALA	C	399	N	GLY	C	409
104	O	THR	C	397	N	GLY	C	409
106	CG	GLU	C	83	N	ASN	C	418
107	C	GLU	C	83	O	ASN	C	418
109	O	GLN	V	34	N	ASN	C	418
26	CD	GLU	A	189	C	ALA	A	344
27	O	ASP	A	342	CA	GLU	A	189
29	O	GLU	A	189	OD1	ASN	A	296
31	O	ALA	A	188	OE1	GLU	A	329
32	OE1	GLU	A	329	O	ASP	A	342
33	N	ASP	A	342	OE2	GLU	A	329
34	O	PRO	A	340	NZ	LYS	V	134
36	O	PRO	A	340	NZ	LYS	V	134
37	N	LYS	U	104	NZ	LYS	V	134
38	CB	LYS	U	104	CB	LYS	V	134
101	NH1	ARG	B	384	N	TYR	U	103
102	NH1	ARG	B	384	OE2	GLU	B	387

Table 6.6: Defining O1 channel by amino acids. ROI is 3.5 Å for all waters.

Water	**Amino acid atom name #1**	**Amino acid resn #1**	**Amino acid chain #1**	**Amino acid resid #1**	**Amino acid atom name #2**	**Amino acid resn #2**	**Amino acid chain #2**	**Amino acid resid #2**
21	OD1	ASP	A	61	NZ	LYS	D	317
40	OE2	GLU	D	312	OD2	ASP	A	61
42	OE1	GLU	A	65	N	ARG	A	334
119	CD	GLU	A	65	C	ASP	O	158
117	O	PRO	D	309	CG	GLU	A	65
121	NH2	ARG	O	152	O	GLU	D	310
125	O	PRO	D	309	OD1	ASP	O	224
126	OE1	GLU	D	310	NH2	ARG	O	152
140	O	ARG	A	64	CG	PRO	O	117
143	CG	PRO	O	117	OH	TYR	A	107
25	OH	TYR	A	161	CD	GLU	A	189
24	CA	ASP	A	170	N	PHE	A	186
61	CA	PHE	A	182	O	SER	A	169
22	ND2	ASN	A	181	OE2	GLU	A	333
60	O	MET	A	331	O	LYS	D	317
66	CD	GLU	D	312	O	PRO	O	159
67	CB	LEU	D	319	N	LYS	O	160
68	OE1	GLU	D	323	CA	LYS	O	160
69	OE1	GLU	D	323	NZ	LYS	O	160
130	OH	TYR	D	296	NH2	ARG	D	326
132	O	GLU	B	364	O	ILE	O	172
133	OD2	ASP	D	297	O	ALA	O	173
135	NH1	ARG	B	326	CG2	ILE	B	442
136	NZ	LYS	B	321	O	GLU	B	364
137	O	LYS	B	321	NH1	ARG	B	326

Table 6.7: Defining Cl1 channel by amino acids. ROI is 3.5 Å for all waters except for W40, W121, W24 and W61, which have a radius of 5 Å.

Water	Amino acid atom name #1	Amino acid resn #1	Amino acid chain #1	Amino acid resid #1	Amino acid atom name #2	Amino acid resn #2	Amino acid chain #2	Amino acid resid #2
19	OD1	ASP	A	61	NH2	ARG	C	357
20	OE2	GLU	C	354	CB	ASP	A	61
48	O	ASP	A	61	ND2	ASN	A	87
49	CG2	ILE	A	63	C	GLY	C	353
50	OG1	THR	C	335	CA	GLY	C	353
51	OG1	THR	C	335	O	LEU	C	337
52	O	ASN	A	335	ND2	ASN	A	338
53	O	GLY	C	338	SD	MET	C	342
71	O	ASN	A	338	OD1	ASN	U	99
72	CD	PRO	C	334	ND2	ASN	D	350
73	ND2	ASN	O	155	OD1	ASP	U	96
74	O	GLN	C	332	OG1	THR	O	153
75	N	ASN	O	155	OD2	ASP	U	96
113	O	ALA	U	90	OD2	ASP	U	100
114	OD1	ASN	O	155	O	ASN	U	100
112	CB	ASN	U	31	ND2	ASN	U	100
115	OH	TYR	U	21	N	ASN	U	31

Table 6.8: Defining O4 channel by amino acids. ROI is 3.5 Å for all waters.

Channel	Number of MD peaks	Number of Crystallographic Waters
O1	30	22
O4	20	19
Cl1	46	35

Table 6.9: Number of MD peaks and crystallographic waters found in each of the three channels, using the amino acid boundaries tabulated in tables 6.6 - 6.8.

6.4 Outlook

There are a variety of future directions one can pursue based on this work. The most logical option involves extending this protocol, which was developed for simulating and analyzing the resting state (S_1), to the other metastable intermediates (S_2, S_3, S_0) as well as the timepoints where we have structures. This will require reparameterization of the manganese cluster.

Through these simulations, we can better investigate the physical mechanism of substrate water entry, track down the cause of disappearance of certain water molecules, study changes in the hydrogen bond network across various S states (through the MADI analysis which infers likely hydrogen bonding connectivity and possible proton transfer pathways), as well as investigate changes in the protein in response to the water network changes.

Using the current S_1 state simulation, one can try a similar restraint titration as has been done already by Wych et al [223]. This type of investigation would be of special interest

if trying to strike the adequate balance between reproducing experimental B factors and recovering crystallographic water positions. To meet the former goal, lower restraints are preferred. To meet the latter goal, higher restraints are preferred.

One can also try comparing/contrasting results from this chapter to those generated when carrying out the simulation in a membrane environment. This would require first artificially constructing a membrane using computational tools, since the PS II protein is separated from the membrane during purification, therefore it is not included in the crystal structure. This sort of investigation has the potential to shed light on the effects of crystal packing on interior waters.

Bibliography

[1] David Attwood and Anne Sakdinawat. *X-Rays and Extreme Ultraviolet Radiation: Principles and Applications*. 2nd ed. Cambridge University Press, 2017.

[2] Philip Willmott. *An introduction to synchrotron radiation: techniques and applications*. John Wiley & Sons, 2019.

[3] Daan Brinks et al. "Ultrafast dynamics of single molecules". *Chemical Society Reviews* 43.8 (2014), pp. 2476–2491.

[4] Yann Fichou. "Hydration water dynamics of the tau protein in its native and amyloid states". PhD thesis. Université Grenoble Alpes, 2015.

[5] Allen M Orville. "Recent results in time resolved serial femtosecond crystallography at XFELs". *Current Opinion in Structural Biology* 65 (2020), pp. 193–208.

[6] Nicoletta Liguori et al. "Molecular dynamics simulations in photosynthesis". *Photosynthesis research* 144.2 (2020), pp. 273–295.

[7] Omar Desouky, Nan Ding, and Guangming Zhou. "Targeted and non-targeted effects of ionizing radiation". *Journal of Radiation Research and Applied Sciences* 8.2 (2015), pp. 247–254.

[8] Karol Nass. "Radiation damage in protein crystallography at X-ray free-electron lasers". *Acta Crystallographica Section D: Structural Biology* 75.2 (2019), pp. 211–218.

[9] Robin L Owen et al. "Outrunning free radicals in room-temperature macromolecular crystallography". *Acta Crystallographica Section D: Biological Crystallography* 68.7 (2012), pp. 810–818.

[10] Vladimir Isachenko et al. "High cryo-resistance of SARS-CoV-2 virus: Increased risk of re-contamination at transplantation of cryopreserved ovarian tissue after COVID-19 pandemic". *Cryobiology* 103 (2021), pp. 1–6.

[11] H Hartmann et al. "Conformational substates in a protein: structure and dynamics of metmyoglobin at 80 K." *Proceedings of the National Academy of Sciences* 79.16 (1982), pp. 4967–4971.

[12] Bertil Halle. "Structural changes during flash cooling". *Proceedings of the National Academy of Sciences* 101.14 (2004), pp. 4793–4798.

[13] Robert F Tilton Jr, John C Dewan, and Gregory A Petsko. "Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K". *Biochemistry* 31.9 (1992), pp. 2469–2481.

[14] JCH Spence. "XFELs for structure and dynamics in biology". *IUCrJ* 4.4 (2017), pp. 322–339.

[15] Geoffrey M Cooper and Kenneth Adams. *The cell: a molecular approach*. Oxford University Press, 2022.

[16] KM Sendra et al. "An ancient metalloenzyme evolves through metal preference modulation". *Nature Ecology & Evolution* 7.5 (2023), pp. 732–744.

[17] Sidsel Birkelund Schmidt and Søren Husted. "The biochemical properties of manganese in plants". *Plants* 8.10 (2019), p. 381.

[18] Nicolas Théodore de Saussure. *Recherches chimiques sur la vegetation*. Nyon, 1804.

[19] JS McHargue. "The Effect of Manganese on the Growth of Wheat: A Source of Manganese for Agricultural Purposes." *Industrial & Engineering Chemistry* 11.4 (1919), pp. 332–335.

[20] Mark Spector and G Douglas Winget. "Purification of a manganese-containing protein involved in photosynthetic oxygen evolution and its use in reconstituting an active membrane". *Proceedings of the National Academy of Sciences* 77.2 (1980), pp. 957–959.

[21] Jean-David Rochaix. "Regulation of photosynthetic electron transport". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1807.3 (2011), pp. 375–383.

[22] Hiroshi Ishikita and Ernst-Walter Knapp. "Function of redox-active tyrosine in photosystem II". *Biophysical journal* 90.11 (2006), pp. 3886–3896.

[23] Floris J Van Eerden et al. "Exchange pathways of plastoquinone and plastoquinol in the photosystem II complex". *Nature communications* 8.1 (2017), p. 15214.

[24] William J Lennarz and M Daniel Lane. *Encyclopedia of biological chemistry*. Academic Press, 2013.

[25] Joachim Weber. "ATP synthase–the structure of the stator stalk". *Trends in biochemical sciences* 32.2 (2007), pp. 53–56.

[26] Jan F Kern, Johannes Messinger, John Whitmarsh, et al. "Photosystem II". *eLS* (2010).

[27] Baker A Martin and Paul D Frymier. "A review of hydrogen production by photosynthetic organisms using whole-cell and cell-free systems". *Applied biochemistry and biotechnology* 183 (2017), pp. 503–519.

[28] Markus Grabolle and Holger Dau. "Energetics of primary and secondary electron transfer in photosystem II membrane particles of spinach revisited on basis of recombination fluorescence measurements". *Biochimica et Biophysica Acta (BBA) Bioenergetics* 1708.2 (2005), pp. 209–218.

[29] RJ Debus et al. "Directed mutagenesis indicates that the donor to P 680+ in photosystem II is tyrosine-161 of the D1 polypeptide". *Biochemistry* 27.26 (1988), pp. 9071–9074.

[30] Johannes Messinger et al. "The S0 state of the oxygen-evolving complex in photosystem II is paramagnetic: detection of an EPR multiline signal". *Journal of the American Chemical Society* 119.46 (1997), pp. 11349–11350.

[31] Anna J Warren, Danny Axford, and Robin L Owen. "Direct measurement of X-ray-induced heating of microcrystals". *Journal of Synchrotron Radiation* 26.4 (2019), pp. 991–997.

[32] Jens Als-Nielsen and Des McMorrow. *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[33] Richard Neutze et al. "Potential for biomolecular imaging with femtosecond X-ray pulses". *Nature* 406.6797 (2000), pp. 752–757.

[34] Henry N. Chapman et al. "Femtosecond X-ray protein nanocrystallography". *Nature* 470 (2011), pp. 73–77. DOI: 10.1038/nature09750.

[35] Sébastien Boutet et al. "High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography". *Science* 337 (2012), pp. 362–364. DOI: 10.1126/science.1217737.

[36] Jan Kern et al. "Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature". *Science* 340.6131 (2013), pp. 491–495.

[37] Roberto Alonso-Mori et al. "Photon-in photon-out hard X-ray spectroscopy at the Linac Coherent Light Source". *Journal of synchrotron radiation* 22.3 (2015), pp. 612–620.

[38] Roberto Alonso-Mori et al. "Energy-dispersive X-ray emission spectroscopy using an X-ray free-electron laser in a shot-by-shot mode". *Proceedings of the National Academy of Sciences* 109.47 (2012), pp. 19103–19107.

[39] James R. Connolly. *The Interaction of X-rays with Matter and Radiation Safety*. Prepared for EPS400-002, Introduction to X-Ray Powder Diffraction, Spring 2007. Material in this document is borrowed from many sources; all original material is ©2007 by James R. Connolly. Revision date: 23-Jan-07 Page 1 of 14. Introduction. 2007.

[40] Uwe Bergmann and Pieter Glatzel. "X-ray emission spectroscopy". *Photosynthesis research* 102 (2009), pp. 255–266.

[41] Rebeca G Castillo et al. "Kβ X-Ray Emission Spectroscopic Study of a Second-Row Transition Metal (Mo) and Its Application to Nitrogenase-Related Model Complexes". *Angewandte Chemie International Edition* 59.31 (2020), pp. 12965–12975.

[42] Uwe Bergmann. "Stimulated X-ray emission spectroscopy". *Photosynthesis Research* (2024), pp. 1–14.

[43] Martin Karplus and J Andrew McCammon. "Molecular dynamics simulations of biomolecules". *Nature structural biology* 9.9 (2002), pp. 646–652.

[44] Miguel A González. "Force fields and molecular dynamics simulations". *École thématique de la Société Française de la Neutronique* 12 (2011), pp. 169–200.

[45] Wojciech Kopec and Vytautas Gapsys. "Periodic boundaries in Molecular Dynamics simulations: why do we need salt?" *bioRxiv* (2022), pp. 2022–10.

[46] Srinivas Ramachandran et al. "Automated minimization of steric clashes in protein structures". *Proteins: Structure, Function, and Bioinformatics* 79.1 (2011), pp. 261–270.

[47] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[48] F. R. Elder et al. "Radiation from Electrons in a Synchrotron". *Phys. Rev.* 71 (11 June 1947), pp. 829–830. DOI: 10.1103/PhysRev.71.829.5. URL: https://link.aps.org/doi/10.1103/PhysRev.71.829.5.

[49] Giorgio Margaritondo. "Who were the founders of synchrotron radiation? Historical facts and misconceptions". *Journal of Vacuum Science & Technology A* 40.3 (2022).

[50] Tetsuya Ishikawa. "Accelerator-based X-ray sources: synchrotron radiation, X-ray free electron lasers and beyond". *Philosophical Transactions of the Royal Society A* 377.2147 (2019), p. 20180231.

[51] DJ Holder, P Quinn, and N Wyles. "The SRS at Daresbury Laboratory: a Eulogy to the World's First Dedicated High-energy Synchrotron Radiation Source". *Cockcroft Preprints* (2008), pp. 2133–2135.

[52] C Pellegrini. *Proceedings of the Workshop on 4th Generation Light Sources*. 1992.

[53] John MJ Madey. "Stimulated emission of bremsstrahlung in a periodic magnetic field". *Journal of Applied Physics* 42.5 (1971), pp. 1906–1913.

[54] C Pellegrini, A Marinelli, and S Reiche. "The physics of x-ray free-electron lasers". *Reviews of Modern Physics* 88.1 (2016), p. 015006.

[55] Sabine Botha and Petra Fromme. "Review of serial femtosecond crystallography including the COVID-19 pandemic impact and future outlook". *Structure* (2023).

[56] Ryan N Coffee et al. "Development of ultrafast capabilities for X-ray free-electron lasers at the linac coherent light source". *Philosophical Transactions of the Royal Society A* 377.2145 (2019), p. 20180386.

[57] Karen Appel et al. "The High Energy Density (HED) science instrument at European XFEL: a new platform to experimentally study planetary matter". *Joint AIRAPT-25th & EHPRG-53rd International Conference on High Pressure Science and Technology*. 2015.

[58] Vladimir Shiltsev. "Particle beams behind physics discoveries". *Physics Today* 73.4 (2020), pp. 32–39.

[59] Iulia Georgescu. "The first decade of XFELs". *Nature Reviews Physics* 2.7 (2020), pp. 345–345.

[60] Johannes Bahrdt. "Shaping photon beams with undulators and wigglers". *Synchrotron Light Sources and Free-Electron Lasers: Accelerator Physics, Instrumentation and Science Applications* (2020), pp. 851–933.

[61] Toru Hara et al. "Two-colour hard X-ray free-electron laser with wide tunability". *Nature communications* 4.1 (2013), p. 2919.

[62] Zhirong Huang and Kwang-Je Kim. "Review of x-ray free-electron laser theory". *Physical Review Special Topics—Accelerators and Beams* 10.3 (2007), p. 034801.

[63] Ichiro Inoue et al. "Generation of narrow-band X-ray free-electron laser via reflection self-seeding". *Nature photonics* 13.5 (2019), pp. 319–322.

[64] J Amann et al. "Demonstration of self-seeding in a hard-X-ray free-electron laser". *Nature photonics* 6.10 (2012), pp. 693–698.

[65] Yeukuang Hwu and Giorgio Margaritondo. "Synchrotron radiation and X-ray free-electron lasers (X-FELs) explained to all users, active and potential". *Journal of Synchrotron Radiation* 28.3 (2021), pp. 1014–1029.

[66] James S Fraser et al. "Accessing protein conformational ensembles using room temperature X-ray crystallography". *Proceedings of the National Academy of Sciences* 108.39 (2011), pp. 16247–16252.

[67] Haiguang Liu and Weontae Lee. *The XFEL protein crystallography: developments and perspectives*. 2019.

[68] M Marvin Seibert et al. "Single mimivirus particles intercepted and imaged with an X-ray laser". *Nature* 470.7332 (2011), pp. 78–81.

[69] Jan Kern et al. "Structures of the intermediates of Kok's photosynthetic water oxidation clock". *Nature* 563.7731 (2018), pp. 421–425.

[70] Wenkai Zhang et al. "Tracking excited-state charge and spin dynamics in iron coordination complexes". *Nature* 509 (2014), pp. 345–348. DOI: 10.1038/nature13252.

[71] M Marvin Seibert et al. "Femtosecond diffractive imaging of biological cells". *Journal of Physics B: Atomic, Molecular and Optical Physics* 43.19 (2010), p. 194015.

[72] Christoph Bostedt et al. "Linac coherent light source: The first five years". *Reviews of Modern Physics* 88.1 (2016), p. 015007.

[73] Luca Giannessi. "Fast and furious X-ray free-electron lasers". *Nature Photonics* (2024).

[74] Andy Extance. "UK begins exploration of whether to build its own billion-pound-plus XFEL". *Chemistry World* (Mar. 2023).

[75] Jon Marangos. "UK XFEL Science Case" (2020).

[76] Sakura Pascarelli, Serguei Molodtsov, Thomas Tschentscher, et al. "The next decade of XFELs" ().

[77] TO Raubenheimer. "Technical challenges of the LCLS-II CW X-ray FEL". *Proc. 6 Int. Particle Accelerator Conf.* 2015.

[78] A. Halavanau et.al. "Population inversion X-ray laser oscillator". *Proceedings of the National Academy of Sciences* 117.27 (2020), pp. 15511–15516. ISSN: 0027-8424. DOI: 10.1073/pnas.2005360117. eprint: https://www.pnas.org/content/117/27/15511.full.pdf. URL: https://www.pnas.org/content/117/27/15511.

[79] Kwang-Je Kim, Yuri Shvyd'ko, and Sven Reiche. "A proposal for an x-ray free-electron laser oscillator with an energy-recovery linac". *Physical review letters* 100.24 (2008), p. 244802.

[80] P Rauer et al. "Integration of an xfelo at the european xfel facility". *energy [mJ]* 10.4 (2019), pp. 10–2.

[81] Sam M Vinko. "The X-ray science frontier is ultra-short and ultra-intense". *Nature Photonics* (2024), pp. 1–2.

[82] Henry N Chapman, Carl Caleman, and Nicusor Timneanu. "Diffraction before destruction". *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1647 (2014), p. 20130313.

[83] Vahid Rahmani et al. "Data reduction for X-ray serial crystallography using machine learning". *Journal of applied crystallography* 56.1 (2023).

[84] Gabriel Blaj et al. "X-ray detectors at the Linac Coherent Light Source". *Journal of synchrotron radiation* 22.3 (2015), pp. 577–583.

[85] Yang-Kyun Park and Gregory C Sharp. "Gain correction for an X-ray imaging system with a movable flat panel detector and intrinsic localization crosshair". *Technology in Cancer Research & Treatment* 15.2 (2016), pp. 387–395.

[86] Frédéric Livet. "Diffraction with a coherent X-ray beam: dynamics and imaging". *Acta Crystallographica Section A: Foundations of Crystallography* 63.2 (2007), pp. 87–107.

[87] Thomas Fransson et al. "X-ray emission spectroscopy as an in situ diagnostic tool for X-ray crystallography of metalloproteins using an X-ray free-electron laser". *Biochemistry* 57.31 (2018), pp. 4629–4637.

[88] Nicholas K Sauter. "XFEL diffraction: developing processing methods to optimize data quality". *Journal of Synchrotron Radiation* 22.2 (2015), pp. 239–248.

[89] Auralee Edelen et al. "Opportunities in machine learning for particle accelerators". *arXiv preprint arXiv:1811.03172* (2018).

[90] Yue Sun and Sandor Brockhauser. "Machine learning applied for spectra classification in X-ray free electorn laser sciences". *Data Science Journal* 21 (2022), pp. 15–15.

[91] Mike Folk et al. "An overview of the HDF5 technology suite and its applications". *Proceedings of the EDBT/ICDT 2011 workshop on array databases*. 2011, pp. 36–47.

[92] Sotiris Kotsiantis, Dimitris Kanellopoulos, Panayiotis Pintelas, et al. "Handling imbalanced datasets: A review". *GESTS international transactions on computer science and engineering* 30.1 (2006), pp. 25–36.

[93] Rikiya Yamashita et al. "Convolutional neural networks: an overview and application in radiology". *Insights into imaging* 9 (2018), pp. 611–629.

[94] Lei Zhao and Zhonglin Zhang. "A improved pooling method for convolutional neural networks". *Scientific Reports* 14.1 (2024), p. 1589.

[95] Chat Room. "Confusion matrix". *Mach. Learn* 6 (2019), p. 27.

[96] Italo Maia. *Building Web Applications with Flask*. Packt Publishing Ltd, 2015.

[97] John N Galayda. *The LCLS-II: A high power upgrade to the LCLS*. Tech. rep. SLAC National Accelerator Lab., Menlo Park, CA (United States), 2018.

[98] Winfried Decking et al. "A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator". *Nature photonics* 14.6 (2020), pp. 391–397.

[99] Jana Thayer et al. "Massive Scale Data Analytics at LCLS-II". *EPJ Web of Conferences*. Vol. 295. EDP Sciences. 2024, p. 13002.

[100] Vahid Rahmani et al. "Robust image descriptor for machine learning based data reduction in serial crystallography". *Applied Crystallography* 57.2 (2024).

[101] G. Aad et al. "Search for New Phenomena in Two-Body Invariant Mass Distributions Using Unsupervised Machine Learning for Anomaly Detection at $\sqrt{s} = 13$ TeV with the ATLAS Detector". *Phys. Rev. Lett.* 132 (8 Feb. 2024), p. 081801.

[102] Uwe Bergmann et al. "Using X-ray free-electron lasers for spectroscopy of molecular catalysts and metalloenzymes". *Nature Reviews Physics* 3.4 (2021), pp. 264–282.

[103] Osvald Lundquist. "Über die kβ-linien in den röntgenemissionsspektra der elemente phosphor und kalium". *Zeitschrift für Physik* 33 (1925), pp. 901–915.

[104] Armin Meisel, Gunter Leonhardt, and Rüdiger Szargan. *X-ray spectra and chemical binding*. Vol. 37. Springer, 1989.

[105] Kenjiro Tsutsumi. "The X-ray Non-diagram Lines K β'of Some Compounds of the Iron Group". *Journal of the Physical Society of Japan* 14.12 (1959), pp. 1696–1706.

[106] Uwe Bergmann et al. "Characterization of the Mn oxidation states in photosystem II by Kβ X-ray fluorescence spectroscopy". *The Journal of Physical Chemistry B* 102.42 (1998), pp. 8350–8352.

[107] L. Young et al. "Femtosecond electronic response of atoms to ultra-intense X-rays". *Nature* 466 (2010), pp. 56–61. DOI: 10.1038/nature09177.

[108] Jose M Martin-Garcia et al. "Serial femtosecond crystallography: A revolution in structural biology". *Archives of biochemistry and biophysics* 602 (2016), pp. 32–47.

[109] Hiroki Makita et al. "Combining on-line spectroscopy with synchrotron and X-ray free electron laser crystallography". *Current opinion in structural biology* 80 (2023), p. 102604.

[110] Franklin D Fuller et al. "Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers". *Nature methods* 14.4 (2017), pp. 443–449.

[111] Jaehyun Park and Ki Hyun Nam. "Sample delivery systems for serial femtosecond crystallography at the PAL-XFEL". *Photonics*. Vol. 10. 5. MDPI. 2023, p. 557.

[112] Ruchira Chatterjee et al. "XANES and EXAFS of dilute solutions of transition metals at XFELs". *Journal of synchrotron radiation* 26.5 (2019), pp. 1716–1724.

[113] Philipp S Simon et al. "Capturing the sequence of events during the water oxidation reaction in photosynthesis using XFELs". *FEBS letters* 597.1 (2023), pp. 30–37.

[114] Roberto Alonso-Mori et al. "A multi-crystal wavelength dispersive x-ray spectrometer". *Review of Scientific Instruments* 83.7 (2012).

[115] Patrick Rabe et al. "X-ray free-electron laser studies reveal correlated motion during isopenicillin N synthase catalysis". *Science Advances* 7.34 (2021), eabh0250.

[116] Vivek Srinivas et al. "High-resolution XFEL structure of the soluble methane monooxygenase hydroxylase complex with its regulatory component at ambient temperature in two oxidation states". *Journal of the American Chemical Society* 142.33 (2020), pp. 14249–14266.

[117] Christopher J Ohmer et al. "XFEL serial crystallography reveals the room temperature structure of methyl-coenzyme M reductase". *Journal of inorganic biochemistry* 230 (2022), p. 111768.

[118] Mohamed Ibrahim et al. "Untangling the sequence of events during the S2→S3 transition in photosystem II and implications for the water oxidation mechanism". *Proceedings of the National Academy of Sciences* 117.23 (2020), pp. 12624–12635.

[119] Sara Lafuerza et al. "Chemical sensitivity of Kβ and Kα X-ray emission from a systematic investigation of iron compounds". *Inorganic Chemistry* 59.17 (2020), pp. 12518–12535.

[120] Dmitry Shevela et al. "Solar energy conversion by photosystem II: principles and structures". *Photosynthesis research* 156.3 (2023), pp. 279–307.

[121] Nathan Nelson and Charles F Yocum. "Structure and function of photosystems I and II". *Annu. Rev. Plant Biol.* 57 (2006), pp. 521–565.

[122] Junko Yano and Vittal Yachandra. "Mn4Ca cluster in photosynthesis: where and how water is oxidized to dioxygen". *Chemical reviews* 114.8 (2014), pp. 4175–4205.

[123] Dmitry Shevela et al. "Photosystem II". *eLS* 2.7 (2021), pp. 1–16.

[124] Bessel Kok, Bliss Forbush, and Marion McGloin. "Cooperation of charges in photosynthetic O2 evolution–I. A linear four step mechanism". *Photochemistry and Photobiology* 11.6 (1970), pp. 457–475.

[125] M Reza Razeghifard and Ronald J Pace. "Electron paramagnetic resonance kinetic studies of the S states in spinach PSII membranes". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1322.2-3 (1997), pp. 141–150.

[126] Hiroki Sakamoto et al. "Monitoring the reaction process during the S2→ S3 transition in photosynthetic water oxidation using time-resolved infrared spectroscopy". *Journal of the American Chemical Society* 139.5 (2017), pp. 2022–2029.

[127] Paul Greife et al. "The electron–proton bottleneck of photosynthetic oxygen evolution". *Nature* 617.7961 (2023), pp. 623–628.

[128] Tatsuki Shimizu, Miwa Sugiura, and Takumi Noguchi. "Mechanism of proton-coupled electron transfer in the S0-to-S1 transition of photosynthetic water oxidation as revealed by time-resolved infrared spectroscopy". *The Journal of Physical Chemistry B* 122.41 (2018), pp. 9460–9470.

[129] M Haumann et al. "Photosynthetic O2 formation tracked by time-resolved X-ray experiments". *Science* 310.5750 (2005), pp. 1019–1021.

[130] André Klauss, Michael Haumann, and Holger Dau. "Alternating electron and proton transfer steps in photosynthetic water oxidation". *Proceedings of the National Academy of Sciences* 109.40 (2012), pp. 16035–16040.

[131] Asmit Bhowmick et al. "Structural evidence for intermediates during O2 formation in photosystem II". *Nature* 617.7961 (2023), pp. 629–636.

[132] Takumi Noguchi. "Fourier transform infrared difference and time-resolved infrared detection of the electron and proton transfer dynamics in photosynthetic water oxidation". *Biochimica et Biophysica Acta (BBA) Bioenergetics* 1847.1 (2015), pp. 35–45.

[133] Takumi Noguchi et al. "Time-resolved infrared detection of the proton and protein dynamics during photosynthetic oxygen evolution". *Biochemistry* 51.15 (2012), pp. 3205–3214.

[134] Ivelina Zaharieva, Holger Dau, and Michael Haumann. "Sequential and coupled proton and electron transfer events in the S2→ S3 transition of photosynthetic water oxidation revealed by time-resolved X-ray absorption spectroscopy". *Biochemistry* 55.50 (2016), pp. 6996–7004.

[135] Fabrice Rappaport, Mireille Blanchard-Desce, and Jérôme Lavergne. "Kinetics of electron transfer and electrochromic change during the redox transitions of the photosynthetic oxygen-evolving complex". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1184.2-3 (1994), pp. 178–192.

[136] Jan P Dekker et al. "Kinetics of the oxygen-evolving complex in salt-washed photosystem II preparations". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 767.3 (1984), pp. 515–523.

[137] Han Bao and Robert L Burnap. "Structural rearrangements preceding dioxygen formation by the water oxidation complex of photosystem II". *Proceedings of the National Academy of Sciences* 112.45 (2015), E6139–E6147.

[138] Preston L Dilbeck et al. "The D1-D61N mutation in Synechocystis sp. PCC 6803 allows the observation of pH-sensitive intermediates in the formation and release of O2 from photosystem II". *Biochemistry* 51.6 (2012), pp. 1079–1091.

[139] Jürgen Clausen, Richard J Debus, and Wolfgang Junge. "Time-resolved oxygen production by PSII: chasing chemical intermediates". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1655 (2004), pp. 184–194.

[140] Felix Rummel et al. "Insights into PSII's S3YZ• State: An Electronic and Magnetic Analysis". *The Journal of Physical Chemistry Letters* 15.2 (2024), pp. 499–506.

[141] Paul Emma et al. "First lasing and operation of an ångstrom-wavelength free-electron laser". *nature photonics* 4.9 (2010), pp. 641–647.

[142] Matthieu Chollet et al. "The x-ray pump–probe instrument at the linac coherent light source". *Journal of synchrotron radiation* 22.3 (2015), pp. 503–507.

[143] Raymond G Sierra et al. "The macromolecular femtosecond crystallography instrument at the linac coherent light source". *Journal of synchrotron radiation* 26.2 (2019), pp. 346–357.

[144] A Baumbaugh et al. "Analysis of full charge reconstruction algorithms for X-ray pixelated detectors". *2011 IEEE Nuclear Science Symposium Conference Record*. IEEE. 2011, pp. 660–667.

[145] Scott C Jensen et al. "X-ray emission spectroscopy of biomimetic Mn coordination complexes". *The Journal of Physical Chemistry Letters* 8.12 (2017), pp. 2584–2589.

[146] Katherine M Davis et al. "X-ray emission spectroscopy of Mn coordination complexes toward interpreting the electronic structure of the oxygen-evolving complex of photosystem II". *The Journal of Physical Chemistry C* 120.6 (2016), pp. 3326–3333.

[147] Johannes Messinger et al. "Absence of Mn-centered oxidation in the S2→ S3 transition: implications for the mechanism of photosynthetic water oxidation". *Journal of the American Chemical Society* 123.32 (2001), pp. 7804–7820.

[148] Ivelina Zaharieva et al. "Room-temperature energy-sampling Kβ X-ray emission spectroscopy of the Mn4Ca complex of photosynthesis reveals three manganese-centered oxidation steps and suggests a coordination change prior to O2 formation". *Biochemistry* 55.30 (2016), pp. 4197–4211.

[149] Katherine M Davis et al. "Rapid evolution of the photosystem II electronic structure during water splitting". *Physical Review X* 8.4 (2018), p. 041014.

[150] Thomas Fransson et al. "Effects of x-ray free-electron laser pulse intensity on the Mn K β 1, 3 x-ray emission spectrum in photosystem II—A case study for metalloprotein crystals and solutions". *Structural Dynamics* 8.6 (2021), p. 064302.

[151] Guangye Han et al. "Molecular basis for turnover inefficiencies (misses) during water oxidation in photosystem II". *Chemical Science* 13.29 (2022), pp. 8667–8678.

[152] Pieter Glatzel et al. "Electronic structural changes of Mn in the oxygen-evolving complex of photosystem II during the catalytic cycle". *Inorganic chemistry* 52.10 (2013), pp. 5642–5644.

[153] Anne Marie March et al. "Feasibility of valence-to-core x-ray emission spectroscopy for tracking transient species". *The Journal of Physical Chemistry C* 119.26 (2015), pp. 14571–14578.

[154] Thomas Kroll et.al. "Stimulated X-Ray Emission Spectroscopy in Transition Metal Complexes". *Phys. Rev. Lett.* 120 (13 Mar. 2018), p. 133203.

[155] Christopher J Pollock et al. "Kβ mainline X-ray emission spectroscopy as an experimental probe of metal–ligand covalency". *Journal of the American Chemical Society* 136.26 (2014), pp. 9453–9463.

[156] Meiyuan Guo et al. "Iron Kβ X-ray Emission Spectroscopy: The Origin of Spectral Features from Atomic to Molecular Systems Using Multi-configurational Calculations". *The Journal of Physical Chemistry A* (2024).

[157] Nina Rohringer et.al. "Atomic inner-shell X-ray laser at 1.46 nanometres pumped by an X-ray free-electron laser". *Nature* 481 (Jan. 2012). URL: https://doi.org/10.1038/nature10721.

[158] M Beye et al. "Stimulated X-ray emission for materials science". *Nature* 501.7466 (2013), pp. 191–194.

[159] Hitoki Yoneda et al. "Atomic inner-shell laser at 1.5-angstrom wavelength pumped by an X-ray free-electron laser". *Nature* 524 (2015), pp. 446–449. DOI: 10.1038/nature14894.

[160] Thomas Kroll et.al. "Observation of Seeded Mn Kβ Stimulated X-Ray Emission Using Two-Color X-Ray Free-Electron Laser Pulses". *Phys. Rev. Lett.* 125 (3 July 2020), p. 037404.

[161] C Pellegrini. "X-ray free-electron lasers: from dreams to reality". *Physica Scripta* 2016.T169 (2017), p. 014004.

[162] Uwe Bergman, Vittal K Yachandra, and Junko Yano. *X-ray free electron lasers: applications in materials, chemistry and biology*. Vol. 18. Royal Society of Chemistry, 2017.

[163] Yu Zhang et al. "Generation of intense phase-stable femtosecond hard X-ray pulse pairs". *Proceedings of the National Academy of Sciences* 119.12 (2022), e2119616119. DOI: 10.1073/pnas.2119616119. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2119616119. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2119616119.

[164] Clemens Weninger and Nina Rohringer. "Transient-gain photoionization x-ray laser". *Physical Review A* 90.6 (2014), p. 063828.

[165] R. Bonifacio, C. Pellegrini, and L. Narducci. "Collective Instabilities and High Gain Regime in a Free Electron Laser". *Opt. Commun.* 50 (1984), pp. 373–378. DOI: 10.1016/0030-4018(84)90105-6.

[166] R Bonifacio et al. "Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise". *Physical review letters* 73.1 (1994), p. 70.

[167] C. Pellegrini, A. Marinelli, and S. Reiche. "The physics of x-ray free-electron lasers". en. *Reviews of Modern Physics* 88.1 (Mar. 2016), p. 015006. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.88.015006. URL: https://link.aps.org/doi/10.1103/RevModPhys.88.015006 (visited on 07/28/2022).

[168] D. Ratner et al. "Experimental Demonstration of a Soft X-Ray Self-Seeded Free-Electron Laser". *Phys. Rev. Lett.* 114 (5 Feb. 2015), p. 054801.

[169] L. H. Yu. "Generation of intense UV radiation by subharmonically seeded single-pass free-electron lasers". *Phys. Rev. A* 44 (8 Oct. 1991), pp. 5178–5193.

[170] Enrico Allaria et al. "Highly coherent and stable pulses from the FERMI seeded free-electron laser in the extreme ultraviolet". *Nat. Photon.* 6 (10 Oct. 2012), pp. 699–704. DOI: 10.1038/nphoton.2012.233. URL: http://dx.doi.org/10.1038/nphoton.2012.233.

[171] J. Amann et al. "Demonstration of self-seeding in a hard-X-ray free-electron laser". *Nature Photonics* 6 (Oct. 2012), pp. 693–698. DOI: 10.1038/nphoton.2012.180.

[172] Alberto A Lutman et al. "Fresh-slice multicolour X-ray free-electron lasers". *Nature Photonics* 10.11 (2016), pp. 745–750.

[173] C. Emma et al. "Experimental demonstration of fresh bunch self-seeding in an X-ray free electron laser". *Applied Physics Letters* 110.15 (2017), p. 154101. DOI: 10.1063/1.4980092. eprint: https://doi.org/10.1063/1.4980092. URL: https://doi.org/10.1063/1.4980092.

[174] Eugenio Ferrari et al. "Widely tunable two-colour seeded free-electron laser source for resonant-pump resonant-probe magnetic scattering". *Nature Communications* 7.1 (2016), p. 10343.

[175] Erik Hemsing, Aliaksei Halavanau, and Zhen Zhang. "Enhanced self-seeding with ultrashort electron beams". *Physical Review Letters* 125.4 (2020), p. 044801.

[176] Kwang-Je Kim et.al. "A Proposal for an X-Ray Free-Electron Laser Oscillator with an Energy-Recovery Linac". *Phys. Rev. Lett.* 100 (24 June 2008), p. 244802. DOI: 10.1103/PhysRevLett.100.244802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.244802.

[177] Gabriel Marcus et.al. "Refractive Guide Switching a Regenerative Amplifier Free-Electron Laser for High Peak and Average Power Hard X Rays". *Phys. Rev. Lett.* 125 (25 Dec. 2020), p. 254801. DOI: 10.1103/PhysRevLett.125.254801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.254801.

[178] Roberto Alonso-Mori et al. "Femtosecond electronic structure response to high intensity XFEL pulses probed by iron X-ray emission spectroscopy". *Scientific reports* 10.1 (2020), pp. 1–7.

[179] Toru Hara et al. "Two-colour hard X-ray free-electron laser with wide tunability". *Nature Communications* 4.1 (Dec. 2013), p. 2919. ISSN: 2041-1723. DOI: 10.1038/ncomms3919. URL: https://doi.org/10.1038/ncomms3919.

[180] Ichiro Inoue et al. "Two-color X-ray free-electron laser consisting of broadband and narrowband beams". *Journal of Synchrotron Radiation* 27.6 (Nov. 2020), pp. 1720–1724. DOI: 10.1107/S1600577520011716. URL: https://doi.org/10.1107/S1600577520011716.

[181] Ichiro Inoue et.al. "Generation of narrow-band X-ray free-electron laser via reflection self-seeding". *Nature Photonics* 13 (May 2019), p. 1. DOI: 10.1038/s41566-019-0365-y.

[182] Taito Osaka et.al. "A micro channel-cut crystal X-ray monochromator for a self-seeded hard X-ray free-electron laser". *Journal of Synchrotron Radiation* 26.5 (Sept. 2019), pp. 1496–1502. DOI: 10.1107/S1600577519008841. URL: https://doi.org/10.1107/S1600577519008841.

[183] Makina Yabashi, Hitoshi Tanaka, and Tetsuya Ishikawa. "Overview of the SACLA facility". *Journal of Synchrotron Radiation* 22.3 (May 2015), pp. 477–484. DOI: 10.1107/S1600577515004658. URL: https://doi.org/10.1107/S1600577515004658.

[184] K Tono et al. "Beamline, experimental stations and photon beam diagnostics for the hard x-ray free electron laser of SACLA". *New Journal of Physics* 15.8 (Aug. 2013), p. 083035. DOI: 10.1088/1367-2630/15/8/083035. URL: https://doi.org/10.1088/1367-2630/15/8/083035.

[185] Hirokatsu Yumoto et al. "Nanofocusing Optics for an X-Ray Free-Electron Laser Generating an Extreme Intensity of 100 EW/cm2 Using Total Reflection Mirrors". *Applied Sciences* 10.7 (2020). ISSN: 2076-3417. DOI: 10.3390/app10072611. URL: https://www.mdpi.com/2076-3417/10/7/2611.

[186] Kensuke Tono. "Fluid sample injectors for x-ray free electron laser at SACLA". *High Power Laser Science and Engineering* 5 (2017), e7. DOI: 10.1017/hpl.2017.6.

[187] Tetsuo Katayama et al. "A beam branching method for timing and spectral characterization of hard X-ray free-electron lasers". *Structural Dynamics* 3.3 (2016), p. 034301. DOI: 10.1063/1.4939655. eprint: https://doi.org/10.1063/1.4939655. URL: https://doi.org/10.1063/1.4939655.

[188] Ichiro Inoue et al. "Shortening X-Ray Pulse Duration via Saturable Absorption". *Phys. Rev. Lett.* 127 (16 Oct. 2021), p. 163903. DOI: 10.1103/PhysRevLett.127.163903. URL: https://link.aps.org/doi/10.1103/PhysRevLett.127.163903.

[189] Takashi Ida, M Ando, and H Toraya. "Extended pseudo-Voigt function for approximating the Voigt profile". *Journal of Applied Crystallography* 33.6 (2000), pp. 1311–1316.

[190] David F Hotz. "Gain narrowing in a laser amplifier". *Applied Optics* 4.5 (1965), pp. 527–530.

[191] A. Halavanau et al. "Progress Report on Population Inversion X-Ray Laser Oscillator at LCLS" (Bangkok, Thailand). International Particle Accelerator Conference 13. JACoW Publishing, Geneva, Switzerland, July 2022, TUPOPT047, pp. 1107–1109. ISBN: 978-3-95450-227-1. DOI: 10.18429/JACoW-IPAC2022-TUPOPT047. URL: https://jacow.org/ipac2022/papers/tupopt047.pdf.

[192] Giovanni De Ninno et al. "Chirped seeded free-electron lasers: self-standing light sources for two-color pump-probe experiments". *Physical review letters* 110.6 (2013), p. 064801.

[193] Taito Osaka et al. "A micro channel-cut crystal X-ray monochromator for a self-seeded hard X-ray free-electron laser". *Journal of Synchrotron Radiation* 26.5 (2019), pp. 1496–1502.

[194] Tao Liu et al. "Matching-based two-color X-ray free-electron laser generation utilizing laser–plasma accelerated electron beam". *High Power Laser Science and Engineering* 10 (2022), e1.

[195] AA Lutman et al. "Experimental demonstration of femtosecond two-color x-ray free-electron lasers". *Physical review letters* 110.13 (2013), p. 134801.

[196] Giuseppe Penco et al. "Two-bunch operation with ns temporal separation at the FERMI FEL facility". *New Journal of Physics* 20.5 (2018), p. 053047.

[197] Jumpei Yamada et al. "Extreme focusing of hard X-ray free-electron laser pulses enables 7 nm focus width and 1022 W cm- 2 intensity". *Nature Photonics* (2024), pp. 1–6.

[198] T. Kroll et al. "Observation of Seeded Mn Kβ Stimulated X-Ray Emission Using Two-Color X-Ray Free-Electron Laser Pulses". *Phys. Rev. Lett.* 125 (2020), p. 037404.

[199] T. Kroll et al. "Stimulated X-Ray Emission Spectroscopy in Transition Metal Complexes". *Phys. Rev. Lett.* 120 (2018), p. 133203.

[200] Julian Limburg et al. "Characterization of the O2-Evolving Reaction Catalyzed by [(terpy)(H2O) MnIII (O) 2MnIV (OH2)(terpy)](NO3) 3 (terpy= 2, 2 ': 6, 2 "Terpyridine)". *Journal of the American Chemical Society* 123.3 (2001), pp. 423–430.

[201] Evgeny L Saldin, EV Schneidmiller, and Mikhail V Yurkov. *The physics of free electron lasers*. Springer Science & Business Media, 2013.

[202] Anthony E Siegman. *Lasers*. University science books, 1986.

[203] Sergey V Kidalov and Fedor M Shakhov. "Thermal conductivity of diamond composites". *Materials* 2.4 (2009), pp. 2467–2495.

[204] Mengning Liang et al. "The coherent X-ray imaging instrument at the Linac Coherent Light Source". *Journal of synchrotron radiation* 22.3 (2015), pp. 514–519.

[205] S Boutet et al. "Design and construction of a population inversion x-ray laser at LCLS" (2023).

[206] Rana Hussein et al. "Structural dynamics in the water and proton channels of photosystem II during the S2 to S3 transition". *Nature communications* 12.1 (2021), p. 6531.

[207] Felix M Ho and Stenbjörn Styring. "Access channels and methanol binding site to the CaMn4 cluster in photosystem II based on solvent accessibility simulations, with implications for substrate water access". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1777.2 (2008), pp. 140–153.

[208] James W Murray and James Barber. "Structural characteristics of channels and pathways in photosystem II including the identification of an oxygen channel". *Journal of structural biology* 159.2 (2007), pp. 228–237.

[209] Albert Guskov et al. "Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride". *Nature structural & molecular biology* 16.3 (2009), pp. 334–342.

[210] Yasufumi Umena et al. "Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å". *Nature* 473.7345 (2011), pp. 55–60.

[211] Keisuke Saito, A William Rutherford, and Hiroshi Ishikita. "Energetics of proton release on the first oxidation step in the water-oxidizing enzyme". *Nature communications* 6.1 (2015), p. 8488.

[212] KN Trueblood et al. "Atomic dispacement parameter nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature". *Acta Crystallographica Section A: Foundations of Crystallography* 52.5 (1996), pp. 770–781.

[213] Hiroshi Ishikita et al. "Energetics of a possible proton exit pathway for water oxidation in photosystem II". *Biochemistry* 45.7 (2006), pp. 2063–2071.

[214] Naoki Sakashita et al. "Origins of water molecules in the photosystem II crystal structure". *Biochemistry* 56.24 (2017), pp. 3049–3057.

[215] Naoki Sakashita, Hiroshi Ishikita, and Keisuke Saito. "Rigidly hydrogen-bonded water molecules facilitate proton transfer in photosystem II". *Physical Chemistry Chemical Physics* 22.28 (2020), pp. 15831–15841.

[216] Serguei Vassiliev et al. "Tracking the flow of water through photosystem II using molecular dynamics and streamline tracing". *Biochemistry* 49.9 (2010), pp. 1873–1881.

[217] Sergey Vassiliev, Tatiana Zaraiskaya, and Doug Bruce. "Exploring the energetics of water permeation in photosystem II by multiple steered molecular dynamics simulations". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1817.9 (2012), pp. 1671–1678.

[218] Michael E Wall et al. "Biomolecular solvation structure revealed by molecular dynamics simulations". *Journal of the American Chemical Society* 141.11 (2019), pp. 4711–4720.

[219] David C Wych et al. "Liquid-like and rigid-body motions in molecular-dynamics simulations of a crystalline protein". *Structural Dynamics* 6.6 (2019).

[220] Sanchaita Rajkhowa and A Nath Jha. "Molecular dynamics simulations: A tool to investigate the interactions between biomolecules and nanoparticles". *Nov Sci Publ Inc* (2019), pp. 65–108.

[221] CompChems. *Molecular Dynamics Ensembles*. Accessed [Insert Date of Access]. Year of Access. URL: `https://www.compchems.com/molecular-dynamics-ensembles/#the-microcanonical-ensemble-nve`.

[222] Lisa J Lapidus. "Protein unfolding mechanisms and their effects on folding experiments". *F1000Research* 6 (2017).

[223] David Charles Wych. "Insights into the Modelling of Crystallographic Structure and Disorder from Molecular Dynamics Simulations of Protein Crystals". Submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Pharmaceutical Sciences. Doctoral dissertation. University of California, Irvine, 2021.

[224] Ralf W Grosse-Kunstleve et al. "The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework". *Journal of Applied Crystallography* 35.1 (2002), pp. 126–136.

[225] Moran Grossman et al. "Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site". *Nature structural & molecular biology* 18.10 (2011), pp. 1102–1108.

[226] Qun Wan et al. "Toward resolving the catalytic mechanism of dihydrofolate reductase using neutron and ultrahigh-resolution X-ray crystallography". *Proceedings of the National Academy of Sciences* 111.51 (2014), pp. 18225–18230.

[227] John CH Spence. *X-ray lasers for structure and dynamics in biology*. 2018.

[228] Henry N Chapman. "X-ray free-electron lasers for the structure and dynamics of macromolecules". *Annual review of biochemistry* 88 (2019), pp. 35–58.

[229] Thomas RM Barends et al. "Serial femtosecond crystallography". *Nature Reviews Methods Primers* 2.1 (2022), p. 59.

[230] Hans Frauenfelder, Gregory A Petsko, and Demetrius Tsernoglou. "Temperature dependent X-ray diffraction as a probe of protein structural dynamics". *Nature* 280 (1979), pp. 558–563.

[231] TP Singh, W Bode, and R Huber. "Low-temperature protein crystallography. Effect on flexibility, temperature factor, mosaic spread, extinction and diffuse scattering in two examples: bovine trypsinogen and Fc fragment". *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* 36.3 (1980), pp. 621–627.

[232] Andreas Ostermann et al. "Ligand binding and conformational motions in myoglobin". *Nature* 404.6774 (2000), pp. 205–208.

[233] Jessica L Thomaston et al. "XFEL structures of the influenza M2 proton channel: Room temperature water networks and insights into proton conduction". *Proceedings of the National Academy of Sciences* 114.51 (2017), pp. 13357–13362.

[234] John F Darby et al. "Water networks can determine the affinity of ligand binding to proteins". *Journal of the American Chemical Society* 141.40 (2019), pp. 15818–15826.

[235] Pawel A Janowski et al. "Peptide crystal simulations reveal hidden dynamics". *Journal of the American Chemical Society* 135.21 (2013), pp. 7938–7948.

[236] James M Holton et al. "The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures". *The FEBS journal* 281.18 (2014), pp. 4046–4060.

[237] Lars Meinhold and Jeremy C Smith. "Correlated dynamics determining X-ray diffuse scattering from a crystalline protein revealed by molecular dynamics simulation". *Physical review letters* 95.21 (2005), p. 218103.

[238] Lars Meinhold and Jeremy C Smith. "Protein dynamics from X-ray crystallography: Anisotropic, global motion in diffuse scattering patterns". *Proteins: Structure, Function, and Bioinformatics* 66.4 (2007), pp. 941–953.

[239] Michael E Wall. "Internal protein motions in molecular-dynamics simulations of Bragg and diffuse X-ray scattering". *IUCrJ* 5.2 (2018), pp. 172–181.

[240] Dorothee Liebschner et al. "Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix". *Acta Crystallographica Section D: Structural Biology* 75.10 (2019), pp. 861–877.

[241] Martyn D Winn et al. "Overview of the CCP4 suite and current developments". *Acta Crystallographica Section D: Biological Crystallography* 67.4 (2011), pp. 235–242.

[242] Ning Liu et al. "Simulating diffraction photographs based on molecular dynamics trajectories of a protein crystal: a new option to examine structure-solving strategies in protein crystallography". *IUCrJ* 10.1 (2023).

[243] P Joliot, G Barbieri, and R Chabaud. "A new model of the system of photochemical centers. II". *Photochemistry and photobiology* (1969).

[244] Azat Gabdulkhakov et al. "Probing the accessibility of the Mn4Ca cluster in photosystem II: channels calculation, noble gas derivatization, and cocrystallization with DMSO". *Structure* 17.9 (2009), pp. 1223–1234.

[245] Koji Ogata et al. "All-atom molecular dynamics simulation of photosystem II embedded in thylakoid membrane". *Journal of the American Chemical Society* 135.42 (2013), pp. 15670–15673.

[246] James Barber et al. "Structural model of the oxygen-evolving centre of photosystem II with mechanistic implications". *Physical Chemistry Chemical Physics* 6.20 (2004), pp. 4737–4742.

[247] Javier De Las Rivas and James Barber. "Analysis of the structure of the PsbO protein and its implications". *Photosynthesis Research* 81 (2004), pp. 329–343.

[248] David C Wych et al. "Molecular-dynamics simulation methods for macromolecular crystallography". *Acta Crystallographica Section D: Structural Biology* 79.1 (2023).

[249] David A Case et al. "The Amber biomolecular simulation programs". *Journal of computational chemistry* 26.16 (2005), pp. 1668–1688.

[250] Eric F Pettersen et al. "UCSF Chimera—a visualization system for exploratory research and analysis". *Journal of computational chemistry* 25.13 (2004), pp. 1605–1612.

[251] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". *SoftwareX* 1 (2015), pp. 19–25.

[252] Jochen S Hub et al. "Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge". *Journal of chemical theory and computation* 10.1 (2014), pp. 381–390.

[253] Gulshat T Ibragimova and Rebecca C Wade. "Importance of explicit salt ions for protein stability in molecular dynamics simulation". *Biophysical journal* 74.6 (1998), pp. 2906–2911.

[254] Miao Zhang et al. "Structural insights into the light-driven auto-assembly process of the water-oxidizing Mn4CaO5-cluster in photosystem II". *Elife* 6 (2017), e26933.

[255] Ernest ES Ong and Jong-Leng Liow. "The temperature dependent structure, hydrogen bonding and other related dynamic properties of the standard TIP3P and CHARMM-modified TIP3P water models". *Fluid Phase Equilibria* 481 (2019), pp. 55–65.

[256] Sereina Riniker. "Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview". *Journal of chemical information and modeling* 58.3 (2018), pp. 565–578.

[257] LLC Schrodinger. "The PyMOL molecular graphics system". *Version* 1 (2015), p. 8.

[258] Eva Chovancova et al. "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures" (2012).

[259] Abhishek Sirohiwal and Dimitrios A Pantazis. "Functional water networks in fully hydrated photosystem II". *Journal of the American Chemical Society* 144.48 (2022), pp. 22035–22050.

[260] Octav Caldararu et al. "Water structure in solution and crystal molecular dynamics simulations compared to protein crystal structures". *RSC advances* 10.14 (2020), pp. 8435–8443.

[261] Rachel J Service, Warwick Hillier, and Richard J Debus. "Evidence from FTIR difference spectroscopy of an extensive network of hydrogen bonds near the oxygen-evolving Mn4Ca cluster of photosystem II involving D1-Glu65, D2-Glu312, and D1-Glu329". *Biochemistry* 49.31 (2010), pp. 6655–6669.

[262] Sebastian Kozuch, Tim Schleif, and Amir Karton. "Quantum mechanical tunnelling: the missing term to achieve sub-kJ mol- 1 barrier heights". *Physical Chemistry Chemical Physics* 23.18 (2021), pp. 10888–10898.

[263] Bridget Salna et al. "Wide-dynamic-range kinetic investigations of deep proton tunnelling in proteins". *Nature Chemistry* 8.9 (2016), pp. 874–880.

[264] Judith P Klinman and Amnon Kohen. "Hydrogen tunneling links protein dynamics to enzyme catalysis". *Annual review of biochemistry* 82 (2013), pp. 471–496.

[265] Divya Kaur et al. "Proton exit pathways surrounding the oxygen evolving complex of photosystem II". *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1862.8 (2021), p. 148446.

[266] Friederike Allgöwer et al. "Molecular principles of redox-coupled protonation dynamics in photosystem II". *Journal of the American Chemical Society* 144.16 (2022), pp. 7171–7180.

[267] Koji Kato et al. "High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams". *Communications Biology* 4.1 (2021), p. 382.

[268] Michihiro Suga et al. "Light-induced structural changes and the site of O= O bond formation in PSII caught by XFEL". *Nature* 543.7643 (2017), pp. 131–135.

[269] Iris D Young et al. "Structure of photosystem II and substrate binding at room temperature". *Nature* 540.7633 (2016), pp. 453–457.

[270] Keisuke Kawakami and Jian-Ren Shen. "Purification of fully active and crystallizable photosystem II from thermophilic cyanobacteria". *Methods in Enzymology*. Vol. 613. Elsevier, 2018, pp. 1–16.

[271] Julia Hellmich et al. "Native-like photosystem II superstructure at 2.44 Å resolution through detergent extraction from the protein crystal". *Structure* 22.11 (2014), pp. 1607–1615.

[272] Paul Emsley et al. "Features and development of Coot". *Acta Crystallographica Section D: Biological Crystallography* 66.4 (2010), pp. 486–501.