Time-Resolved Crystallography using X-ray Free-Electron Laser

by

Shibom Basu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Approved November 2015 by the
Graduate Supervisory Committee

Petra Fromme, Chair
John C.H. Spence
Robert Ros
George Wolf
Raimund Fromme

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

Photosystem II (PSII) is a large protein-cofactor complex. The first step in photosynthesis involves the harvesting of light energy from the sun by the antenna (made of pigments) of the PSII trans-membrane complex. The harvested excitation energy is transferred from the antenna complex to the reaction center of the PSII, which leads to a light-driven charge separation event, from water to plastoquinone. This phenomenal process has been producing the oxygen that maintains the oxygenic environment of our planet for the past 2.5 billion years.

The oxygen molecule formation involves the light-driven extraction of 4 electrons and protons from two water molecules through a multistep reaction, in which the **O**xygen **E**volving **C**enter (OEC) of PSII cycles through 5 different oxidation states, S_0 to S_4. Unraveling the water-splitting mechanism remains as a grant challenge in the field of photosynthesis research. This requires the development of an entirely new capability, the ability to produce molecular movies. This dissertation advances a novel technique, **S**erial **F**emtosecond **X**-ray crystallography (SFX), into a new realm whereby such time-resolved molecular movies may be attained. The ultimate goal is to make a "molecular movie" that reveals the dynamics of the water splitting mechanism using time-resolved SFX (TR-SFX) experiments and the uniquely enabling features of **X**-ray **F**ree-**E**lectron **L**aser (XFEL) for the study of biological processes.

This thesis presents the development of SFX techniques, including development of new methods to analyze millions of diffraction patterns (~100 terabytes of data per XFEL experiment) with the goal of solving the X-ray structures in different transition states.

The research comprises significant advancements to XFEL software packages (e.g., *Cheetah* and *CrystFEL*). Initially these programs could evaluate only 8-10% of all the data acquired successfully. This research demonstrates that with manual optimizations, the evaluation success rate was enhanced to 40-50%. These improvements have enabled TR-SFX, for the first time, to examine the double excited state (S_3) of PSII at 5.5-Å. This breakthrough demonstrated the first indication of conformational changes between the ground (S_1) and the double-excited (S_3) states, a result fully consistent with theoretical predictions.

The power of the TR-SFX technique was further demonstrated with proof-of-principle experiments on Photoactive Yellow Protein (PYP) micro-crystals that high temporal (10-ns) and spatial (1.5-Å) resolution structures could be achieved.

In summary, this dissertation research heralds the development of the TR-SFX technique, protocols, and associated data analysis methods that will usher into practice a new era in structural biology for the recording of 'molecular movies' of any biomolecular process.

I dedicate this thesis to my parents and my mathematics teacher, Anindya Bhattacharyya (Gora Sir), without whom, I would have never pursued a career in science. Thank you for igniting my scientific imagination and passion to explore new frontiers. To my parents (especially my mother), thank you for being supportive, trusting me, and allowing me to pursue my dreams.

ACKNOWLEDGMENTS

First and foremost, I acknowledge my supervisor and mentor, Prof. Petra Fromme for her support and guidance throughout my Ph.D. She taught me what is meant by "being passionate" and "dedicated" to a project. Her immense enthusiasm and unfathomed passion for scientific research, brainstorming ideas, and her motherly caring for students motivated me constantly to produce best of my efforts. I shall always remember those sleepless nights that we spent during our Nature paper's review process. I aspire to follow the example she has shown me with her unceasing dedication of heart and soul to the pursuit of scientific excellence.

I am extremely grateful to Prof. John C. H. Spence for his support, advice, and his generous sharing of his laboratory space. Through his mentorship I learned from his vast knowledge of diffraction Physics. I am thankful for his open sharing of valuable suggestions and brainstorming ideas. I am honored to work with him as a scientist.

I have heartfelt appreciation for my Committee members – Dr. George Wolf and Dr. Robert Ros. I am deeply thankful to Dr. Nadia Zatsepin and Dr. Richard A. Kirian, who helped introduce me to this new field. I am profoundly thankful to Dr. Raimund Fromme for being my mentor, colleague, and advisor. For many more reasons, I will always be in his debt, and am proud to be his colleague.

Although I was a member of the Fromme Laboratory, throughout my Ph.D. program I primarily shared space within the Spence Laboratory in the physics department. Hence, my primary lab mates were from Spence Laboratory – Ganesh Subramanian, Yun Zhou, and Chufeng Li. We shared many conversations about various topics – from science,

TABLE OF CONTENTS

Table   Page

1. INTRODUCTION

Macromolecular X-ray crystallography (MX) is currently the most prolific technique used for the determination of molecular structures at the atomic resolution. There are 99,759 (~90% of total depositions) X-ray structure depositions in the Protein Data Bank (PDB) (Source: PDB; http://www.rcsb.org/pdb/statistics/holdings.do). By comparison, only ~10% of total deposited structures were determined by NMR (Nuclear Magnetic Resonance) and ~1% were determined by cryo electron microscopy (cryo-EM). This dissertation is focused on the development a new, state-of-the-art, technique for time-resolved crystallography, time-resolved Serial Femtosecond X-ray crystallography (TR-SFX). In this introductory chapter I present, 1) a brief review on X-ray crystallography, and 2) an overview of the intersection of X-ray free-electron laser (XFEL) technology. These technologies are applied to structural biology and herald the advent of the revolutionary technique for time-resolved structure determination and the making of "molecular movies," time-resolved serial femtosecond X-ray crystallography (TR-SFX).

1.1. X-ray Crystallography – from Macro to Micro

In the following sub-sections, the principles of macro-molecular crystallography (MX) will be discussed in brief. Then, a background and a brief history for the advent of micro-molecular crystallography are presented.

1.1.1. Principles of Macromolecular Crystallography

A crystal is a periodic array of identically ordered molecules in all three dimensions. The scientific study of crystals growth and the method to determine molecular structures based on crystals is called **crystallography**. Macromolecular crystallography (MX) is a technique, which uses high-energy hard X-rays to probe a crystal of biomolecules/protein molecules. Unlike other light sources (e.g., visible or infrared lights), X-rays are so penetrative and highly energetic that they cannot be focused by lenses. X-rays, having a wavelength between 1- to 2-Å, can resonate to the dimension of atomic distances in molecules. This property enables scientists to use X-ray sources and the diffraction power of crystals to solve structures in the field of MX. The success of the MX method requires two major ingredients – a high-energy X-ray source and large crystals of pure biomolecules. Crystals are generally identified by their specific geometry (i.e., with embedded symmetry) and they often form objects of specific shapes with sharp edges. A crystal is imagined as a translational periodic arrangement of smallest repetitive units, called the **unit cells**. The smallest unit of a crystal structure to which symmetry operations can be applied to generate the complete unit cell (repetitive unit in crystal) is called an **Asymmetric unit**. Based on the geometry (i.e., lengths of the sides and adjacent angles) of the unit cell, 7 fundamental lattice types or crystallographic systems can be considered (See, Table 1.1). These 7 systems can be divided into 32 **point-groups** based on different degrees of symmetry. The crystallographic point-group is a set of symmetry operations, such as rotations and reflections that move the faces and other directions of
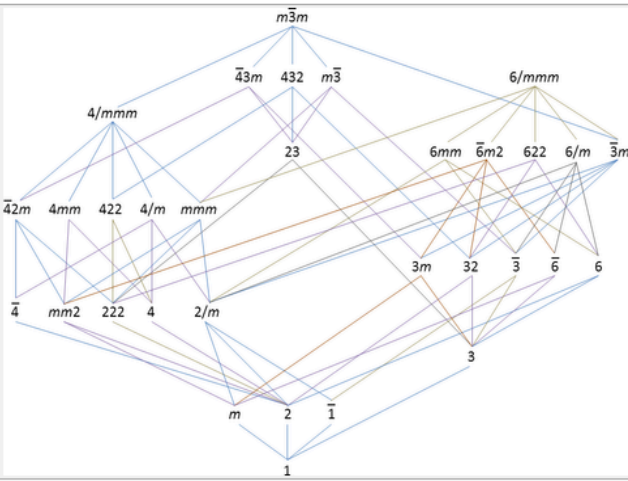
the crystal around a central fixed point. There are several types of notation conventions to represent point-groups. Here, Hermann-Mauguin notations are provided, Table 1.2. The symmetry elements, such as center of symmetry, mirror-plane, gliding plane, translational axis, rotational axis, and screw axis, can be combined in various combinations to generate 230 different arrangements, called space groups. Further details of symmetry and space groups are beyond the scope of this dissertation. The interested reader is referred to the "International Tables for Crystallography", Volume A, 2006 (Hahn, 2006), for a comprehensive treatment of the subject.

Table 1.1: Seven Crystallographic Lattice Types. This table indicates differences in length of sides and angles. It also points out the number of lattices in each lattice system (Hahn, 2006)(ICT, 2011).

Sr. No.	Crystal System	Axial length of Unit Cell	Inter axial angles	Number of Lattice in the system
1	Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^0$	3
2	Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^0$	2
3	Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^0$	4
4	Monoclinic	$a \neq b \neq c$	$\alpha = \beta = 90^0 \neq \gamma$	2
5	Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^0$	1
6	Trigonal	$a = b = c$	$\alpha = \beta = \gamma < 120^0, \neq 90^0$	1
7	Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^0, and \ \gamma = 120^0$	1

Table 1.2: Table for the Subgroup Relationships of 32 Crystallographic Point Groups. Point groups are shown along with corresponding seven crystallographic lattice types (Source: Wikipedia).

Class	Group names						
Cubic	23	$m\bar{3}$		432	$\bar{4}3m$	$m\bar{3}m$	
Hexagonal	6	$\bar{6}$	$6/m$	622	6mm	$\bar{6}2m$	$6/mmm$
Trigonal	3	$\bar{3}$		32	3m	$\bar{3}m$	
Tetragonal	4	$\bar{4}$	$4/m$	422	4mm	$\bar{4}2m$	$4/mmm$
Monoclinic Orthorhombic	2		$2/m$	222	m	mm2	mmm
Triclinic	1	$\bar{1}$					

Subgroup relations of the 32 crystallographic point groups
(rows represent group orders from bottom to top as: 1,2,3,4,6,8,12,16,24, and 48.)

When large biomolecules (e.g., proteins, DNA, or viruses, etc.) are studied in the form of crystals (with typical size ranging from 100-μm to 1-mm), the corresponding field of study is called **Macro-molecular X-ray crystallography (MX)**. For large biomolecules, composed primarily of C, H, O, N, and S atoms, crystallization is challenging, mainly because crystals of these molecules rely on van-der Waals' interactions, weak hydrogen bonds, or electrostatic interactions (c.f., Section 1.1.2 for physical chemistry of protein crystallization) in order to nucleate and grow. MX often uses high-energy X-ray sources (synchrotrons) which provide a photon flux of ~10^{11}-10^{13} photons/sec and brightness of ~10^{18} photons/mm^2/mrad^2/sec/0.1%BW (bandwidth) with an energy range from 6-keV to 15-keV, with a maximum brightness at 12-keV (Robinson et al., 2010). At these energies, MX relies on the amplifying diffraction power of crystals, which is a result of the additive constructive interference from elastic scattering of X-ray

photons by valence electrons of the atoms of the repeat unit molecules constituting the crystals. In general, an entire diffraction data set is collected from a single *large* crystal (~10 to 100 μm in all dimensions), mounted on a goniometer head that rotates the crystal. Therefore, the size of the beam determines the required size of the crystal. However, recent advances in X-ray beam focusing capabilities have produced micro-focus beamlines at various synchrotron facilities worldwide (e.g., Diamond light source in UK, Petra III at DESY in Germany, Advanced Photon Source at Chicago, USA, and Swiss Light Source in Switzerland). These micro-focus beamlines, have a very narrow X-ray beam focus of ~10-μm. This allows dataset collection from crystals as small as 5-μm to 15-μm size. In all cases, to collect complete data set (i.e., to sample the full reciprocal volume), the crystal is rotated by 180° (in ideal case), while collecting X-ray diffraction data. This rotational oscillation process ensures that each and every Miller plane (i.e., set of *hkl* planes; c.f., Section 1.3) has satisfied Bragg's condition (i.e., all Bragg reflections have been recorded; c.f., Section 1.3). In MX, to protect crystals from X-ray radiation damage, crystals are cryo-cooled (~100K) throughout data collection. Ada Yonath was the first to apply the method of cryo-crystallography (Yonath et al., 1987), which she used to determine the three-dimensional architecture of ribosomal particles (i.e., the complex responsible for synthesis of proteins cells). Later, Hopes et al. (Hope, 1988)(Hope et al., 1989) further advanced the method. Before, Yonath's work, X-ray diffraction data was collected at room temperature with limited success. However, with the many technical advances achieved since 1989, recently studies are revisiting the utility and efficacy of room temperature data collection (Fraser et al., 2011).

1.1.2. Physical Chemistry of Protein Crystallization

Nucleation and growth of crystals of proteins/biomolecules depend on many factors. The most important factor is to achieve a supersaturation state where crystal nucleation is thermodynamically favorable. Later in the crystal growth phase, as the crystal traverses the phase space to a saturated solution condition, the concentration of solute reaches equilibrium with the solution. At an equilibrium condition the amount of solute present in the solution is such that neither growth nor dissolution is favored. As such, a saturated solution corresponds to a thermodynamic equilibrium between the solution and solid phases. A saturated solution can be expressed in terms of chemical potentials of each species, represented by 'i', as,

$$\mu_{ic} = \mu_{is} = \mu_0 + RTln(\gamma c_i),\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\text{(Eq. 1.1.2a)}$$

where:

μ_{ic} = Chemical potential of species 'i', in crystal phase,

μ_{is} = Chemical potential of species 'i' in solution phase,

γ = Activity coefficient of species 'i',

μ_o = Standard chemical potential for the species 'i', and

c_i = Concentration of species 'i'.

Equation 1.1.2a describes a saturated solution, where the chemical potentials are equal for each species in two different phases. Supersaturation occurs when the chemical

potential of a solute in solution (i.e., μ_{is}) is greater than the crystal phase chemical potential (i.e., μ_{ic}). A supersaturated state may be achieved within a phase space by a number of paths by varying key process parameters that affect chemical potential of the solute (e.g., temperature, protein concentration, pressure, activity coefficient, salt concentration, etc.). Supersaturation is the driving force for crystal nucleation and growth and can only be reached by traversing the saturated phase state. Therefore, protein solubility is a necessary pre-requisite for supersaturation and methods that enable its attainment must therefore be determined. The unique properties of every protein make the discovery of these methods an informed, but challenging process in many cases.

The process of crystal growth can be most effectively presented with the assistance of the representative phase diagram shown in Fig. 1.1. Protein concentration is presented on the ordinate, and salt/precipitant concentration on the abscissa. Four phase space zones are identified: 1) the unsaturated zone, where everything is stable; 2) the metastable zone, where crystal growth occurs; 3) the nucleation zone, where crystal originally start forming; and 4) the precipitation zone, where proteins agglomerate together in an amorphous form.
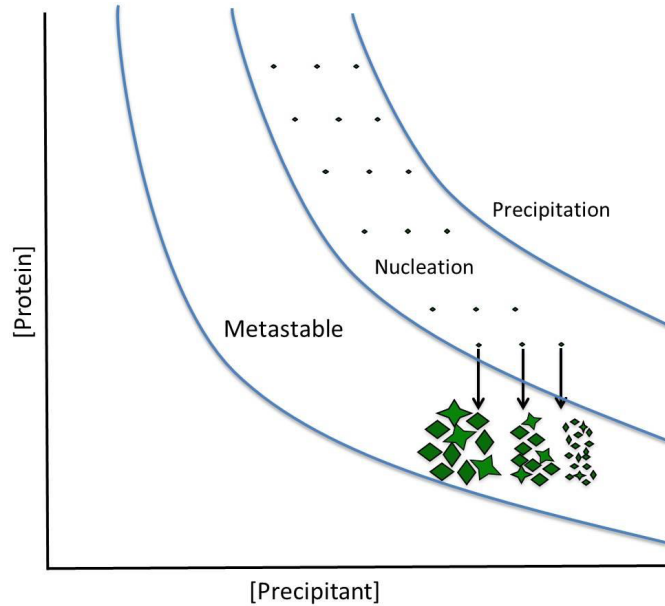
Figure 1.1: Schematic phase diagram for Protein Crystallization. This is shown with suitable starting point in nucleation zone. The nucleation frequency increases with increasing supersaturation, resulting in small crystals (Source: Kupitz et al., 2014).

The metastable phase is a supersaturation state, where crystal growth begins. However, crystal nucleation and growth will not spontaneously occur for any supersaturated solution. To form crystalline phase, the solvent-solute system must reach a point when degree of supersaturation is high enough to cross a critical activation free energy barrier (ΔG_g). This free energy activation barrier is represented by the following equation,

$$\Delta G_g = \left[-\frac{kT(4\pi r^3)}{V ln\beta} \right] + 4\pi\gamma r^2, \quad \text{.. (Eq. 1.1.2b)}$$

where:

V = Volume of the molecule inside the crystal,

β = Degree of supersaturation,

r = Radius of nucleus formation,

γ = Interfacial free energy between nucleus and solution, and

ΔG_g is a function of negative volume term and a positive surface free energy term.

In order to obtain nucleation and crystal growth, the degree of supersaturation (i.e., β) must increase, or the interfacial (crystal/solution) free energy term must decrease. Either of these process changes will decrease ΔG_g, making crystallization thermodynamically favorable. As such, a high level of supersaturation decreasing ΔG_g favors nucleation and growth of a new crystalline phase. The presence of foreign particles further reduces the interfacial surface free energy term, and increases the frequency of nucleation. For this reason, the use of nucleating agent particles has been frequently employed in crystallography.

However, higher supersaturation state levels are not always better. At higher supersaturation states, when nuclei are larger than the critical radius (i.e., r term in Eq. 1.1.2b), nucleation is certainly observed. Unfortunately, it is difficult to grow large well-ordered crystals as nucleation and crystal growth both compete for available solution phase protein. By way of example, nucleation starts in nucleation zone, followed by crystal growth in metastable state. If supersaturation is just high enough to reach nucleation zone, the nuclei formed will be unstable and while nucleation will not occur, only pre-formed crystals can grow. Conversely, if a phase space trajectory reaches a supersaturation state on the borderline between the nucleation and precipitation zones, this condition will lead to very fast disordered crystal growth. Another undesirable outcome occure, if supersaturation is reached at low protein concentration which leads to amorphous agglomeration of protein precipitants. Therefore, it is critical to reach the

correct degree of supersaturation by the right phase space trajectory, to cross the activation barrier ΔG_g, and manage the subsequent phase space trajectory throughout the crystal growth process in order to obtain high quality (low disorder) crystals. This convergence of process parameters that must be precisely controlled in order to achieve "optimal" crystal quality is a challenging problem in biological materials science that has traditionally be highly artisan and time-consuming in its practice. In recent years and with the advent of high-throughput screening method this art has been greatly reduced to a scientific practice. Nevertheless, the essential process of attaining high quality protein crystals for structure determination using X-ray crystallography methods remains a vigorous and challenging field that remains rich in scientific discovery opportunity.

1.1.3. X-ray Radiation Damage – Limitations in X-ray Crystallography

MX is the most mature, ubiquitous, and proven technique for protein structure determination using X-ray diffraction. While MX has demonstrated powerful capabilities, MX also has detracting technical attributes that limit its full utility. The most notorious of these detractors is X-ray radiation damage. While its effect is clearly observed, the mechanism of X-ray induced protein crystal damage is not completely understood. Current interpretations suggest that the effects of radiation damage are stochastic in nature (Nave, 1995), with additional studies (Ravelli and McSweeney, 2000)(Nave, 1995)(Henderson, 1995)(Singh and Singh, 1982) presenting research that aims to classify

the types of radiation damage observed in biomolecules exposed to X-ray radiation. In all cases X-ray induced damage is classified into two major groups – primary and secondary. When X-rays interact with protein crystals, diffraction, absorption, and scattering phenomenon are observed. In the first case, X-rays interact by elastic Rayleigh scattering, which provides diffraction patterns. The remaining phenomenon, photoelectric absorption, Auger emission, and Compton scattering are the cause of 'primary' radiation damage which occurs at a very fast time-scale, i.e., in the order of femtoseconds (Henderson, 1995). The absorbed energy is due to inelastic scattering events that are not uniformly distributed within the crystal. The photoelectric/photoionization events generate free-radicals and release electrons which are destructive to biological molecules, and which stay solvated in the crystals. This process occurs at picosecond time-scale (Singh and Singh, 1982). These free-radicals (e.g., hydroxyl, hydrogen radicals) and solvated electrons can diffuse and break various bonds (e.g. disulphide bonds, decarboxylation of acid group in amino acids) within the macromolecular crystals in microsecond to millisecond time-scale at room temperature (Henderson, 1995). The breaking of disulphide bonds, decarboxylation of acid groups in the macromolecular crystals by free-radicals are considered 'secondary' damage (Ravelli and McSweeney, 2000). The effects of primary damage in biomolecular structures is not understood and considered to be non-specific. On the contrary, the effects of 'secondary' damage are more specific. Ravelli and McSweeney (Ravelli and McSweeney, 2000) have shown in their work that X-ray exposure can bias structures by leaving 'damage induced marks' by breaking disulphide bonds, decarboxylation of acid group of amino acids, and generating

11

secondary radicals, among other suspected but lesser affects. They showed (Ravelli and McSweeney, 2000) that the secondary damage occurs before the diffraction power of crystals decays, even at cryogenic temperatures. Secondary damage is most pronounced in the heavy atoms region in the biomolecules. Heavy atoms, containing more electrons in the valence shell, interact more strongly with X-rays, resulting in fast Auger decay. Due to photo-reduction, the oxidation states of affected heavy atoms change, leading to photo-damaged structures (Yano et al., 2005)(Garman, 2010). For these reasons, MX data collection is most commonly performed under cryo-cooled conditions to slow down the diffusion of free-radicals (i.e., secondary damage). Nevertheless, recent spectroscopic and computational studies (Garman, 2010)(Isobe et al., 2012)(Galstyan et al., 2012) show that most of the metallo-protein structures determined by MX still suffer from secondary radiation damage, most prominently at the sites of electron rich metal atoms due to the longer exposure time and cumulative nature of dose deposition. Further complicating structural biology considerations, cryogenic condition also leads to 'freezing artifacts' that can hamper the attainment of native protein molecule conformations. Minimizing radiation damage is therefore a primary motivator for the use of the extremely brilliant and fast pulses (10-fs to 100-fs) provided by X-ray Free-Electron Laser (XFEL) facilities. These properties enable X-ray diffraction patters to be recorded before 'secondary' damage caused by reactive free-radicals has occurred. Within this context, XFEL produced diffraction patterns are said to "outrun" damage.

1.1.4. Diffract before Destroy – Advent of FELs

The 'diffract before destroy' principle was first conceptualized by Solem in 1986 (Solem, 1986). He theoretically proposed that for a biological specimen (e.g., cell, viruses), it would be possible to obtain their images/diffraction patterns with extremely intense X-rays if the exposure time was sufficiently short, thereby enabling the recording of diffraction patterns before the biological molecules were obliterated by the beam intensity. Building on this concept, an *in silico* simulation was performed using the protein lysozyme as an exemplar case (Neutze et al., 2000). The goal of this *in silico* work was to simulate the time-course of energy deposition on single molecules. In their simulation, they considered photon flux of 3 x 10^{12}/pulse (12-keV) with an interaction spot 100-nm diameter. This photon flux is much higher than that of strongest synchrotron facility (5 x 10^{12}/second at a focal point of 5 to 10-μm). They considered the representative lysozyme molecule to be bound with solvent molecules in a gas phase. At a given deposited energy, the velocities of photoelectrons (43-nm/fs) and Auger electrons (7-nm/fs) are high enough to cause 'primary' damage or Coulomb explosion on the femtosecond order of magnitude timescale (Neutze et al., 2000). In Fig. 1.2, Neutze et al., have shown that at the given deposited energy, a 2-fs exposure time would be short enough such that kinetic energy of the photoelectrons would not grow appreciably. On the other hand, for a 10-fs or 50-fs exposure time, kinetic energy of the photo and Auger electrons would be large enough to destroy the lysozyme molecule. This simulation provided a theoretical foundational basis for Solem's 'diffract-before-destroy' concept. Of importance, the parameters of this simulation are a representative of a single lysozyme

13

molecule, not an extended crystal. A biomolecule crystals can tolerate relatively longer exposure times of ~40 to 50-fs (Chapman et al., 2011)(Barty et al., 2012)(Boutet et al., 2012) when exposed to the energy and fluence of current XFEL beamlines.
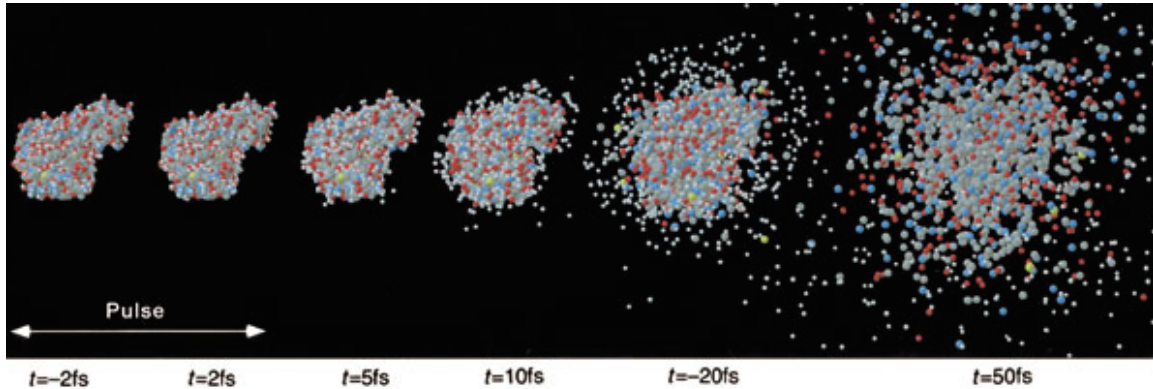
Figure 1.2: Progression of X-ray Induced Radiation Damage on a Lysozyme Molecule. A lysozyme protein molecule was exposed to X-ray pulse with an FWHM 2-fs and disintegration followed in time. The integrated X-ray photon flux of 3 x 1012/pulse (12 keV) at 100-nm diameter interaction region was considered. The first two structures (before and after 2-fs pulse) are identical because of inertial delay in explosion (Source: Neutze et al., 2000).

The first experimental evidence of 'diffract-before-destroy' was provided by the pioneering work of Henry Chapman et al., (Chapman et al., 2006). They performed an experiment at a soft X-ray free-electron laser (FLASH, Germany) showing that if the radiation dose is quickly (i.e., on the order of femtosecond exposure) delivered to an object, a single diffraction pattern can be obtained before the object is destroyed (Chapman et al., 2006). This was the first experimental proof of the 'diffract-before-destroy' concept. In 2009, when World's first hard X-ray free-electron laser (XFEL) became operational, the principle of 'diffract-before-destroy' was proven experimentally on one of the largest membrane protein complexes, Photosystem I (Chapman et al., 2011). The outcome of this experiment is discussed in Section 1.1.5.

14

1.1.5. Motivation behind X-ray Free-Electron Lasers

It is a dream of every structural biologist to study biomolecules in action and to elucidate fundamental biological processes. This dream has created demand for a new tool which can realize molecular movies and merge with other analytic and simulation oriented techniques. Realizing a hard X-ray free-electron laser is requisite to this objective. The world's first femtosecond pulsed hard X-ray laser, LCLS (Linac Coherent Light Source), began operating at the SLAC (Stanford Linear Accelerator Center) in 2009. The LCLS produces a hard X-ray laser (i.e., extremely coherent; See section 1.1.6 for description of XFELs), with 10^{12} times higher peak brilliance compared to the most advanced 3rd generation synchrotron (See Fig. 1.3). This means that when focused, any material placed into the X-ray beam path is converted to a plasma state hotter than the surface of the sun. Emma et al. published the capabilities and instrument specifications of the LCLS in 2010 (Emma et al., 2010). The LCLS XFEL provides hard X-ray laser pulses with a pulse length of 10 – 100-fs. This exposure time is too short for 'secondary' damage to be detected (See section 1.1.2) even though the deposited radiation dose is much higher than in MX (Barty et al., 2012)(Lomb et al., 2012)(Spence et al., 2012). Therefore, 'secondary' damage in inclusive of breaking of disulphide bonds, decarboxylation, photo-reduction of heavy elements, and formation of reactive free-radicals is not seen at the XFEL data. The promise of use of XFELs in structural biology has already been demonstrated in several studies (Chapman et al., 2011)(Boutet et al., 2012)(Redecke et al., 2013) using submicron protein crystals.

15

XFEL-based crystallography techniques have numerous advantages over MX techniques. As previously discussed, MX techniques suffers from 'secondary' radiation damage because of the long exposure time. In XFEL, use of femtosecond pulses of X-ray ensures very short exposure time. Therefore, diffraction patterns are collected before 'secondary' damage sets in, it is said that the detection process 'outruns damage.' It is noted that XFEL cannot overcome 'primary' damage, caused by photoionization, Compton scattering, auger emission, etc., and the effect of 'primary' damage is a very active field of study in XFEL research (Erk et al., 2014). As a further advantage, XFEL works at room temperature. As such, protein crystals are closer to the native conformation of the protein structures, unlike MX which is typically done in cryogenic conditions. In contrast to MX, XFEL requires only tiny submicron crystals because the XFEL beam is micro-focused (typically ~1-µm to 100-nm). The ability to use submicron scale crystals avoids the 'hassle' of growing large, uniform, high-quality single protein crystals. For example: Photosystem I, one of the largest membrane protein complexes, required thirteen years from the time of first trace of micro crystals (Witt et al., 1988) to the time when large good diffracting crystals (Jordan et al., 2001) were attained. The advent of XFEL technology renewed the interests and hopes of structural biologists that the process of rapidly solving challenging and unknown protein structures based on small (nano/micro) crystals would be possible. Fortifying this hope, in 2009, the first SFX experiment was carried out at the LCLS (SLAC) using submicron sized crystals (200-nm to 2-µm) of Photosystem I (PSI), one of the largest membrane protein complexes. This work was performed at a photon energy of at 1.8 keV ($\lambda = 6.9$-Å) (Chapman et al., 2011).

The PSI experiment was remarkable for many reasons: (a) it proved that the 'diffract-before-destroy' principle for protein crystals; (b) the diffraction data was collected using three different x-ray pulse lengths (10, 70, and 200-fs); (c) it also proved that unlike MX, collecting data from many 'partial' Bragg reflections can be integrated over many diffraction 'snapshots' to achieve a complete data set or to summarize a partial to full Bragg reflection dataset (Kirian et al., 2011). In a unique way, this experiment established that this new type of data could be analyzed to generate meaningful structure factor information (See data analysis Section 1.3). The interested reader is encouraged to engage excellent reviews of this extended subject (Fromme and Spence, 2012)(Spence et al., 2012) for a more in-depth treatment of the subject.
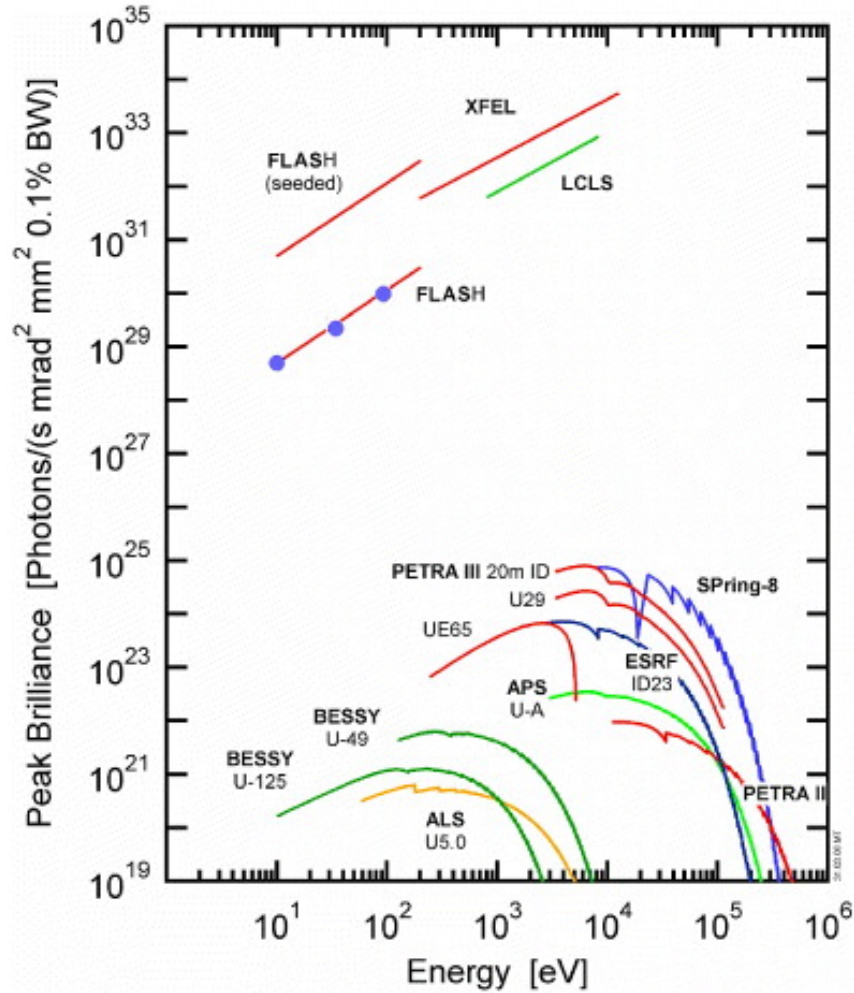
Figure 1.3: Comparison of the Peak-Brilliances of Various Synchrotron Sources with XFEL Facilities (Source: Robinson et al., 2010).

1.1.6. X-ray Free-Electron Lasers – a Brief Description

Like conventional optical lasers, electromagnetic radiation, produced by X-ray Free-Electron Lasers (XFEL) possesses high brightness, narrow bandwidth, and high coherence. However, there are also significant differences. The differences originate from the distinct mechanisms of generating XFEL radiation (Huang and Kim, 2007).

18

In an XFEL, a Copper cathode, an electron emitter, releases electrons as a result of photoemission. These photo-emitted electrons are accelerated close to the speed of light (i.e., at relativistic velocities) in a linear particle accelerator (LINAC) in the presence of an electromagnetic field. Accelerated electrons then enter the undulator section (See Fig. 1.4). The undulator is a long periodic array of dipole magnets with alternating polarity as shown in Fig. 1.4a. Electrons interact with the magnetic field of the undulator and experience a periodic Lorentz force. As a result, while traversing through the undulator region, they wiggle along their paths and emit electromagnetic radiation on every bend of their path. The emission and amplification of X-ray radiation is created by the formation of microbunches of electrons by Self-Amplified Spontaneous Emission (SASE) (Huang and Kim, 2007). While traveling through the undulator, electrons, oscillating in transverse direction (i.e., transverse velocity V_T), interact with the transverse magnetic field (B-field) of the previously emitted wave, causing longitudinal Lorentz force. This Lorentz force modifies the longitudinal velocity of the electron and pushes them to form micron-sliced bunches, so called micro-bunches (See Fig. 1.4b), with the periodicity being λ (i.e., correlated to the emitted wavelength). The electrons within a micro-bunch oscillate together under the effect of undulator and emit radiation in a correlated way (See Fig. 1.4d). However, to sustain the micro-bunch formation, the speed of electrons has to be slightly slower than the speed of emitted waves ($u < c$). Otherwise, after one-half of the magnetic period, the transverse velocity of the electrons and the Lorentz force direction would be reversed preventing the micro-bunching (Margaritondo and Rebernik Ribic, 2011). Continuous formation of electron micro-bunches through the undulator

region produce a correlated emission of waves that leads to a gradual amplification of wave intensities. The amplification of wave intensities cannot increase indefinitely and saturation does occur (See Fig. 1.4e). In order to make an X-ray laser, one method (adopted for the LCLS) is to reach the saturation limit before the undulator length ends (Emma et al., 2010). In a typical laser, the saturation limit or amplification intensity is obtained by using an external optical cavity (e.g., made of optical mirrors). For X-rays, it is 'ineffective' to create an optical cavity with mirrors. Therefore, a 'one pass' strategy of reaching the saturation limit is implemented with a very long undulator - the optimal strategy for obtaining lasing action from an XFEL. At LCLS, the undulator length is 134-m, which is much longer than the undulator for any synchrotron facility (typically 1 to 4-m).
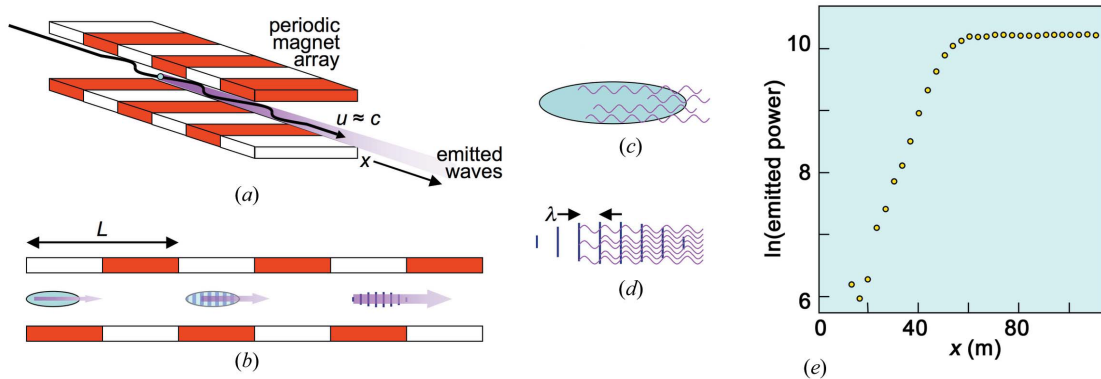
Figure 1.4: Illustration of X-ray FEL Generation and Mechanism. (a) Optical amplification and emission of X-ray waves, produced by electrons traveling through the periodic array of dipole magnets (undulator) at the relativistic velocities. (b) The first wave emitted from the single bunch of randomly oriented electrons lead to form micro-bunches gradually through the undulator. (c) Uncorrelated waves emitted from a single electron bunch. (d) Micro-bunches with electrons oscillating altogether lead to emit waves coherently. (e) As the microbunches of electrons advance through the undulator length, the amplification of emitted wave intensities increases exponentially and reaches a saturation limit (Source: Margaritondo and Rebernik Ribic, 2011).

1.1.7. Serial Femtosecond Crystallography (SFX) – a New Technique

Serial femtosecond crystallography (SFX) is a new crystallographic technique developed specifically for the data collection at XFEL instruments. The X-ray peak brilliance, offered by the LCLS, is 10^{12} higher than the brightest synchrotron facility. XFEL pulses have a very narrow focus of ~1-μm or 100-nm can and obliterate (i.e., convert to the plasma state of matter) any materials, that are placed into the beam path. Therefore, data collection from a single crystal is completely impractical, unless the beam is attenuated by several orders of magnitude. That is why, SFX techniques use a completely new sample delivery systems that are either fluidic-based (DePonte et al., 2008)(Weierstall et al., 2012) or electro-spray-based (Sierra et al., 2012). These delivery systems provide a stream of crystals to the X-ray interaction region. 'Serial', by definition, means that crystals will be delivered in a fully-hydrated state within the stream in a serial fashion. Figure 1.5 illustrates a typical SFX experimental set up that has been used for numerous experiments by the international community. 'Femtosecond' implies that the exposure time to the X-ray is in the fs time scale and that diffractions snapshots acquired outrun 'secondary' damage. There are several studies, which showed that even though the deposited amount of dose in XFEL is almost at tolerable dose limit, still the exposure time is too short to observe the effects of damage (Lomb et al., 2011)(Barty et al., 2012).

The injector system, which delivers the crystals in random orientations, uses a High Performance Liquid Chromatography (HPLC) pump to create the pressure difference

(Weierstall et al., 2012). The flow of the crystals is co-axially focused with helium gas flow, leading to form the thin jet (diameter = ~4-μm) of crystals.

In MX, time-resolved crystallography can be performed for only reversible processes through Laue crystallography (discussed in later section). But, in SFX method, irreversible reactions with very high temporal resolution (in order of femtosecond time) can be studied (Aquila et al., 2012). Because most biomolecular processes are dynamically very fast and also irreversible, SFX provides a uniquely enabling analytic capability for studying such ultra-fast processes. Thus, the SFX technique overcomes the limitations of MX. Unlike MX, the SFX method is based on collecting diffraction patterns from the stream of crystals, delivered into the interaction region in random orientation. This means that each diffraction snapshot is collected from each different crystal in its random orientation. This produces a "new" type of data, requiring new algorithm development for the analysis of crystallographic datasets (Spence and Chapman, 2014). Thus, with the advent of XFEL technology, and SFX, a new avenue of structural biology has emerged (Fromme and Spence, 2011) with many promises and capabilities.
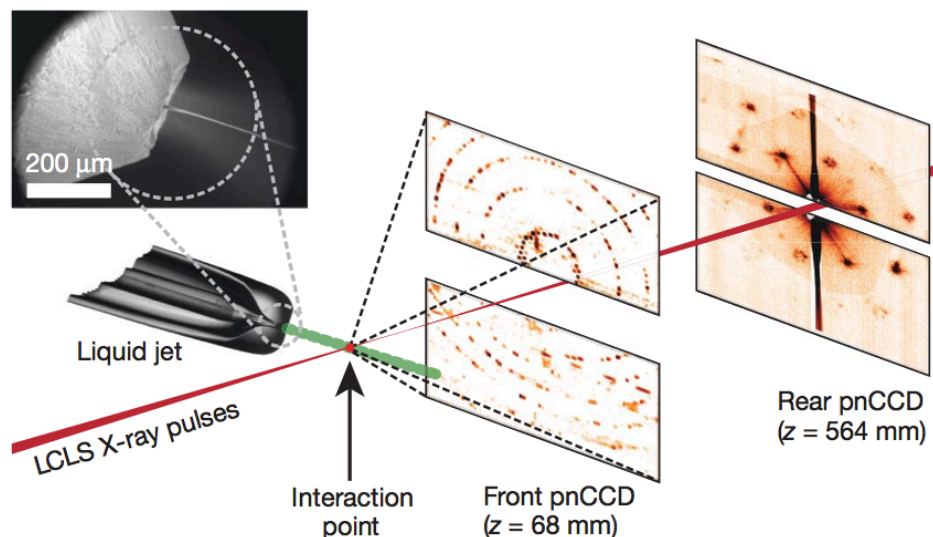
Figure 1.5: The Experimental Set-up for the Serial Femtosecond Crystallography Experiment (Chapman et al., 2011). The gas focused liquid injector (with 4-µm diameter jet), developed by Weierstall et al., 2012 is used to deliver sample continuously at a speed of 10-m/s. Inset, environmental scanning electron micrograph of the nozzle, flowing jet, and focusing gas. The diffraction data is collected on two pairs of pnCCD detectors. Nano-crystals arrive at random time and orientations in the beam and the probability of hitting one varied with the crystal concentration.

1.1.8. Comparison Existing and Upcoming XFELs Facilities

As previously mentioned, the first XFEL, i.e., LCLS, was commissioned in 2009 at SLAC, California. Soon after this, Japan built an XFEL and commissioned this faculty in 2011, SACLA. However, at the time of this writing (i.e., till 2015) there are only two XFELs facilities available and working. Due to the prospect and growing number scientific breakthroughs enabled by XFELs, many other countries are now working to build XFEL facilities. At present there are four XFELs under construction (PSI-SwissFEL in Switzerland, the European XFEL in Germany, PAL in South Korea, and LCLS-II in the USA). These facilities will become operational in the coming three or

four years. Each of these XFELs will provide some unique capabilities. Table 1.3 provides a comparison of the capabilities of existing and upcoming XFELs (circa 2015).

Table 1.3: Comparison of Various Existing and Upcoming XFEL Facilities.

	LCLS	SACLA	Eu-XFEL	PAL XFEL	PSI-XFEL
Max. E-beam energy (GeV)	4.5 – 14.5	6-8	10-17.5	10/3	2.1-5.8
Wavelength (Å)	1.3 - 15	0.8 - 15	0.5-60	0.6 - 12	1 -7 7-70
Time-structure (Hz)	120	60	27,000	60	100
Pulse length (fs)	10-100 (FWHM)	30-200 (FWHM)	2-100 (FWHM)	50-200	20-100
Bandwidth (%)	0.3	-	0.08–0.65	0.11-0.9	0.03-0.5 0.15-0.25
Detectors	CS-PAD	MPCCD	AGIPD	PAL	JUNGFRAU
Detector pixel size (μm x μm)	110x110	50x50	200x200	40x40	75x75
# of pixel/unit	184x195	512x1024	64x64	512x128	512x1024
Single photon sensitivity	0	0	0	0	0
Maximum # of photons (counts)	3500 @8-keV	2700 @6-keV	10,000 @12.4-keV	15,000 @12-keV	10,000 @12-keV
Frame-rate (Hz)	120	60	2700	60	2,000
Commissioning	2009	2011	2017	2015	2016

1.2. Time-Resolved Crystallography – a Concept

In general, X-ray crystallography is applied to the determination of structures of biomolecules, averaged over time and space (i.e., providing a static picture of the molecule). On the other hand, conformational dynamics are key to any deeper understanding of biological processes. Therefore, without the ability to access structural dynamics involved in a biological process is investigated it will not be possible to the

intricate and essential workings of nature. This need, particularly relevant for photoactive

proteins, has motivated scientists to conceptualize the 'pump-probe' experimental scheme

in order to explicitly introduce time as a 4th dimension for solving structures of the

reaction intermediates, and to follow the reaction co-ordinates as the process progresses.

In time-resolved Laue crystallographic experiments (Moffat and Ren, 1997)(Moffat,

2014), the protein molecules in the crystal are excited with a laser (i.e., the pump) to

initiate the light-driven reaction, and after a delay (to allow the conformational changes to

occur), and then the dynamic state of that crystal is probed by an X-ray pulse (i.e., probe).

The time delay between 'pump' and 'probe' is varied to sample the reaction dynamics as

a function of time.

In the following sub-sections, a brief overview, progress within the field, and the

challenges that persist in pump-probe Laue crystallography are discussed.

1.2.1. Brief History of Success in Time-Resolved Laue Crystallography

In early days, time-resolved crystallography was done in two ways – physical

trapping and chemical trapping (Messecar et al., 1997)(Bourgeois et al., 2007). In both

cases, the diffraction data were collected using monochromatic X-ray source at the

synchrotron facility. In case of physical trapping, the reaction was "trapped" by flash

freezing of the crystals and X-ray data were collected at cryogenic temperature (i.e.,

liquid helium or liquid nitrogen temp.). Thus, a reaction intermediate was trapped or

cryo-frozen, followed by usual crystallographic data collection. In 1994-95, a work by

Ilme Schlichting (Schlichting et al., 1994) on myoglobin and carboxymono dynamics

defined the pioneering leading-edge. In case of chemical trapping, the protein residues are mutated in such a way that the reaction intermediate can be made long-lived with the hope of maintaining native functionality (Bolduc et al., 1995).

The Laue pump-probe approach (described in the section 1.2.2) leads to visualize the reaction in a native system at the room temperature. In 1996-97, the study of myglobin-carboxymono dynamics (Srajer et al., 1996), photoactive yellow protein (PYP) photo-cycle studies (Genick et al., 1997)(Genick et al., 1998)(Perman et al., 1998) established the potential for time-resolved Laue crystallography with a series of landmark publications. Nanosecond time-resolution was achieved through the use of a 'tailored' undulator in the ID09B beamline at the ESRF facility in 1998 (Bourgeois et al., 2003)(Bourgeois et al., 2007). However, to achieve time-resolution in the nanosecond to picosecond time-range, required another seven years. The first X-ray structures of 150-ps and 100-ps lifetime intermediates of the PYP photo-cycle were published in 2003-04 by Schotte et al., (Schotte, 2003; Schotte et al., 2004). Currently, using a very fast pulse-rotating X-ray chopper, 100-ps time-resolution can be achieved at the ESRF (European Synchrotron and Radiation Facility) and the BioCARS at the APS (Advanced Photon Source) synchrotron facilities (Bourgeois et al., 2007).

1.2.2. Laue Crystallographic Approach for Time-Resolved Study

The first time-resolved Laue crystallography was done in 1994 using large Myoglobin crystals as model system (Schlichting et al., 1994). In Fig. 1.6, a schematic of a Laue crystallographic setup (Bourgeois et al., 2007) for time-resolved experiment is

shown. In this method, an intense polychromatic (with an energy bandwidth of $\Delta E/E =$ ~4%) synchrotron derived X-ray beam (Neutze and Moffat, 2012) is used for X-ray data collection. A single large ($\geq$ 200-µm) crystal is mounted in a capillary on a goniometer head, which allows for the data collection at room temperature. Use of a polychromatic X-ray beam minimizes the number of crystal oscillation steps, which in turn helps to make the collection more efficient, requiring fewer diffraction patterns, and therefore enabling dataset acquisition before the crystal is destroyed. In the Laue method, the mounted single crystal is excited with a visible 'pump' laser to initiate the reaction at the molecular level, followed by the 'probe' X-ray pulse (after a time delay) to collect diffraction pattern from that crystal at the given orientation. This same process of 'pump-probe' is followed for all crystal orientations, enabling full coverage of the reciprocal volume and the acquisition of a complete data set. In this method, a complete data set is collected for each time-delay.
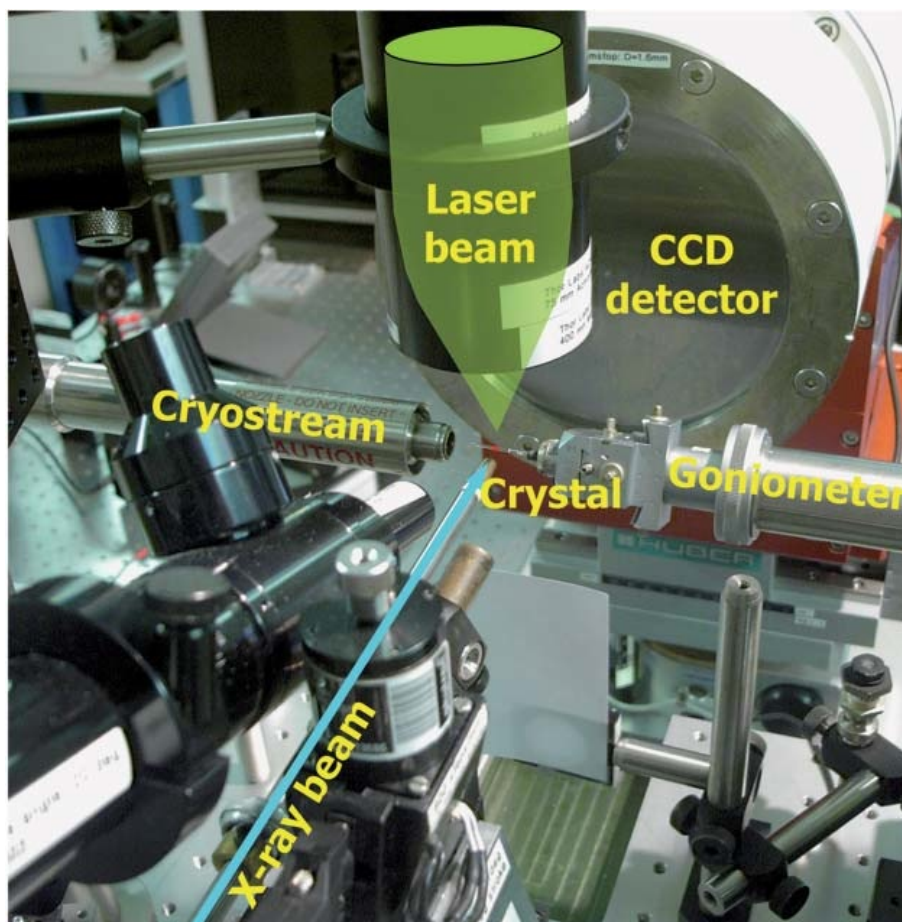
Figure 1.6: Time-Resolved Laue Experimental Set-up at ID09-B Beamline at the ESRF (Bourgeois et al., 2007).

1.2.3. Challenges of Time-Resolved Laue Crystallography

There are many practical challenges encountered when performing time-resolved Laue experiments, such as: choice of pump laser wavelength; laser power; pulse duration to minimize the pump induced heating effect; optical density of the chromophore or protein inside the crystal; and penetration depth of light or reactant within the crystal (i.e., correlated to the size of the crystals). In pump-probe experiments penetration depth is determines the extent to which a reaction progresses within crystals (Tenboer et al.,

2014). More recently, 'pink' X-ray beams are becoming more widely used in Laue experiments. A 'pink' X-ray beam is created with a narrow bandpass (~3%) that yields a 'less' polychromatic beam. With a 'pink' X-ray beam the diffraction patterns from a larger set of crystal orientations is required in order to cover the reciprocal volume completely (Bourgeois et al, 2007). In classical MX this makes the crystals under test more susceptible to X-ray induced damage. In addition to that, Laue data processing is challenging because it is difficult to accurately account for wavelength scaling (Bourgeois et al, 2000), and it is also difficult to index weaker reflections at the higher resolution (Yang et al., 1998). In addition to these practical challenges, the time-resolved Laue technique cannot be applied to study irreversible reactions. Fundamental limitations of the Laue instrumental setup step from the requirement that protein molecules in the crystal must relax back to their ground state after each excitation in order that X-ray data can be collected from both the ground and excited states in an alternating fashion. In Laue, the diffraction data alternatively collected from the ground and excited states ensures that both states are treated equally. This approach is taken to eliminate systematic errors introduced by instrumental variance, X-ray induced damage, pump-laser induced ionization, oxidation/reduction or heating effects, where the latter crystal damage artifacts hamper the reaction initiation which is the desired excited state observation (Bourgeois et al., 2007). In order to avoid these errors in a 'pump-probe' set up data have to be collected as fast as possible. These practical constraints limit the utility of time-resolved Laue experiments to reversible processes.

1.3. Data Analysis Methods

In this section and following sub-sections, the theoretical foundation of 3D protein crystallography will be briefly described. Details of approaches to analyze two-dimensional diffraction patterns to determine three-dimensional structures of molecules will follow in subsequent sub-sections. In this section, the fundamental differences between the data analysis approach used in conventional *macro-molecular crystallography* (MX) and that used in *serial femtosecond crystallography* (SFX), followed by technical details and challenges in the "new" field of SFX will be presented.

1.3.1. Theoretical Background in X-ray Crystallography

A crystal is composed of molecules in their smallest repetitive units that form an ordered lattice, called unit cells (see Section 1.1). Each unit cell can be described with three translational vectors, called cell axes, *a, b, c*. If a crystal contains n_1, n_2, and n_3 unit cells in the directions defined by the vectors *a, b, c*, then for that crystal, a unit cell origin will be at position ($R = ua + vb + wc$) with respect to global crystal origin, where u, v, w are integers, indicating nth unit cell.

X-ray scattering within crystals is based on the interaction between X-ray electromagnetic waves and the electrons of the atoms comprising the crystal. In X-ray diffraction, X-ray scattering by a crystal can be considered as superposition of many scattered waves from the elastic interactions between X-rays and electrons. Here, it is

noted that there are various types of scattering events – elastic, inelastic, Compton, Rayleigh and so on. Readers desiring a comprehensive treatment are referred to Jackson, 1999.

The scattering from a crystal depends on the number of electrons in the system and the position of the electrons (i.e., the maximum probability of the location of an electron, i.e., the "electron clouds"). Thus, the **atomic scattering factor** can be defined as,

$$f = \int \rho(r)e^{2\pi i r.q} \, dr \,,$$ ....(Eq. 1.3.1)

where:

$\rho(r)$ = electron density at r, and

q = scattering vector.

The scattering vector, q, is the difference between incident (k_i) and scattering (k_o) vectors ($q = k_i - k_o$). The amplitude of the scattering vector q is defined as,

$$|q| = \frac{2\sin\theta}{\lambda} \,,$$ .... (Eq. 1.3.2)

where: λ is the X-ray wavelength, and θ is an angle, made by k_o vector with the reflecting lattice plane.

Thereby, the scattering by a unit cell ($F^{cell}(q)$), composed of n atoms, can be described as superposition of "n" scattering events from "n" atoms.

$$F^{cell}(q) = \sum_{i=1}^{n} f_i e^{2\pi i r_i.q} \,,$$ .... (Eq. 1.3.3)

$F^{cell}(q)$ is called the **structure factor**.

Bragg's Law:

The scattering vector *q* is directed to the perpendicular direction of the crystal's reflecting lattice plane, defined by (*h, k, l*) Miller indices (See Atkins and De Paula, 2002)). If the incident wave vector k_i and scattered wave vector k_o make an angle of θ with reflecting plane (*h, k, l*); then, from 1st Laue condition, it is may be derived that,

$$\frac{a}{h} \bullet q = 1.$$

The same relation can be obtained for vectors *b* and *c*. This scalar product result implies that the projection of *a/h* onto *q* vector, multiplied by |*q*| will yield unity, under the condition of constructive interference. Using $|q| = 2\sin(\theta)/\lambda$ (Eq. 1.3.2) and $|q| = 1/d$, where *d* is the inter-planer distance between two reflecting lattice planes, it can be shown that

$$n\lambda = 2d\sin\theta \quad \text{..} \text{(Eq. 1.3.4)}$$

This is the famous 'Bragg's Law' equation, where *n* is an integer value representing the order of reflection. In order to obtain constructive interference (thereby satisfying the Bragg's Law condition), *n* must be an integer. It is noted that Bragg's Law only uses the magnitudes of the vectors (or the physically measurable quantities).

The Diffraction Condition and Ewald Sphere Construction:

In reciprocal space, each set of equidistant reflecting lattice planes is represented as one discrete reciprocal lattice point. In 1921, Paul Ewald provided the geometric representation of the Bragg's equation (see above), i.e., the diffraction condition. In reciprocal space, the lattice vectors are *a*, b*,* and *c**, defined as ("X-ray diffraction" by B. E. Warren 1990),

32

$$a^* = 2\pi \frac{b \times c}{a \bullet (b \times c)}, b^* = 2\pi \frac{c \times a}{b \bullet (c \times a)}, c^* = 2\pi \frac{a \times b}{c \bullet (a \times b)} , \dots\dots\dots\dots\dots\dots\dots\dots \text{(Eq. 1.3.5)}$$

where: a^* is directed in the plane perpendicular to b and c, i.e., parallel to a. Similarly, b^* is directed in the perpendicular plane of a and c, i.e., parallel to b and c^* is also in the parallel direction of c. The triple product, shown in the denominator of the above equation, represents the unit cell volume. Therefore, the scattering vector *S(hkl)*, can written as a function of unit vectors in reciprocal space, i.e., *a*, b*,* and *c*;* $S(hkl) = h \bullet a^* + k \bullet b^* + l \bullet c^*$. Now, considering the scattering from the crystal in all directions in three-dimensional space, a sphere can be constructed connecting the end-points of the scattering vectors (i.e., *S*(*hkl*)). This sphere is named the "Ewald sphere". If the reciprocal lattice is superimposed with the Ewald sphere, it can be shown that only those reflections which lie on the surface of the Ewald sphere will fulfill the Bragg condition. Accordingly, (Fig.1.7) not all reciprocal lattice points or *hkl* planes can lie on the surface of the Ewald sphere at the same time. Therefore, to fully sample or cover the reciprocal volume, the crystal is rotated during diffraction data collection such that each and every *hkl* planes can be brought into the diffraction condition or brought onto the surface of the Ewald sphere (P.P. Ewald, 1969)(Rupp, 2010).
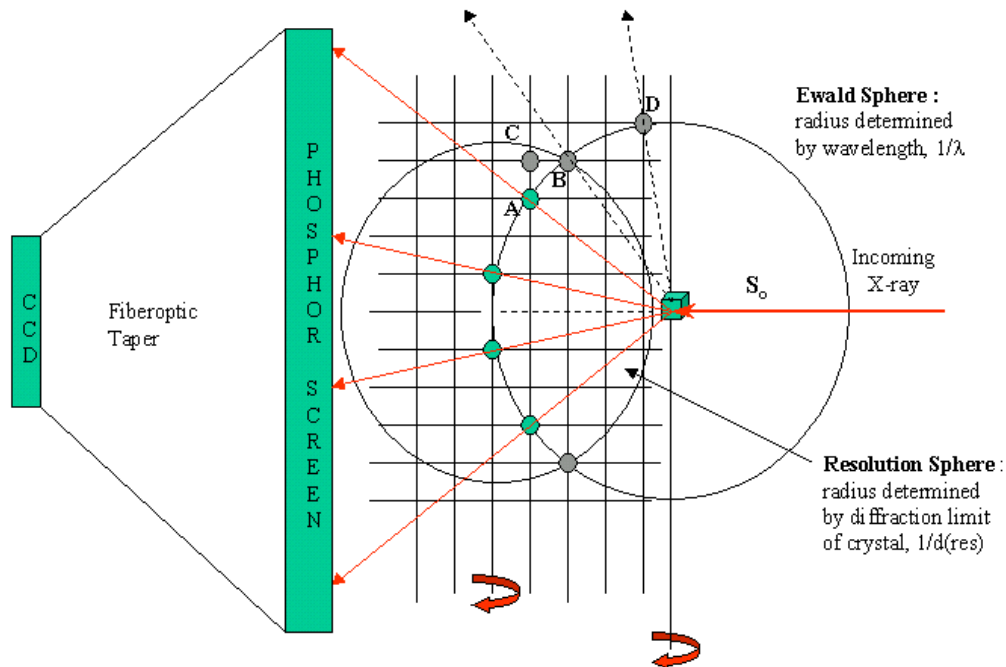
Figure 1.7: X-ray Diffraction from a Crystal and Ewald Sphere Construction. In order to record X-ray diffraction intensities from the crystal, a detector is placed in the forward scattering direction (i.e., behind the crystal). When the crystal is rotated (along the indicated axis), the reciprocal lattice points also rotate and are brought into the diffraction condition, intersecting the Ewald sphere (Source: Rupp, 2010).

Temperature B-factor and Wilson Plot:

Atoms in the crystal are not stationary in one position. Rather they vibrate around a mean position. The atoms can also be displaced in the lattice due to crystal disorder. This implies that atoms are prone to occupy slightly different position in each unit cell. This can attenuate the diffracted intensities from the atoms. Therefore, a vibrational component of the atom should be considered in atomic scattering factor definition. Because vibration is temperature dependent at higher temperature atoms are displaced to a higher degree, which is reflected in the degree of crystal disorder. This vibrational component is called **Debye-Weller temperature factor T_{iso},** (Debye, 1913) defined as,

34

$$T_{iso} = \exp[-B_{iso}(\sin\theta/\lambda)^2]. \quad\text{..} \text{(Eq. 1.3.6)}$$

This formula is based on the assumption that vibrational motion of the atom is harmonic. The Debye-Weller atomic vibration around a mean position, which is isotropic in nature, is called the **isotropic B-factor**, B_{iso}.

The thermal parameter B_{iso} is related to mean square isotropic displacement $<u^2_{iso}>$ given by, $B_{iso} = 8\pi^2 \times \langle u^2_{iso} \rangle$. It can be shown that $B_{iso} = 79\text{-Å}^2$ corresponds to a root mean square isotropic displacement of 1-Å. Nevertheless, the modified atomic scattering factor (defined earlier), is given as,

$$f^B = f^0 \exp[-B_{iso}(\sin\theta/\lambda)^2]. \quad\text{..} \text{(Eq. 1.3.7)}$$

In practice, during processing diffraction data, it is customary to scale the experimentally obtained diffracted intensities with respect to the theoretical model, or with respect to an absolute scale. This scaling requires a linear scale factor along with a correction term for the overall B-factor, accounting for the attenuation in the intensities. This isotropic B-factor scaling is called Wilson scaling (Wilson, 1942)).

The measured average intensity is directly proportional to the absolute intensity, $\langle I_{obs} \rangle \propto \langle I_{abs} \rangle$. Re-writing the "modified" atomic scattering factor for each atom with Debye-Weller term in the squared form, it can be shown that,

$$(f_i^B)^2 = (f_i^0)^2 \exp[-2B_{iso}(\sin\theta/\lambda)^2], \quad\text{...} \text{(Eq. 1.3.8)}$$

where: "i" is the index of ith atom in the crystal lattice. From the structure factor statistics in the absolute scale, it can be shown that $\langle I_{abs} \rangle = \sum_{i}^{atoms} (f_i^0)^2$. Summing the atomic

scattering factors from all atoms in the crystal lattice and replacing the scattering factors by the intensity term, resolves to become,

$$\langle I_{obs} \rangle = \langle I_{abs} \rangle k \exp[-2B_{iso}(\sin\theta / \lambda)^2] \, .$$ .... (Eq. 1.3.9)

A linearized form of this equation is obtained by taking the logarithm on both sides, yielding,

$$\ln \frac{\langle I_{obs} \rangle}{\langle I_{abs} \rangle} = \ln k - 2B_{iso}(\sin\theta / \lambda)^2 \, .$$ .... (Eq. 1.3.10)

The plot of $\ln \dfrac{\langle I_{obs} \rangle}{\langle I_{abs} \rangle}$ versus $2(\sin\theta / \lambda)^2$ is called **Wilson plot** (Wilson, 1942). The slope of this linear plot provides the estimate of B_{iso} for a particular diffraction data set that is specific to a crystal (Rupp, 2010).

Calculation of Electron Density Map:

Considering the X-ray diffraction by crystals as an elastic scattering event, then within *Kinematic limit* (referred as first order Born Approximation (Guinier, 1994)(Jackson, 1999)), it can be shown that diffracted intensity is directly proportional to the square of the structure factor (See, (Warren, 1990)). In the previous section, the structure factor term $F^{cell}(q)$ and the atomic scattering factor f have been introduced (See, Eq. 1.3.1 and 1.3.3). Using these definitions the $F^{cell}(q)$ term can be re-written as a function of electron density integrated over the unit cell, yielding,

$$F^{cell}(q) = \int_{cell} \rho(r)e^{2\pi i r \cdot q} dv \, ,$$ .... (Eq. 1.3.11)

36

where: $\rho(r)$ = electron density at the position r in the coordinate of unit cell, defined by the fractional coordinate system of x, y, z. Here, the position, r, can be expressed in terms of unit cell coordinate with respect to origin of the crystal, as $r = a \bullet x + b \bullet y + c \bullet z$. And, the volume element dv as,

$$dv = Vdxdydz \text{.} \quad\text{.. (Eq. 1.3.12)}$$

Using the expression for r as the position vector and the Laue conditions, it can be shown that,

$$r \bullet q = (a \bullet x + b \bullet y + c \bullet z) \bullet q = hx + ky + lz \text{.} \quad\text{.. (Eq. 1.3.13)}$$

Combining Eq. 1.3.12 and 1.3.13, Eq. 1.3.11, $F^{cell}(q)$ can be re-written as,

$$F^{cell}(q) = F(hkl) = V \iiint\limits_{x,y,z} \rho(xyz) e^{2\pi i(hx+ky+lz)} dxdydz \text{.} \quad\text{... (Eq. 1.3.14)}$$

Eq. 1.3.14 can be interpreted as $F(hkl)$, or the structure factor (i.e., the term in reciprocal space), which is the Fourier transform of $\rho(xyz)$, i.e., the electron density of atoms. The inverse Fourier transformation yields,

$$\rho(xyz) = \frac{1}{V} \sum_{h,k,l} F(hkl) e^{-2\pi i(hx+ky+lz)} \text{.} \quad\text{.. (Eq. 1.3.15)}$$

Here, the integration is replaced by summation because diffraction occurs in discrete directions rather than continuously in all directions. $F(hkl)$, i.e., the structure factor is a complex vector quantity, defined in reciprocal space. So, $F(hkl) = |F(hkl)| e^{i\varphi(hkl)}$ and Eq.

1.3.15 can be re-written as, $\rho(xyz) = \dfrac{1}{V} \sum\limits_{h,k,l} |F(hkl)| * e^{-2\pi i(hx+ky+lz)} e^{i\varphi(hkl)}$ . ... (Eq. 1.3.16)

In Eq. 1.3.16, the $\varphi(hkl)$ term can be interpreted as phase angle for each set of hkl planes in reciprocal space, which cannot be determined directly from the diffraction

37

patterns. This exposes the so-called "phase problem" in the field of X-ray crystallography. There are various ways to address the phase problem, such as molecular replacement, and experimental phasing methods (Details can be found in (Rupp, 2010) book).

1.3.2. Data Processing Approach in Macromolecular Crystallography

The data processing approach in macromolecular crystallography encompasses all steps from the data acquisition, indexing, merging, phasing to obtaining electron density map and model building. MX data evaluation is a very intense and mature subject. New data evaluation methods also have to be developed specifically for the SFX technique. In this thesis, I will briefly discuss these subjects. Interested readers will be directed to relevant references for in-depth treatments.

In MX, data collection and sampling of the reciprocal volume have been briefly outlined in the previous section (i.e., Section 1.3.1). Given a complete data set from crystal diffraction, data processing starts with the identification of "Bragg" peaks and auto-indexing. In MX, detectors are well calibrated to tackle electronic and shot noises. Thereby, detecting "Bragg" peaks based on diffracted intensities is straightforward. However, in SFX data processing, identifying Bragg peaks based on diffracted intensities is extremely challenging and new methodology are required. This section begins with the introduction of the indexing step.

Indexing Algorithm:

Indexing means literally to identify the location of each Bragg peak in a diffraction pattern and to then assign these peaks to a set of *hkl* indices; thereby, the indexing step determines the orientation and unit cell parameters of the crystal. There are various software packages such as MOSFLM (Steller et al., 1997), DIRAX (Duisenberg, 1992), XDS (Kabsch, 1993) etc., that are commonly used to perform the indexing step in MX. Most of these programs use a multi-slice Fourier Transform (FT) approach (i.e., multiple 1D-FTs instead of one 3D-FT), as theorized by (Rossmann and van Beek, 1999). In this method, a diffraction image contains concentric circles (or, "lunes") of spots, which satisfy the diffraction condition and lie on the surface of the Ewald sphere. Each spot is associated with a 3D scattering vector, perpendicular to the reflecting plane direction. If these scattering vectors are projected onto the principle zone-axis for a particular crystal orientation, all the spots, lying on the same "lune" (or circle) will give rise to a projected vector of same length (i.e., can be thought of a resultant vector). Therefore, projected scattering vectors from all spots will fall into different clusters. Each cluster, containing a certain number of scattering vectors from a certain numbers of spots, will have a cluster-centroid or a resultant vector, representing the axis length in one direction in reciprocal space. From these results, a histogram can be plotted, showing how many candidate spots are assigned to a particular cluster. 1D-FT of that histogram will represent the axis-length in real space (See, (Rossmann and van Beek, 1999)). Generally, the program will consider the cluster which has highest number of spots. This refers to the cluster with highest order of periodicity. Thus, the program will virtually rotate the crystal along a

39

zone axis, very similar to the experimental rotation of the crystal on a goniometer. Thus, combining the candidate axis-lengths from all possible rotations (covering the hemisphere of the Ewald sphere) in all three-directions, a list of unit cell axes is formed. The candidate axis-lengths are then sorted in descending order of periodicity. Finally, the candidate axis-length combination (i.e., combination of *a, b, c* parameters) that gives highest possible periodicity in a data set, is chosen as the unit cell parameters for that data set. Based on those axis length vectors in reciprocal space, an orientation matrix is formed, representing the crystal orientation.

$$A = \begin{pmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{pmatrix}$$ .... (Eq. 1.3.17)

Merging and Scaling of the Data Set:

The indexing step determines unit cell dimensions and crystal orientation (i.e., A matrix in Eq. 1.3.17). The Miller indices (*hkl*) are determined based on scattering vector q and A, i.e., orientation matrix, following the equation,

$$\begin{bmatrix} h^{frac} \\ k^{frac} \\ l^{frac} \end{bmatrix} = \begin{bmatrix} a_x^* & b_x^* & c_x^* \\ a_y^* & b_y^* & c_y^* \\ a_z^* & b_z^* & c_z^* \end{bmatrix}^{-1} \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}$$ .... (Eq. 1.3.18)

According to Eq. 1.3.18, initially, the Miller indices are determined as fractions, corresponding to each detector pixel. Then, the nearest reciprocal lattice vector to each detector pixel is obtained by rounding the fractional Miller indices (Kirian et al., 2010) to the nearest integer values of *hkl*.

$$g_i = Ah_i.$$ .... (Eq. 1.3.19)

It is noted that the matrices are represented here symbolically by one letter instead of writing out entire large matrices for convenience. A = orientation matrix, h_i = integer Miller indices for detector pixel i, and g_i = nearest reciprocal lattice point.

The reciprocal space distance δ_i' between q_i and the nearest reciprocal lattice point, i.e., g_i is given as, $\delta_i' = |q_i - g_i|$. .... (Eq. 1.3.20)

The merging and scaling step involves integration of intensities from each Bragg reflection over 'm' number of crystals, which can be given as,

$$I_{hkl}^{exp}(m, \delta') = \sum_{j=1}^{m} \sum_{\{i\}_{j,hkl,\delta'}} I_j^{correct}(q_i),$$ .... (Eq. 1.3.21)

$I_j^{correct}(q_i)$= Diffracted intensity over $\{i\}_{j,hkl,\delta'}$, a set of pixels in diffraction pattern j after background subtraction and correction of polarization factor.

$$I_j^{correct}(q_i) = \frac{I_j(q_i) - I_{bg}(q_i)}{P_i \Delta\Omega}.$$ .... (Eq. 1.3.22)

The experimental structure factor amplitude is obtained from integrated intensities (i.e., Eq. 1.3.21), using the following equation,

$$|F_{hkl}^{exp}|^2 = \frac{I_{hkl}^{exp}(m, \delta')}{M_{hkl}(m, \delta')},$$ .... (Eq. 1.3.23)

where: $M_{hkl}(m, \delta')$ is a scaling factor to normalize each term to a single absolute scale.

The abridged derivation above, taken together as a whole, provides the mathematical construct for the merging and scaling process. For detail derivations, readers are referred to read (Kirian et al., 2010) and (Kirian et al., 2011) articles. It is noted that during merging step, measurements of the symmetry equivalent reflections are combined. In addition, reflections that are measured partially in different diffraction snapshots (particularly in SFX) are merged to confirm the Bragg reflection (using Eq. 1.3.21 and

Section 1.3.5). During the scaling step, the intensity for each HKL reflection is put in an absolute scale. In addition, the Lorentz factor, and the polarization factor are also corrected for geometrical effects. Each diffraction pattern is different to some extent based on "effective exposure time," accumulated radiation dose, variable intensity of the X-ray beam, absorption by crystal, and so on. Therefore, every HKL reflection intensity is normalized to an absolute scale along with the temperature B-factor correction. In the case of SFX, this merging step is also performed based on the abridged derivation presented above. Interested readers are referred to Section 1.3.5 for Monte-Carlo integration, and the work by (Kirian et al., 2011) for an in-depth treatment of the subject.

Molecular Replacement:

From the diffraction of crystals, only intensities of the scattered waves can be obtained, but the phase information is lost. These intensities are equal to the square of the structure factor amplitudes. The structure factor is a complex vector defined in reciprocal space (See Eq. 1.3.16). It is related to the scattered X-ray waves, having both amplitude and phase. Structure factor amplitudes can be extracted from the diffracted intensities experimentally. Phase (i.e., *φ(hkl)* term) angle cannot be obtained directly, leading to so-called "phase problem." In order to determine *de novo* structures, of proteins/biomolecules, this phase information has to be retrieved. There are various experimental methods of determining phases, which require collection of multiple data sets. These experimental phasing methods are: Single-Anomalous-diffraction (SAD), Single-Isomorphous-Replacement (SIR), Multiple Isomorphous Replacement (MIR), and

Multi-wavelengths Anomalous diffraction (MAD). Experimental phasing is a very broad and intensively studies topic.

Molecular Replacement (MR) is used to obtain phase information without doing experimental phasing experiments. The method is based on prior knowledge from existing similar structures with 20% or greater homology in amino acid sequence. MR cannot determine per say *ab initio* structures. It can only work if some prior structural information is available. As such, Molecular Replacement (MR) is the most frequently used method to determine the phase when some prior structural information is known (such as homologous models). In this section, MR will be briefly explained. The concept of MR was first introduced by Rossmann and Blow in 1962 (Rossmann and Blow, 1962).

The MR method, in principle, is very simple. If there is a known model, which is very similar to the protein of interest, then the phase information can be extracted using that 'known' model as a seed for a search for a model that fits the experimentally obtained structure factors from the crystal of the protein of interest. Using translational and rotational motions, phase information will be modified to determine the structure factors, followed by rebuilding the model for the structure of interest. The similarity between the 'known or search' model and the unknown model can be defined as the similarity in protein (i.e., amino acids) sequences. Common protocols consider a minimum cut-off similarity identity $\geq 30\%$, or require belonging to the same class or family of proteins by having some conserved regions in the sequence or similarity in protein folding.

In order to find a MR solution, at first, structure factors are extracted from the known model, called (F_{calc}) and a Patterson function is calculated from F_{calc}. Similarly, using the

43

F_{obs} (i.e., experimentally obtained structure factor amplitudes from crystal of protein of interest), another Patterson function is calculated. The Patterson function is very useful because it estimates the position of one atom with respect to others in Patterson space.

$$P(uvw) = \frac{1}{V} \sum_{h,k,l} |F(hkl)|^2 \cos[2\pi(hu + kv + lw)] .$$ .... (Eq. 1.3.24)

Eq. 1.3.24 is the Patterson function equation, where: $u = x_1 - x_2; v = y_1 - y_2; w = z_1 - z_2$. Using u, v, w coordinates, any pair of atoms in the crystal can be shown in Patterson space. After extracting Patterson functions from both F_{calc} and F_{obs}, the Patterson function from the observed data is superimposed onto that from the known model after a rotation and translational operation in such a manner that maximum overlap is ensured. To estimate a 'score' for the overlap or fitting, a correlation metric is calculated as follows:

$$C = \frac{\sum_{hkl}\left(|F_{obs}|^2 - \langle|F_{obs}|^2\rangle\right) \bullet \left(|F_{calc}|^2 - \langle|F_{calc}|^2\rangle\right)}{\sqrt{\sum_{hkl}\left(|F_{obs}|^2 - \langle|F_{obs}|^2\rangle\right)^2}\sqrt{\sum_{hkl}\left(|F_{calc}|^2 - \langle|F_{calc}|^2\rangle\right)^2}} .$$ .... (Eq. 1.3.25)

In another approach a maximum-likelihood probability is calculated in order to obtain best orientation and translational function.

The major drawback of MR method is that it can introduce 'model bias', i.e., it can force the unknown model to be very much similar looking as the 'known or search' model. Therefore, crystallographers are extremely cautious in refining the structures from the MR solutions. They use many different control tests to cross-validate (e.g., difference electron density maps, omit maps, Simulated Annealing omit maps, etc.) or circumvent the 'model-bias' issue (some of them are used later in this thesis).

Experimental Phasing Method – Isomorphous Replacement:

As mentioned in the previous section, experimental phasing methods are widely used to determine unknown or novel structures. There are various ways of performing phasing experiments. Since, thorough discussion of experimental phasing methods is beyond the scope of this thesis, here only one method (i.e., isomorphous replacement) will be introduced and advanced interested readers are referred to the comprehensive treatments of (Rhodes, 2006) and (Rupp, 2010).

In Isomorphous Replacement (IR) experiments, at least two (single IR or SIR) or multiple (multiple IR or MIR) X-ray diffraction data sets need to be collected, in which heavy atoms are soaked into protein crystals, so that the structure factors F_{PH} become a summation of structure factors of native proteins, F_P (i.e., without heavy atoms) and additional structure factors from heavy atoms, F_H:

$$F_{PH} = F_P + F_H \quad \text{... (Eq. 1.3.26)}$$

Given such IR data sets, a Harker diagram has been shown (Fig. 1.8) to explain the solution to the phase problem. This diagram shows the relationship between native (F_P) and derivative structure factors (F_{PH}) in the complex plane. From the experimentally measured $|F_{PH}|$ and $|F_P|$ values, two circles of radii $|F_{PH}|$ and $|F_P|$ can be drawn. If we know both the magnitude and phase of F_H, we can determine the offset on both circles by vector F_H and obtain two possible phase values for F_P (magenta arrows). F_H can be determined easily by locating the position of the heavy atoms in the structure. Since single IR measurement will provide two phase values for F_P, a third data set can be obtained (i.e., multiple IR) to resolve this "phase ambiguity."
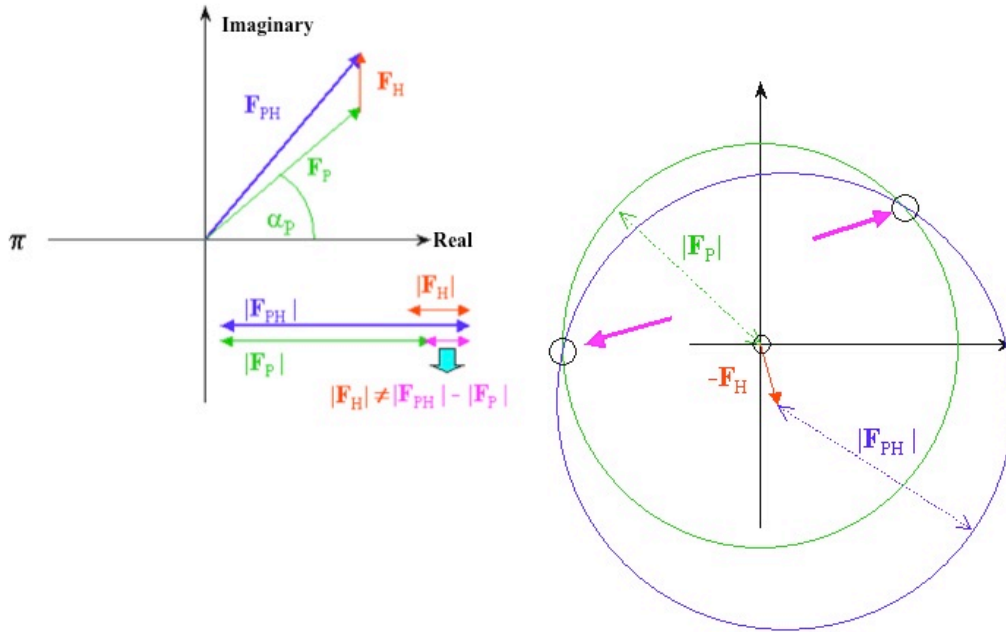
Figure 1.8: Harker Diagram for Single Isomorphous Replacement (SIR). It shows the solution to the phase problem (Rupp, 2010).

Model Refinement:

In the data processing steps of MX, model refinement is the very last step. Before refining the MR solution there are certain data quality metrics which need to be checked. Data completeness and $I/\sigma(I)$ are two very important metrics. Data completeness is defined as the percentage ratio of unique *hkl* reflections to the possible numbers of symmetry allowed reflections at a given unit cell dimension. For a reasonably good data set, completeness should be very high (100% - 98%). $I/\sigma(I)$ metric is very useful for determining the maximum resolution possible for a given data set. It is defined as the average intensity of a set of reflections at a given resolution shell (above the noise level), and is represented by the standard deviation of the intensities from the same set of reflections at that resolution shell. Typically, $I/\sigma(I) \geq 1.5$ to 3 is used as one of the criteria

46

for the highest resolution limit of a MX data set. There is another parameter, recently, introduced by (Karplus and Diederichs 2012), called $CC_{1/2}$. This is basically a Pearson correlation term (Chapter 3 methods section) for measuring the consistency of the intensities for each HKL reflection. This metric is more robust and serves better in efforts directed at 'pushing' the resolution limit to the highest end. Generally, $CC_{1/2} \geq 0.3$ at the highest resolution shell can be used as resolution limit. All of these metrics help to estimate the quality of the structure factor amplitudes. Once solved, final model and data quality can be better interpreted in terms of R-factor (particularly, R-free (Brünger, 1992)).

In the SFX technique, there are two very crucial metrics Rsplit and multiplicity (See, Section 1.3.5). Multiplicity is defined as the number of measurement of each *hkl* reflection on an average. Because the crystal is not rotated in the SFX technique not all reflections are brought onto the diffraction condition and also each crystal cuts a very thin slice of the Ewald sphere leading to 'partial' Bragg reflection instead of 'full' Bragg reflection (See section 1.3.5, for details).

During model refinement, a good agreement between R-factor and R-free must be obtained to produce a reliable structure. R-free is a type of R-factor, introduced by Brunger in 1992 (Brünger, 1992). R-free is defined as the R-factor calculated based on randomly selected (say, 5% - 10%) structure factor amplitudes as a cross-validation set. R-free would not be used during the phasing or model refinement step. A good agreement

between R-factor and R-free ensures that the 'solved' protein structure does not suffer

from the so-called 'over-refinement issue.' R-factor is defined as, $R = \dfrac{\sum_{hkl} \| F^{obs}_{hkl} | - | F^{calc}_{hkl} \|}{\sum_{hkl} | F^{obs}_{hkl} |}$.

1.3.3. Serial Femtosecond Crystallography (SFX) – Birth of 'New' Type of Data

In section 1.1.7, Serial femtosecond crystallography (SFX) was introduced and SFX

data collection was discussed in Section 1.1.7. In this section, I delve more deeply in the

subject in an attempt to get reveal new insights pertinent to the technique from the

perspective of data processing.

As shown in Fig. 1.5, SFX data collection is quite different than that in MX. In SFX

the crystals are embedded in their mother liquor and are delivered as a stream into the

interaction region with X-ray beam. Each diffraction pattern is collected from a different

crystal. These crystals are generally on the order of a micron or submicron in size.

Therefore, these crystals are anticipated to have short-range order, unlike crystals in MX.

In a stream of crystals, not every crystal is necessarily homogeneous in size and/or shape.

Further, even if crystals are homogeneous in size and shape, X-ray pulses may not hit the

same volume of the crystals in each shot. In addition, because of the nature of crystal

delivery, each diffraction 'snapshot' is not only from a different crystal, but is also

random in orientation (Kirian et al., 2010). Therefore, unlike MX, where a complete data

set is collected from one single crystal, in SFX, a data set comprises of diffraction

patterns from many tiny crystals. Since, the crystals interacts with X-ray in random

orientations, millions of diffraction patterns may need to be collected in order to ensure

that sufficient diffraction patterns have been collected to sample all possible orientations and to achieve data information content completeness. In addition to this, XFEL pulses are extremely coherent with small divergence (~1-μrad), leading to the observed detection of a very thin slice of the Ewald sphere on interaction between crystal and X-ray beam. This thin slice of the Ewald sphere cannot satisfy the Bragg condition completely, resulting in Bragg reflections with very low partiality. This is the typical situation of compromise between coherence and divergence of the beam (Kirian, 2011). However, in order to ensure each *hkl* reflection is measured completely, each reflection one has to be measured more frequently than in MX (i.e., high multiplicity of each reflection must be ensured), resulting in a large number of diffraction patterns. Moreover, due to the nature of the crystal-delivery, there are always probabilities of "not hitting a crystal", "hitting of multiple crystals" or the desired intercept of "hitting a single crystal." Considering all of these technical limitations and factors, each complete SFX dataset comprises of several thousands to millions of diffraction patterns (i.e., at least 20-50 terabytes of data). This is a completely different situation than that experienced in MX, where a maximum of ~400 diffraction images (i.e., 1 or 2 gigabytes of data) are collected. Thus, SFX generates a completely "new" type of data. In addition, the SFX method needs a detector which can function under vacuum condition and can read out data at the speed of 120-Hz, the repetition rate of the X-ray pulses. The use of a "new" type of detector adds further data acquisition and handling complexity because the data itself is "new" as well (Spence and Chapman, 2014). As a result, the standard protocols and algorithms used in MX for standard detectors could not be applied directly. Thus, the

SFX technique is not only unique in terms of the required experimental set up, but is also new in the type of data produced (Spence and Chapman, 2014). This inherent novelty revealed the need for "new" approaches, algorithms, and new analytical tools to process these large and substantially unique datasets in an efficient manner.

1.3.4. Challenges in Data Analysis

SFX is different than MX and produces a "new" type of data. Unlike MX, during SFX data processing, indexing and merging of thousands of diffraction patterns poses a demanding technical challenge which is defined by the required need to quickly analyze many terabytes of raw data, collected at the CS-PAD (Cornell-Stanford Pixel Array Detector) at the repetition rate of 120-Hz. The unique SFX challenges that must be addressed in order to make data analysis as mature as in MX are:

a. No automated 'intelligent' software exists at LCLS which can distinguish "on the fly" between diffraction snapshot from the crystal and empty snapshots.

b. SFX data is recorded on a CSPAD detector (Hart et al., 2012), which comprises of 64 panels. Each panel contains 194x185-pixels, totaling of 1752x1752-pixels, including gaps in between panels. Each pixel is 110x110-μm in size (See Fig. 1.4). CSPAD is not calibrated like detectors in used in the MX method, where the detectors are well calibrated and are not paneled with varying sizes of gaps in between. This poses multiple challenge in terms of accurately correcting the electronic noise; pixel gain correction; and the ability to mask out bad, hot, saturated pixels, which has to be done manually. Adding further complexity, scattering from the liquid injector jet is

possible, as well as injected noise due to delivery medium. Combining all these, it is clear that careful pre-processing of patterns to identify images from actual crystals and throwing away the empty and useless noisy ones (note: a process that must be very carefully done) requires a huge level of manual effort in the absence of robust automated programs.

c. The indexing step can also be challenging, as it requires a corrected detector "metrology". Since, CSPAD is multi-paneled and each panel can move and tilt independently, knowledge of accurate "metrology" of the detector is very important in terms of locating peak position and indexing them accurately. Incorrect "metrology" of the detector can lead to erroneous indexing and incorrect unit cell dimensions. That is why refining the metrology provides significant improvement in peak integration and final data quality.

d. Use of nano-/micro- crystals complicates the analysis. Due to finite truncation of the crystals, diffracted spots on the detector can be 'smeared' or produce 'radial streaks.' These streaks not only make the indexing step harder, but also complicate the peak integration process by introducing 'artificial' mosaicity.

e. After the indexing step, merging and scaling of reflections are needed (as mentioned in section 1.3.2). Each diffraction pattern contains 'partial' Bragg reflections (See section 1.3.3), unlike in MX data processing. It is noted that individual MX diffraction patterns may also contains 'partial' Bragg reflection, however a full oscillation series that is obtained during data collection ensured completeness of the Bragg reflections. Crystal shots produced by X-ray pulses with varying intensities,

and heterogeneity in diffraction volume further complicate the analysis. With varying crystal sizes the number of unit cells present in each crystal is different, resulting in different scattering powers (Holton and Frankel, 2010). With recorded reflections being 'partials,' extraction of 'full' structure factor amplitudes would benefit from 3D modeling of intensity profile around each Bragg condition. If each reflection is measured a sufficient number of times over many diffraction patterns, those 'partial' intensities for each HKL reflection can be summed to achieve "full" Bragg intensities, adopting a "Monte Carlo" approach to integrate (Kirian et al., 2011) over all varying parameters.

f. Therefore, estimating "partiality" of each measured reflection or the fraction of the 'full' Bragg reflection still poses a significant challenge in the field. Recently published works begin to examine this problem and its potential solutions (White et al., 2014)(Ginn et al., 2015)(Uervirojnangkoom et al., 2015) for details.
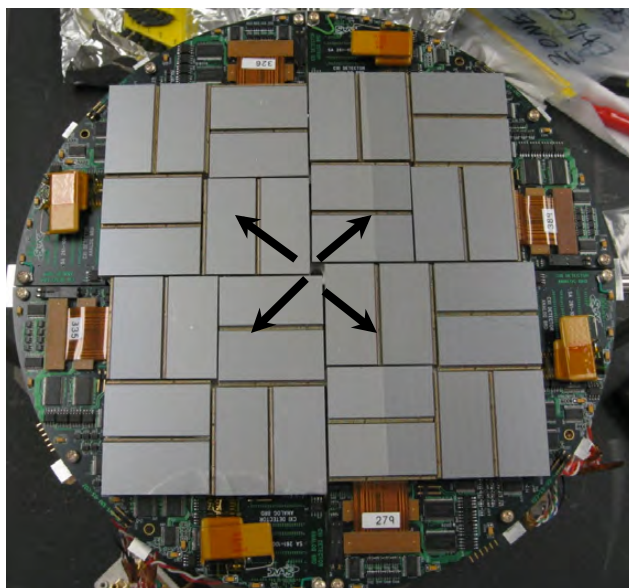
Figure 1.9: Image of a Cornell-SLAC Pixel Array Detector (CSPAD). It has 4 quadrants, each of which has 16 panels. The possible movements of the quadrants are shown in black arrows.

1.3.5. Data Processing for Time-Resolved SFX

Data processing steps for SFX data sets are in principle similar to those for MX data, (discussed in Section 1.3.2). The only difference arising between these two data types results from the way the data is handled. The data processing steps for time-resolved SFX (TR-SFX) diffraction data has been provided as a flow chart in Fig. 1.10. In principle SFX data needs to be pre-processed before the indexing step. This pre-processing step involves background noise subtraction, dark, hot, dead pixels masking, detector electronic noise (i.e., dark current) subtraction for each pattern in order to find out which patterns contain diffraction spots from crystals. For the purpose of pre-processing, the Cheetah (Barty et al., 2014) program has been developed. An image showing electronic noise level of the detector, bad, hot, and saturated pixels is provided in Fig. 1.11. This

electronic noise, bad and saturated pixels (as shown in Fig. 1.11) are subtracted from each diffraction pattern. Then, based on threshold intensity of the diffracted spots, minimum number of spots expected, and threshold signal-to-noise ratio (SNR), useful diffraction patterns containing diffraction spots are extracted. An example diffraction pattern (after pre-processing) is shown in Fig. 1.12.

In TR-SFX, after pre-processing, it is important to sort diffraction patterns based on pump laser information (i.e., ON/OFF information) into two groups – one for ground state and another for excited state (Further details are given in Chapter 3 and 5). In SFX, indexing algorithms borrowed from MX are used. Since, in SFX, data is not collected using an oscillation series, SFX programs (e.g., CrystFEL (White et al., 2012)) create a new platform for SFX data evaluation. Conceptually, the indexing algorithm in CrystFEL considers each pattern independent from the others because each pattern in SFX comes from different crystals.
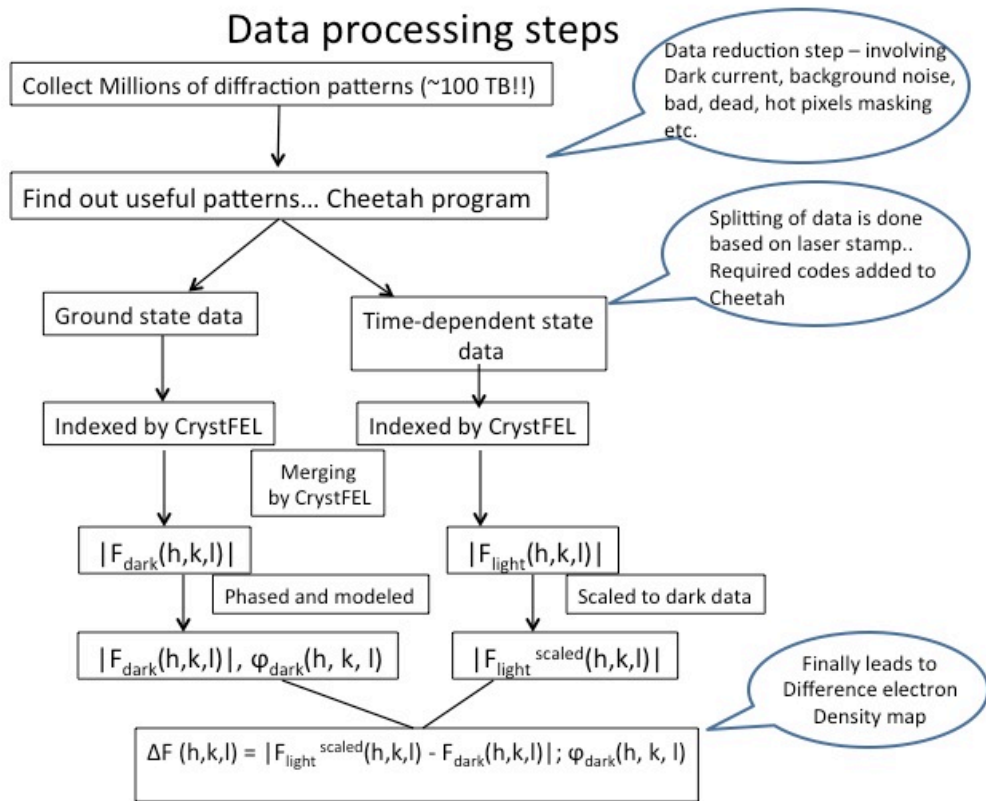
Figure 1.10: A Flow-Chart for Time-Resolved SFX Data Processing. It shows all steps involved in the analysis pipeline.
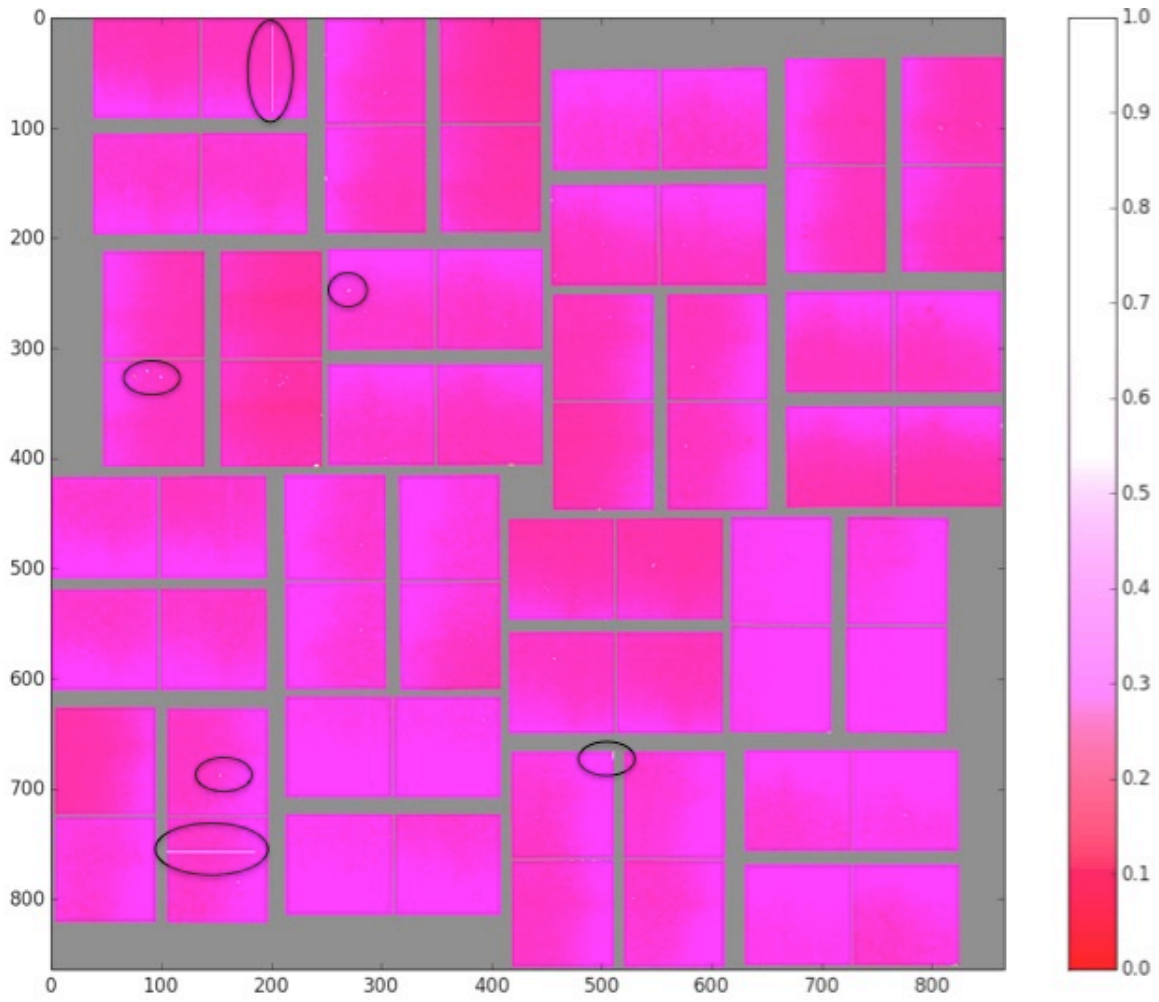
Figure 1.11: Illustration of the Detector's Electronic Noise (i.e., dark current), bad, hot, and saturated pixels are located with 'black' circles.
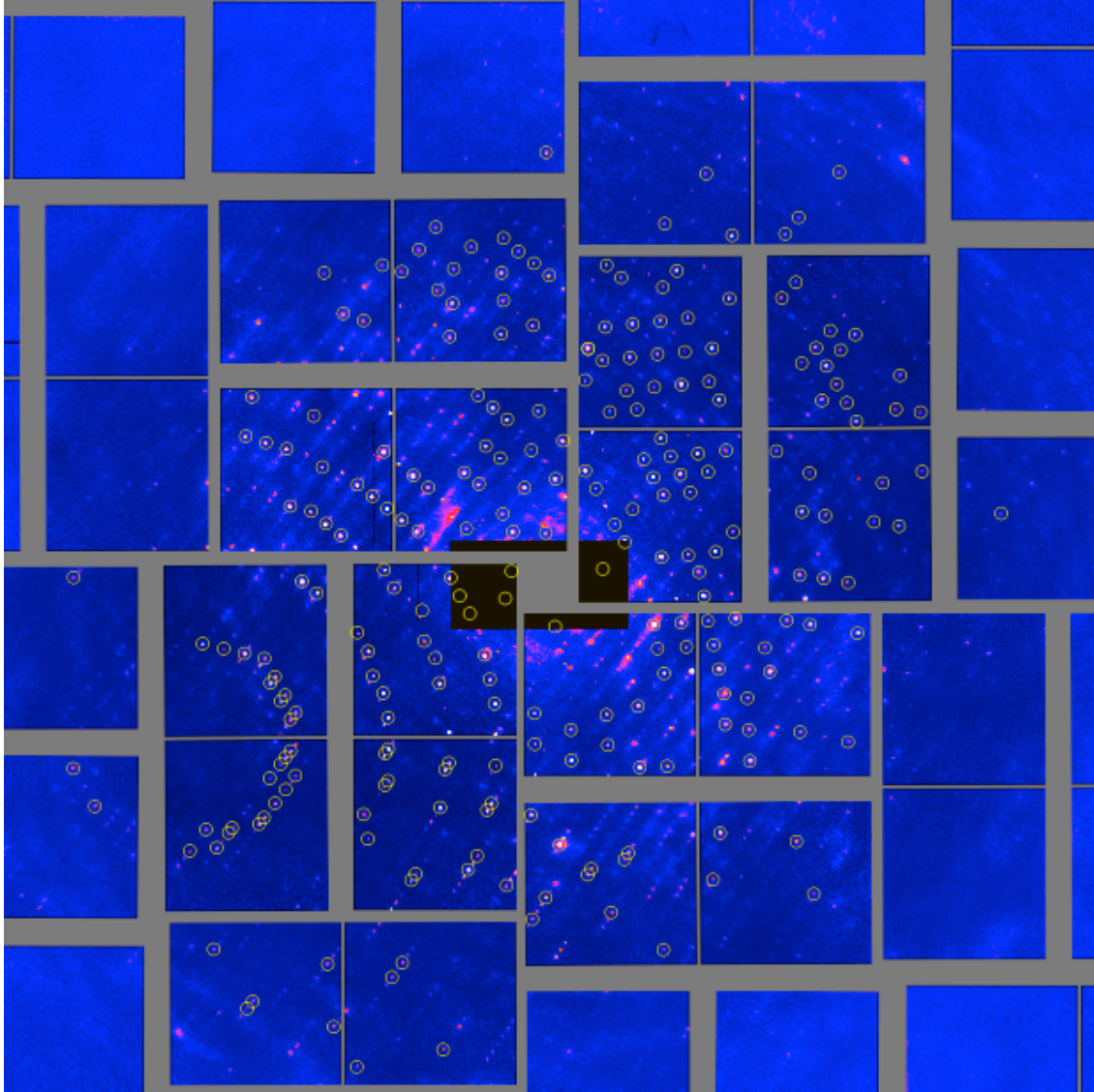
Figure 1.12: A Typical Good Diffraction Pattern from Photosystem II Data. Such patterns are extracted after hundreds of iterations of optimization of parameters during pre-processing step in Cheetah program (Barty et al., 2014).

Merging and scaling of intensities are performed in a similar fashion using "Monte Carlo integration" methods. To describe this method of intensity integration, scattered intensity $(I_n(q))$ of the scattering vector q from the nth of many randomly-oriented crystals can be written in kinematic approximation as:

$$I_n(q) = J_0 r_e^2 P |F(q)|^2 |S_n(q)|^2 \Delta\Omega, \quad \text{... (Eq. 1.3.26)}$$

where:

J_0 = Incident photon flux density (photons/pulse/area),

r_e = radius of an electron,

$F(q)$ = structure factor amplitudes of the unit cell, and

$S_n(q)$ = shape transform.

The shape transform is the fourier transform of the shape of the finite size crystal lattice corresponding to the nth nano-crystal, $\Delta\Omega$ = scattering solid angle subtended by the detector pixels, and P is the polarization factor.

In Eq. 1.3.26, for a given pixel of the detector, the scattering vector q can be determined from the crystal orientation and pixel location. $S_n(q)$, the shape or lattice transform, depends on size and shape of the nth crystal. It will obey translational symmetry, resulting in $S_n(q) = S_n(q + g_{hkl})$, where g_{hkl} is the reciprocal lattice vector at corresponding integer Miller indices of hkl. For a perfect crystal, $S_n(g_{hkl})$ is proportional to the number of unit cells in the nth crystal and the corresponding $I_n(g_{hkl})$ is proportional to the square of the number of unit cells of the nth crystal (Kirian et al., 2010).

In SFX, with the availability of several thousands of indexed diffraction patterns (i.e., containing crystal orientation information), the diffracted intensities can be merged into 3D reciprocal space intensity-map (averaging over size and shape of the crystals that differ over shot to shot).

$$\langle I_n(q) \rangle_n = J_0 r_e^2 P |F(q)|^2 \langle |S_n(q)|^2 \rangle_n \Delta\Omega. \quad \text{.. (Eq. 1.3.27)}$$

For integration of diffracted intensities, it is required that the average of the intensities fall in a small domain of radius $\delta^{'}$ (i.e., $|q - g_{hkl}|$, See Eq. 1.3.19 and 1.3.20) because the probability of obtaining diffracted intensity precisely at the reciprocal lattice point g_{hkl} is zero. To extract structure factors, the integration domain of radius $\delta^{'}$ is chosen such that all intensities satisfying the criteria $|q - g_{hkl}| < \delta^{'}$, is considered the average. Considering the effect of $\delta^{'}$, the Eq. 1.3.27 can be re-written as,

$$\langle I_n(q) \rangle_{n,hkl,\delta'} = J_0 r_e^2 P |F(q)|^2 \langle |S_n(q)|^2 \rangle_{n,hkl,\delta'} \Delta\Omega. \quad\text{................................... (Eq. 1.3.28)}$$

Assuming a narrow distribution of sizes and shapes of the crystals, an accurate mean value of $\langle |S_n(q)|^2 \rangle_{n,hkl,\delta'}$ can be obtained. Moreover, since the shape transform is identical when translated by a reciprocal lattice vector $ghkl$, $\langle |S_n(q)|^2 \rangle_{n,hkl,\delta'}$ will become constant. Then, the structure factor of unit cells can be expressed without the crystal size and shape information (because for data processing, only relative structure factors are needed) as,

$$|F_{hkl}(q)|^2 \propto \frac{\langle I_n(q) \rangle_{n,hkl,\delta'}}{J_0 r_e^2 P \Delta\Omega}. \quad\text{.. (Eq. 1.3.29)}$$

It is important to determine optimum $\delta^{'}$ value, which depends on beam divergence, spectral width, crystal disorder, mosicity, and 'partiality' factor of the Bragg reflection. That is why it is necessary to measure each HKL reflection as many times as possible (i.e., high multiplicity). The effect of $\delta^{'}$, and the quality of merged intensities after indexing are evaluated using the R-factor, expressed as,

$$R(m, \delta) = \frac{\sum_{hkl} \left| |F_{hkl}^{calc}| - \eta |F_{hkl}^{calc}| \right|}{\sum_{hkl} |F_{hkl}^{calc}|}. \quad\text{... (Eq. 1.3.30)}$$

$R(m, \delta)$= R-factor over '*m*' number of crystals or diffraction patterns and δ values. A plot (Fig. 1.13) of such R-factor against number of diffraction patterns from a simulated data set is given after (Kirian et al., 2011). For further details, interested readers are referred to (Kirian et al., 2010)(Kirian et al., 2011).
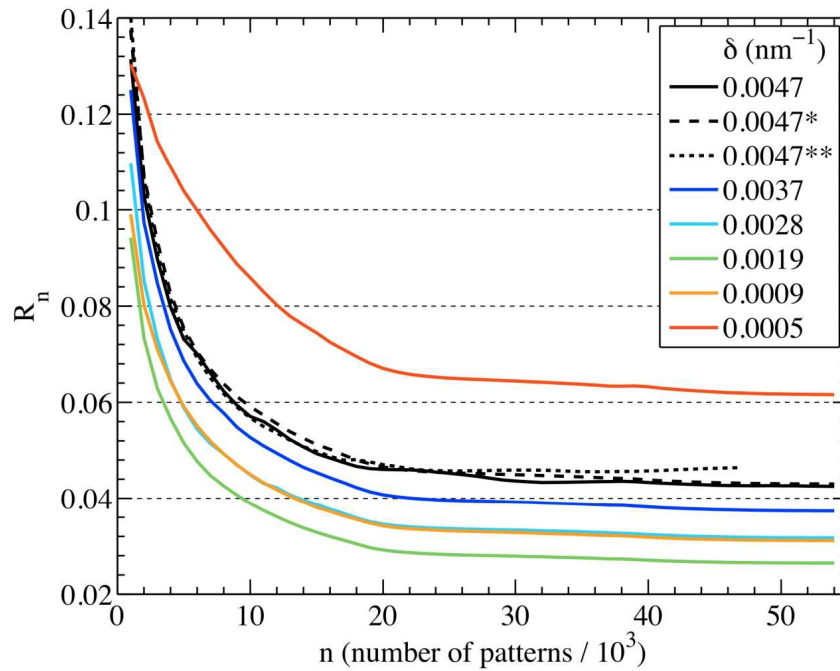
Figure 1.13: Crystallographic R-factor Plotted against Number of Crystallites in Random Orientations. Colors indicate the threshold value d. Solid lines represent simulations including Gaussian size distribution, photon noise, and water background. Dashed lines are for comparison to simulations without photon noise (*) and without noise or size distribution (**) (Source: Kirian et al., 2010).

Thus, the Monte-Carlo integration approach is used to obtain a final reflection list containing HKL reflections and corresponding Bragg intensities. Then, regular steps such as molecular replacement and refinements are carried out. The fact that SFX data processing tools under development are not fully automated poses significant practical and technical challenges (discussed in previous Section 1.3.4) making the processing of

datasets tedious and compelling users to engage in tedious manual optimization of the data evaluation steps.

When considering TR-SFX data, it is important to understand that structure factor amplitude is a 4-dimensional variable. Therefore, the structure factor is a function of space (i.e., h, k, l) and time (i.e., $F(h,k,l,t)$). In case of TR-SFX, the general aim is to see the structural differences between an excited state and ground state of the molecule. Therefore, it is important to determine the difference between two structure factors (See flow-chart in Fig. 1.10) – excited and ground states are required in order to calculate the real space difference Fourier map. The accurate calculation and interpretation of the difference Fourier map is non-trivial, mostly suffering from low signal-to-noise value due to error in difference structure factor amplitudes in the ground state structure phase, and difference Fourier truncation error (Henderson and Moffat, 1971). Moreover, no excitation process is 100% efficient, resulting in a mixture of various excited states and ground state. This means that the extracted structure factor amplitudes from the excited state diffraction patterns (i.e., $F(h,k,l,t)$) are often heterogeneous mixtures of various states. It is important to deconvolute this mixture and extract $F(h,k,l,t)$, which is pure representation of each excited state. In a classic work by Schmidt (Schmidt et al., 2003) for Laue time-resolved crystallographic data a mathematical approach was demonstrated, called **SVD (Singluar Value Decomposition)** to decouple the mixture of $F(h,k,l,t)$ and to obtain structure factors of time-independent intermediates. This was originally developed for the time-resolved Laue crystallography but can easily be transferred into time-resolved SFX (TR-SFX).

61

SVD is a technique in the field of principal component analysis (PCA). It takes a data matrix and performs matrix-diagonalization to result in separation of variables into two independent matrices. When time-resolved data, *F(h,k,l,t)* is obtained as a data-matrix, *A* of the size of *M*x*N* where each column represents *F(h,k,l)* from each time-point or excited state. After SVD operation, the *A* matrix will result in the (*M*x*N*) matrix *U* and the (*N*x*N*) *V* (See Eq. 1.3.31). The *U* matrix will contain time-independent structure factors (i.e., *F(h,k,l)*), namely LSVs (i.e., left singular vectors). *V* (i.e., RSVs) will contain time-dependent variations of the corresponding LSV term from *U* matrix. The (*N*x*N*) *S* matrix is the diagonal matrix, containing singular values or 'weighting' factors, by which *U* and V^T are correlated.

$$A \Rightarrow [MXN] \rightarrow (SVD) \rightarrow U \bullet S \bullet V^T .$$ .... (Eq. 1.3.31)

Because this approach draws upon the PCA method, Eq. 1.3.20 shows a 'reduced' representation, resulting in only principal terms and throwing away "noisy" terms (i.e., terms, having negligible values). The difference structure factors will be obtained as (Schmidt et al., 2003),

$$\Delta F(hkl,t) = |F(h,k,l,t)| - |F^D(h,k,l,t=0)|; \varphi^D(h,k,l) .$$ .... (Eq. 1.3.32)

where: $|F^D(h,k,l,t=0)|$ represents the structure factor amplitudes in the dark or ground state and $\varphi^D(h,k,l)$ term is the accurate phase information from the dark or ground state. Using $\Delta F(hkl,t)$ from Eq. 1.3.32, the difference electron density map, i.e., *Δρ(x,y,z,t)* can be calculated. Thus, the LSVs determine $\Delta \rho^{calc}$, the time-independent difference electron density. Whereas the RSVs determine $\Delta \rho^{svd}(t)$, the time-dependent difference electron density.

In order to determine kinetics or mechanisms of a reaction using time-resolved crystallographic data, $F(h,k,l,t)$ should be expressed as a linear combination of each excited state (i.e., taking account of the full population set in each excited state) and the corresponding time-independent $F(h,k,l)$ (obtained from LSVs after SVD) for that particular excited state (See Eq. 1.3.33; Schmidt et al., 2002). Therefore,

$$F(h,k,l,t) = \left(\sum_{j=1}^{J} C_{lj}(t) \bullet F_{lj}(h,k,l) \right) + \left(1 - \sum_{j=1}^{J} C_{lj}(t) \right) \bullet F_{lj=0}(h,k,l) \; \dots\dots\dots\dots\dots\dots \text{(Eq.1.3.33)}$$

where: $F_{lj=0}(h,k,l) =$ structure factor from the dark or ground state, and $C_{lj}(t)$ concentration term that is not only dependent on time but also depends on the reaction rate constant, k, which is a characteristic property of each of each proposed reaction mechanism. Thereby, a least square fitting with respect to k, can determine the correct reaction mechanism based on structure factors.

Thus, time-resolved crystallography (including TR-SFX) can be applied to the study of reaction dynamics through structural insights in a unique fashion. Since, SFX uses ultra-fast X-ray FEL pulses, it offers unique opportunities to reach very high temporal resolution and study ultra-fast processes.

1.4. Photosynthesis: Conversion of Light to Chemical Energy

The continuing existence of life on earth heavily depends on the photosynthesis, performed by plants, algae, and cyanobacteria. This is the process, in which Carbon dioxide (CO_2) and water (H_2O) are converted to sugar and oxygen (O_2) driven by sunlight. Thus, the absorbed light energy is converted to chemical energy. Photosynthetic

organisms can be divided into two groups – oxygenic (performed by plants, algae, and cyanobacteria) and anoxygenic (performed by green sulfur bacteria, green non-sulfur bacteria, and heliobacteria). While the principles of light conversion are similar, oxygenic photosynthesis provided all oxygen in the atmosphere by using water as substrate for electron-transport chain. This dissertation will focus on the oxygenic photosynthesis. In the following sections, oxygenic photosynthesis and its important components will be discussed.

1.4.1. Overview of Photosynthesis

In higher plants and eukaryotic algae, photosynthesis takes place in the cell organelle, called chloroplast. This organelle contains a complex membrane system, the thylakoids. The first part of the process i.e., photosynthetic electron transport and associated ATP synthesis take place at the thylakoid membrane. Pigments such as chlorophylls absorb photons from the sunlight and transfer their excitation energy to the Photosystem I and II, where the reaction center pigments P680 and P700, chlorophylls in special environment of pigment/protein complexes, perform the charge separation event. The photosynthetic process is driven by two major membrane protein complexes – Photosystem I (PSI) and II (PSII), which act in series to convert light energy to the chemical energy. P680 is the reaction center pigment complex of PSII and P700 is the reaction center pigment complex of PSI. Both PSI and PSII contain a core antenna complex and additional peripheral antenna system, which differ among species.

In oxygenic photosynthesis, the light energy being absorbed by antenna complex is first transferred as excitation energy to the P680 pigment, located at the center of the PSII, membrane protein complex. Receiving the excitation energy, P680, i.e., chlorophyll-*a* complex in PSII, undergoes a charge separation and forms the primary donor, $P680^+$. After charge separation, $P680^+$ is reduced by the electrons, extracted from water molecule, bound to the metal cluster, forming Oxygen Evolving Complex (OEC) in PSII. The OEC cycles through five steps of the Kok cycle (See Chapter 3; Fig. 3.1a) - S_0 to S_4, involving four charge separation events to oxidize the water molecules to oxygen. In each charge separation event, $P680^+$ is formed and reduced by one electron, extracted from water molecule via Tyr_z and metal cluster. Thus, metal cluster being the catalyst of water oxidation process also undergoes changes in oxidation states during the Kok cycle. As a result of four charge separation events, two water molecules convert to one O_2, four protons, and four electrons. In each charge separation event, the electron ejected from P680 traverses through the electron transport chain and arrives at the acceptor side, i.e., plastoquinone (PQ_B), which is a mobile charge carrier. After receiving two electrons and being reduced twice, PQ^{2-} will bind to two protons, producing plastoquinol (PQH_2), which is then released into thylakoid membrane. This process is continually repeated with PQH_2, being replaced with PQ in the thylakoid membrane (Renger, 2012)(Loll et al., 2005).

PQH_2 diffuses within the membrane and docks onto the cytochrome b6f complex, releasing two protons into the interior of the thylakoid (i.e., lumen). PQ^{2-} releases two electrons to the b6f complex, which pumps additional protons through the membrane.

The electrons will then reduce plastocyanin (PC) or cytochrome c6, contributing to the formation of the electrochemical gradient. Then, PC or cytochrome c6 will transfer the electron to $P700^+$, in the reaction center of the PSI. PSI uses the photon from the sunlight for the second charge separation event that transfers electrons from the inside of the thylakoid membrane to the outside of the membrane, where electrons will reduce ferredoxin. Once, ferredoxin receives the electrons, it undocks from PSI and binds to ferredoxin-$NADP^+$-reductase (FNR) (Sétif and Bottin 1995), which then reduces $NADP^+$ to NADPH. Fig 1.14 has shown the entire electron transport process along with protein structures embedded into thylakoid membrane schematically.
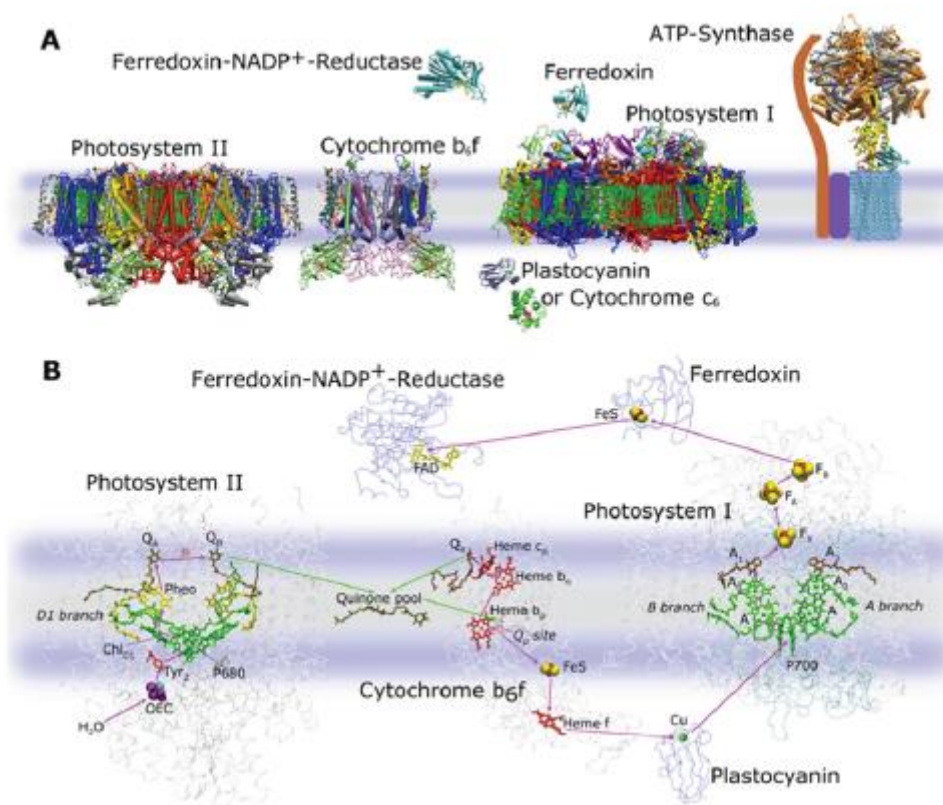
Figure 1.14: Overview of Photosynthesis - the Electron Transport Process. The electrochemical gradient generated during this entire process is used to produce ATP from ADP and inorganic phosphate by ATP-synthase, another membrane protein complex. However, amongst all the membrane protein complexes mentioned above in the photosynthetic overview, this dissertation will only focus on Photosystem II (PSII). (Source: (Fromme, 2008)).

1.4.2. Structure and Function of Photosystem I

The photosystem I (PSI) large membrane protein complex undergoes a complex light driven reaction involving multiple charge separation events. During the light driven reaction electrons are pumped from the inner part of the membrane (lumen) to the outer part of the membrane (stroma). These electrons eventually reduce $NADP^+$ to NADPH. PSI can exists in multiple oligomeric states. In higher green plants, it exists as a

monomer, bound to four light harvesting complexes (LHC). On the other hand, in cyanobacteria, PSI exists as trimer (Jordan et al., 2001)(Fromme and Grotjohann, 2008). PSI from cyanobacteria is the largest membrane protein complex, being first crystallized in 1998 (Fromme and Witt, 1998). From those crystals, first X-ray structure of PSI was determined in 2001 at 2.5-Å (Jordan et al., 2001).

Each monomer of PSI trimer structure contains 12 protein subunits and 127 non-covalently bonded cofactors or ligands. This cofactors composition comprises of 96 chlorophylls, 22 carotenoids, 3 Fe_4-S_4 clusters, 3 lipids, 2 phylloquinones, and 1 Ca^{2+} ion (Jordan et al., 2001). The structural model of PSI trimer from 2.5-Å structure is shown in Fig. 1.15. Each monomer in the trimer structure is composed of PsaA and PsaB subunits, which are related via a pseudo-C_2-axis (Jordan et al., 2001) and 10 non-symmetry related proteins. All three monomers are related to each other via the C_3 axis.
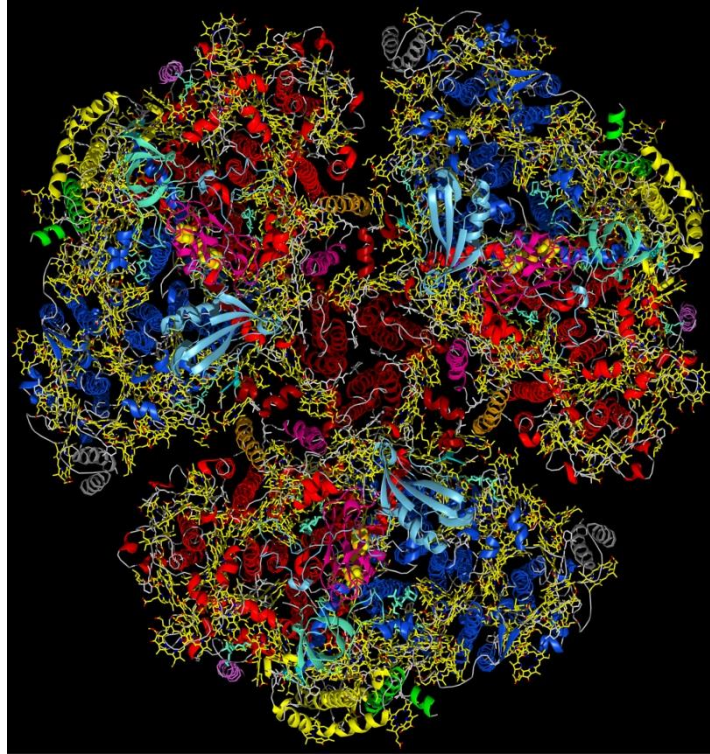
Figure 1.15: Structural Model of PSI Trimer from the 2.5-Å X-ray Structure (Jordan et al., 2001). It shows a view, perpendicular to the thylakoid membrane.

The core of the PSI contains 7 protein chains – PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM, and PsaX. Amongst these subunits, PsaI, PsaL, and PsaM are placed in the trimer contact region. PsaC, PsaD, and PsaE subunits are placed at the region, which extends into the stromal side. PsaC coordinates with two terminal electron acceptors, i.e., iron-sulfur clusters – F_A and F_B. PsaD is required for assembling and both – PsaD and PsaE are involved in the binding site of ferrodoxin (Fischer et al., 1998).

The PSI electron transport chain (ETC) is composed of 6 chlorophylls, 2 phylloquinones, and three iron-sulfur clusters (Fe_4-S_4). The P700 special pair of chlorophylls absorbs the photon from the sunlight and undergoes charge separation, forming $P700^{+\bullet}$. The released electron travels through the ETC via phylloquinone to the

iron-sulfur clusters (See Fig. 1.16). The electron goes through the pathway of F_X and F_A to the terminal cluster, F_B, from where ferrodoxin, docked onto the PSI molecule, takes up the electron. Ferredoxin, then, releases the electron to FNR enzyme, which catalyzes the reduction of $NADP^+$ to NADPH (Fromme et al., 2003).

Although scientists have a basic understanding of many of these electron transport chains, much remains unclear with regard to the intricate details and focused structural and functional (or kinetics) studies are needed to obtain insights into how the ETC functions, and in addition, the tools developed will assist in the elucidation of many hidden molecular mechanism in a vast array of different organisms.
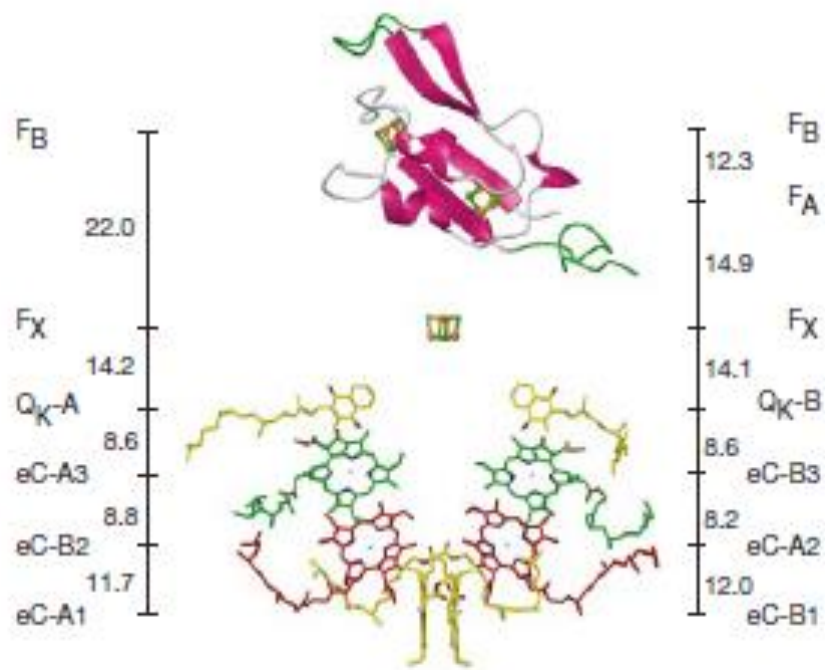
Figure 1.16: Electron Transport Chain of PSI Monomer. It shows a view parallel to the membrane. It shows two subunits – PsaA and PsaB. The label, eC, denotes the chlorophyll, and QK denotes phylloquinone. Image is taken from (Jordan et al., 2001) with modification.

1.4.3. Structure and Function of Photosystem II

70

Photosystem II (PSII) is the only membrane protein complex, embedded into thylakoid membrane, capable of catalyzing the oxidation of water to oxygen driven by sunlight. This part of the photosynthesis process is solely responsible for maintaining the oxygenic atmosphere on Earth. In *Thermosynechococcus elongates,* it contains 19 protein subunits, roughly 50 cofactors, including 36 chlorophylls. In crystallized form and in the membrane, it exists as a homo-dimer. The first X-ray structure of PSII was determined in 2001 at 3.8-Å (Zouni et al., 2001). It took almost 10 years for the scientists to obtain a X-ray structure at near atomic resolution, i.e., 1.9-Å (Umena et al., 2011). Fig. 1.17 shows an overview of PSII structural model from 1.9-Å structure ((Umena et al., 2011); PDB entry 3ARC). Four subunits from each monomer of PSII form the core of the reaction center and antenna system. They are denoted as PsbA (D1), PsbD (D2), PsbC (CP43), and PsbB (CP47). The D1 and CP43 subunits are very important for the oxygen evolution, because they provide the relevant ligands, coordinating with the metal cluster of the oxygen-evolving complex (OEC).
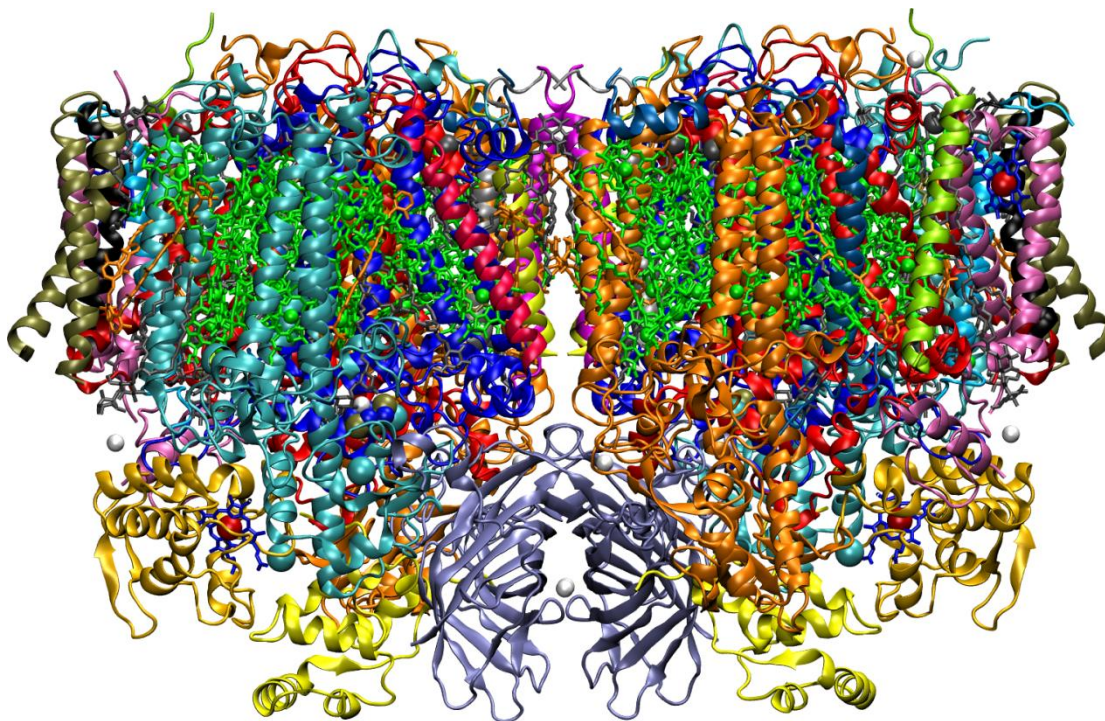
Figure 1.17: Overall Structure of PSII Dimer. The protein subunits are colored individually based on their gene code. Image is taken from the 3ARC structural model (Umena et al., 2011).

One PSII monomer contains 36 trans-membrane (TM) helices, out of which 10 TMs belong to D1 and D2 subunits, and 12 TMs belong to CP43 and CP47 subunits. The membrane loop regions of D1, D2, CP43, and CP47 host the OEC (Umena et al., 2011).
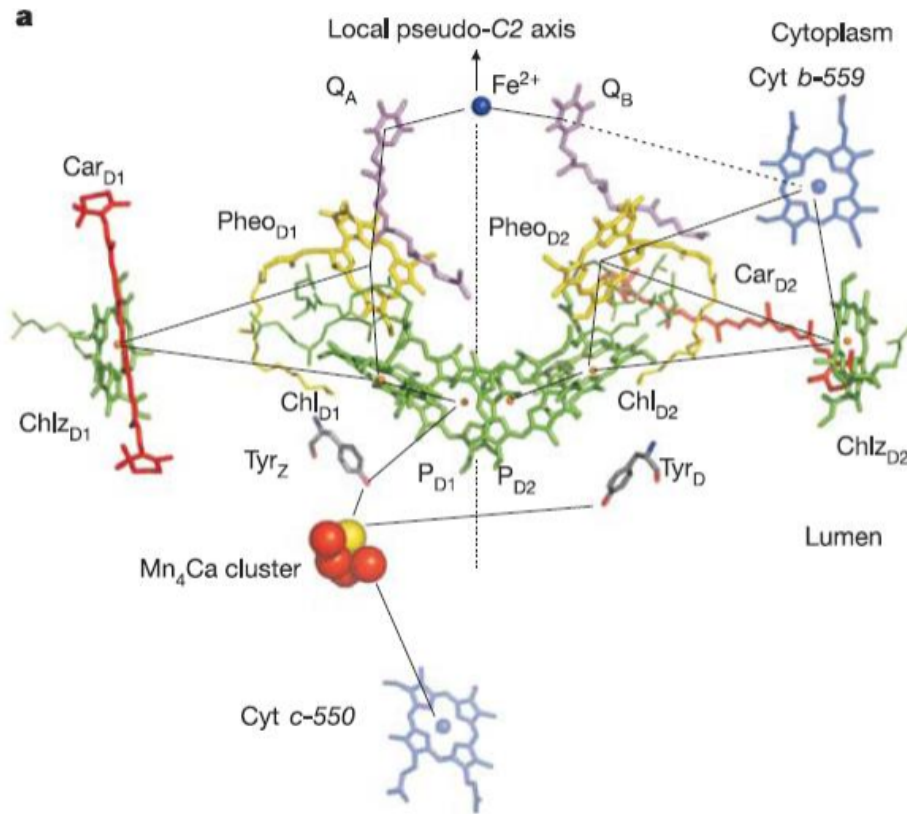
Figure 1.18: Electron Transport Chain in PSII. This view is parallel to the membrane. The image is taken with modifications from (Loll et al., 2005).

The electron transport chain (ETC) of PSII contains 4 chlorophylls-a, 2 pheophytins, 2 plastoquinones, and one non-Heme Fe atom. Non-heme Fe atom is structurally important to coordinate between PQ_A and PQ_B but not involved in ETC. Only one branch of this chain (i.e., D1 chain) is actually active for the electron transport process (See, Fig. 1.18). After absorbing photon from the sunlight, P680, the special chlorophyll-a complex, located at the reaction center (i.e., D1 branch) undergoes charge separation event, producing $P680^+$ and release the electron. The ejected electron reaches pheophytin of chain D1 in 1-5 ps (very fast) (Holzwarth et al., 2006). The electron further travels from pheophytin to immobile plastoquinone (PQ_A), in 200-ps (Holzwarth et al., 2006), forming

73

PQ_A^-. PQ_A^- transfers the electron to PQ_B, mobile plastoquinone, forming semiquinone (i.e., PQ_B^-) via non-Heme Fe atom. Once two electrons are accumulated (i.e., after two charge separation events) at PQ_B, PQ_B^{2-} is formed. Then, PQ_B^{2-} leaves the binding pocket and is released to the membrane as PQH_2. Non-Heme Fe atom holds the structural integrity and coordinates between two plastoquinones. This step is the slowest step (~200 to 400-μs) in the entire electron transfer process (de Wijn and van Gorkom, 2001). On the other hand, $P680^+$, outcome of charge separation event, is extremely oxidizing ($E_{P680+/P680} > 1.1$ eV). Thereby, $P680^+$ needs to be reduced immediately. It is reduced by Tyr_z which then extracts electron from the water molecule via OEC or the Mn_4CaO_5 metal cluster, forming the electron donor side of the process. The electron extraction from the substrate water molecule to the metal cluster takes 70-μs (Dekker et al., 1984). The extracted electron finally reduce P680+ to P680 via Tyr_z in 200-ns (Renger and Schlodder, 2011), resulting in $Tyr_z^{+\cdot}$, which extracts electron from the OEC.

To oxidize water to oxygen, 4 electrons need to be extracted from 2 molecules of substrate water. Thereby, four cycles of charge separation events need to take place. This becomes only possible because Mn atoms in the metal cluster (Mn_4CaO_5) of the OEC can store multiple charges and afford higher oxidation states, following the Kok cycle (Renger et al., 2012).

1.4.4. Photosystem II and Motivation for Time-Resolved SFX

Serial femtosecond crystallography (SFX) can take the understanding of the molecular mechanism, involved in photosynthesis to a new level. SFX and time-resolved SFX (TR-SFX) together can help to obtain structural insights into PSII membrane protein, while it is active in evolving oxygen. Since, PSII undergoes four sequential complicated charge-separation events in order to produce oxygen, TR-SFX with a capability of very high temporal resolution along with undamaged structures can reveal the 'subtle' structural changes, involved during the process. Thus, putting all those sequential structural changes together, a molecular movie can be produced to elucidate the secret mechanism of water oxidizing process, performed only by plants, algae, and cyanobacteria. Interestingly, water oxidation produces oxygen as a "side product". The aim of the OEC reaction is to provide electrons to the PSI, which reduces ferredoxin, used for the reduction of $NADP^+$ to NADPH, thereby reducing proton to hydrogen. Therefore, the knowledge of how photosystem splits water molecule along with mechanistic understanding will accelerate the research for alternative source energy. In addition to that, TR-SFX can also further investigate the dynamical process involved in very complicated electron transfer process. Thus, all these important and fundamental queries, exposing new discoveries related to the basis of life on earth, will motivate and lead to technical developments enabling the success of time-resolved SFX experiments.

2. MOTIVATION

With the birth of the new field of Serial Femtosecond Crystallography (SFX) in the research field of structural biology, it has become necessary to developing new and every increasingly efficient strategies for the execution of experiments, the analysis of massive datasets, and the interpretation of results. Indeed, the need for tools to aid in the interpretation of data and extract of structural information and functional correlations in most efficient manner has become extremely urgent. As described in Section 1.3 of the Introduction, SFX experiments produce a "new type" of data using a unique type of detector. The understanding of this "new type" of data and unique ways of collecting X-ray data, which is entirely different from conventional protein crystallography, poses new challenges.

The major challenge/motivation of this thesis was to discover and develop new methods, algorithms, and approaches to look at large datasets (~100 terabytes of data per SFX experiment). In this research, new methods for data reduction, processing, and analysis to efficiently determine structures of the large complex biomolecules were developed.

This thesis also comprises of the development of time-resolved SFX (TR-SFX). TR-SFX at provides the unique opportunity to obtain very high time and spatial resolution information, offering the expansion of analytical capabilities to the study of ultra-fast biomolecular dynamics. New capabilities often pose new challenges in the early phases of development. The challenges of establishing TR-SFX technique were engaged with proof-of-principle experimental strategies coupled with method development. This

research enabled analysis and interpretation of TR-SFX data, both major aspects of this thesis.

With the development of TR-SFX, it was of the utmost importance to apply the TR-SFX technique to a very fundamental and challenging biological question. PSII, one of the most important membrane protein complexes, facilitates the most fundamental process on earth, i.e., the water splitting reaction that produces oxygen and hydrogen. PSII is in all photosynthetic organisms, providing oxygen-evolving capacity that is critical to all high life forms on Earth. Of equal future importance to potential sustainability, the hydrogen produced may be used as renewable alternate source of fuel.

The potential is clear, however the intricacies of nature's way of splitting water to produce oxygen and hydrogen have not yet been unraveled experimentally. Unlocking the mystery of water splitting is a dream of humankind. The traditional time-resolved Laue method used in Macro-molecular crystallography is incapable of studying such reaction dynamics because of sensitivity and susceptibility of the metal cluster towards irreversible radiation damage. Because the metal cluster undergoes non-recoverable damage, it is not possible to obtain any structural insights that revealing the secretive mechanism of the water splitting reaction.

The scientific breakthroughs leading to the TR-SFX method have now, for the first time, provided analytical tools capable of producing undamaged structural information along with ability to study irreversible reaction dynamics. TR-SFX is the perfect technique for the most challenging of biological questions, including the water splitting process performed by PSII. In order to achieve reveal the mechanism of water splitting

by PSII it is necessary to conceive and implement significant contributions to the field of XFEL science that prepare the way for making 'molecular movies' of protein complexes in action. This dissertation is motivated by this grand challenge, building the platform for elucidation of PSII function, and advancing fundamental tools for structural biology research.

3. TIME-RESOLVED CRYSTALLOGRAPHY OF PHOTOSYSTEM II USING A FEMTOSECOND X-RAY LASER

This chapter will describe the time-resolved experiments performed on Photosystem II micro-crystals to investigate the 2nd excited state, attained during water splitting, at the LCLS facility using Serial Femtosecond X-ray crystallography (SFX) method. This time-resolved work demonstrates the first successful time-resolved SFX experiment, exhibiting significant conformational changes between ground and the excited states at a moderate resolution. This work was first published in *Nature* journal in July 2014 (Kupitz and Basu et al., 2014).

3.1. Overview

Photosynthesis converts sunlight into the chemical energy needed to sustain life on Earth. Two large membrane protein complexes, photosystem I and II (PSI and PSII) act in series to catalyze the light-driven reactions in photosynthesis. PSII catalyzes the light-driven water splitting process, which maintains the Earth's oxygenic atmosphere (Renger, 2012). In this process, the oxygen-evolving complex (OEC) of PSII cycles through five states, S_0 to S_4, in which four electrons are sequentially extracted from the OEC in four light-driven charge-separation events. Here we describe serial femtosecond crystallography (SFX), a recently developed technique (Chapman et al., 2011) applied to time-resolved experiments on PSII nano/microcrystals. Structures have been determined from PSII in the dark S_1 state and after double laser excitation (putative S_3 state) at 5 and 5.5-Å resolution, respectively. The results provide evidence that PSII undergoes significant conformational changes at the electron acceptor side and at the Mn_4CaO_5 core

of the OEC. These include an elongation of the metal cluster, accompanied by changes in the protein environment, which could allow for binding of the second substrate water molecule between the "dangler Mn" and the cubane in the S_2 to S_3 transition, as predicted by spectroscopic and computational studies (Navarro-Garcia et al., 1999)(Isobe et al., 2012) This work shows the potential for time-resolved serial femtosecond crystallography (TR-SFX) for investigation of catalytic processes in biomolecules.

The first X-ray structure of PSII was determined to a resolution of 3.8-Å in 2001 (Zouni et al., 2001), revealing the protein's architecture and the overall shape and location of the OEC. In 2011, Shen and co-workers achieved a breakthrough in the structural elucidation by dramatically improving crystal quality, enabling determination at 1.9-Å resolution (Umena et al., 2011). This structure showed the OEC at near atomic resolution. However, the OEC was probably affected by X-ray damage, a fundamental problem in X-ray crystallography.

The X-ray damage problem may be overcome through the use of serial femtosecond crystallography (SFX), (Chapman et al., 2011)(Boutet et al., 2012)(Redecke et al., 2013) an advancement enabled by the advent of the X-ray free-electron laser (XFEL). In SFX, a stream of microcrystals in their mother liquor is exposed to intense femtosecond XFEL pulses, thereby collecting millions of X-ray diffraction 'snapshots' in a time-frame of hours. Each FEL pulse is so intense that it destroys the sample; however, the pulse duration is so short that diffraction is observed before destruction occurs (Barty et al., 2012).

Conventional X-ray structures correspond to a time and spatially averaged representation of biomolecules, leading to a 'static' picture. To capture dynamic processes such as water oxidation in PSII, time-resolved X-ray data can be collected using SFX (Aquila et al., 2012)(Neutze and Moffat et al., 2012). Conformational changes may be observed at a time-resolution ranging from femtoseconds to microseconds by combining visible laser excitation with the SFX setup and varying time delays between the optical pump and the X-ray probe snapshot. As partial reflections from crystals in random orientations are recorded, many snapshots must be collected for adequate sampling of the full reflections to determine structure factors and three-dimensional reconstruction. A time-resolved pump-probe experiment was performed in 2010 using PSI-ferredoxin crystals as a model system, in which changes in diffraction intensities, consistent with a light-induced electron transfer process in the PSI-ferredoxin complex and dissociation of the PSI-ferredoxin complex were seen (Aquila et al., 2012).

The catalytic reaction in PSII is a dynamic process. The oxygen evolution reaction is catalyzed by the oxygen evolving complex, in which the electrons are extracted from the Oxygen-Evolving Complex (OEC) in four sequential charge separation events through the S-state cycle (the Kok cycle), as shown in Fig. 3.1a (see (Renger, 2012) for a review). Recently, pump-probe simultaneous SFX diffraction and X-ray emission spectroscopy (XES) was reported (Kern et al., 2013) to investigate the dark S_1 state and the single flash (S_2-state) of PSII. The XES data show that the electronic structure of the highly radiation sensitive Mn_4CaO_5 cluster does not change during femtosecond X-ray exposure (Kern et

al., 2013). However, the quantity and quality of X-ray diffraction data was insufficient to determine if any structural changes occurred.

We report on microsecond time-resolved SFX experiments conducted at the CXI instrument (Boutet and Willams, 2010) at the Linac Coherent Light Source (LCLS) (Emma et al., 2010). The experimental setup is shown in Fig. 3.1b and c. We developed a multiple-laser illumination scheme that progressively excites the OEC in dark-adapted PSII nano/micro crystals by two laser pulses from the dark S_1 state via the S_2 state to the double-flash putative S_3 state. Not all PSII centers progress to the next S-state by a single saturating flash, which could lead to heterogeneities. Therefore the S-state reached in the double-flash experiment is indicated as 'putative S_3 state' throughout the manuscript.

a

2 H⁺ + PQ$_B^{2-}$ → ☐ + PQ$_B$ → PQ$_B$

PQH$_2$

S$_1$
Dark

+3, +3, +4, +4
Mn$_4$Ca

PQ$_B^-$ S$_0$ S$_2$ PQ$_B$

+3, +3, +3, +4   +3, +4, +4, +4

O$_2$

H$_2$O

+4, +4, +4, +5*   +4, +4, +4, +4

S$_4$ S$_3$ H$_2$O

PQ$_B^-$

PQ$_B$ PQ + ← ☐ ← PQ$_B^{2-}$ + 2H⁺

PQH$_2$

b

527 nm pump laser (60 Hz)

Pump laser trigger

Illuminated region 400 µm FWHM

Liquid jet

Interaction region

~50 fs

KB mirrors

XFEL pulses (120 Hz)

120 Hz

Dark   Light   Dark   Light

c

527 nm pump laser pulses   XFEL pulse

90 ns   150 ns   50 fs

Trigger   110 µs   210 µs   570 µs   Time
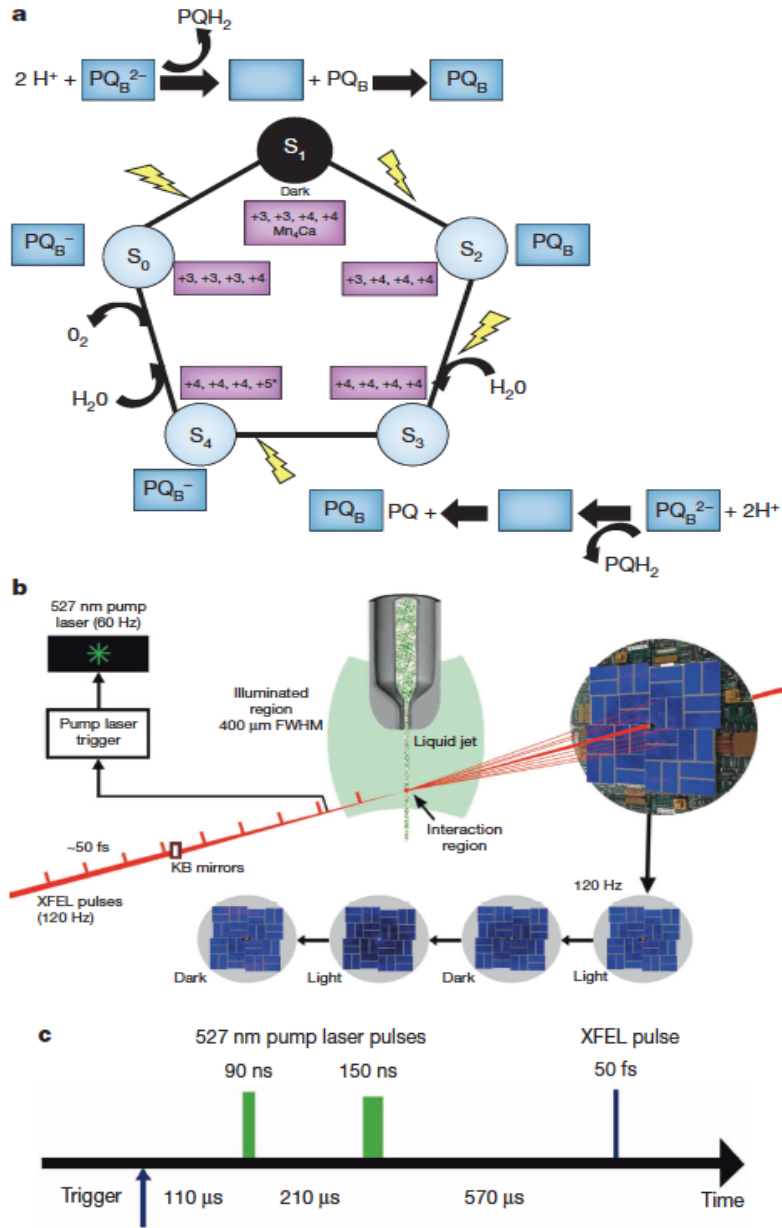
Figure 3.1: Experimental Schemes for the Time-Resolved Serial Femtosecond Crystallography Experiments on Photosystem II. a, S-state scheme of the oxygen-evolving complex depicting changes the oxidation state of the 4 manganese ions of the Mn_4CaO_5 cluster in the S-state cycle and indicating the reduction of the plastoquinone (PQ) to plastoquinol (PQH_2) in the Q_B site. b, Experimental setup. The crystal-steam of photosystem II was exposed to two subsequent optical laser pulses at 527 nm before interacting with the femtosecond X-ray FEL pulses. With a FEL frequency of 120 Hz and triggering of the laser at 60 Hz, X-ray diffraction patterns from crystals in the dark state and 'light' double-flash state alternate. c, Laser excitation scheme. The first 527 nm laser pulse excited the crystals 110 µs after the trigger pulse. The delay time between the first and second 527 nm laser pulse was 210 µs, with X-ray diffraction data collected 570 µs after the second laser pulse.

The diffraction patterns collected from dark and illuminated crystals were sorted into two data sets. Using the 'hit finding' program Cheetah (Barty et al., 2014), 71,628 PSII diffraction images were identified from the dark diffraction patterns and 63,363 were identified from the double-flash patterns. From these hits, 34,554 dark state patterns and 18,772 double-flash patterns were indexed and merged to reduce stochastic errors using the *CrystFEL* software suite (White et al., 2012) (see Table 3.1a,b). The data were indexed as orthorhombic, with unit-cell parameters of $a = 133$-Å, $b = 226$-Å, and $c = 307$-Å for the dark state, and $a = 136$-Å, $b = 228$-Å, and $c = 308$-Å for the double-flash state. The distributions of unit cell dimensions are shown in Fig. 3.2 and Table 3.1a,b. The data clearly supports an increase in unit cell dimensions in the double-flash state, with the largest difference detected for the a-axis. Two factors may explain the change in unit cell constants, lower indexing rates, and a slight decrease in resolution of diffraction: crystal degradation upon laser illumination or significant structural changes upon the transition to the double flash state, which may represent the putative S_1 to S_3 transition. To distinguish between these two possibilities, we collected data with triple-

flash excitation of the PSII crystals, where at least part of the PSII centers may have reached the putative transient S_4 state. Preliminary data evaluation of the triple-flash data set (i.e., putative S_4 state) shows similar unit cell dimensions and crystal quality as the dark S_1 state (see Fig. 3.2 and Table 3.1a). This suggests that conformational changes induced in PSII by the double-flash excitation (i.e., during the putative S_1 to S_3 transition) are reversed after excitation with the third flash (in the putative S_3 to S_4 transition). This unit cell changes may be caused by dissociation of the mobile plastoquinone PQ_B from the Q_B binding pocket after double-flash excitation, when PSII may reach the S_3-state (see Fig. 3.1a). The structural changes leading to the difference in unit cell constants are likely most significant at the stromal side of PSII where the quinone bindings sites are located. To avoid structural heterogeneity at the acceptor side by partial re-occupation of the PQ_B binding site, no quinone was added to the crystals for the double pump experiments. We thereby may have "trapped" PSII in the double flash experiment in the putative S_3-state conformation with an empty Q_B binding pocket. In order to transition from $S_3 \rightarrow S_4$, an electron acceptor must replenish the empty Q_B binding site. Therefore, the plastoquinone derivative $PQ_{decyl,}$ which diffuses into the Q_B pocket, was added to the crystals used for the triple-flash excitation data set. With the Q_B binding site re-occupied, the change in unit cell constants is reversed.

Diffraction data from the dark and double-flash states were evaluated to 5-Å and 5.5-Å resolution, respectively; the data refinement statistics are shown in Table 3.3 Since each diffraction pattern represents a thin cut through the Ewald sphere, only partial reflections were recorded. A high multiplicity of observations is therefore needed for

each Bragg reflection to obtain full, accurate structure factors. The average multiplicity per reflection was 617 for the dark state data set and 383 for the double-flash data set over the whole resolution range (see Table 3.2a,b). Table 3.1c (Details in Methods section) shows a comparison of the data statistics of this work with the S_1 and S_2 data in (Kern et al., 2013).
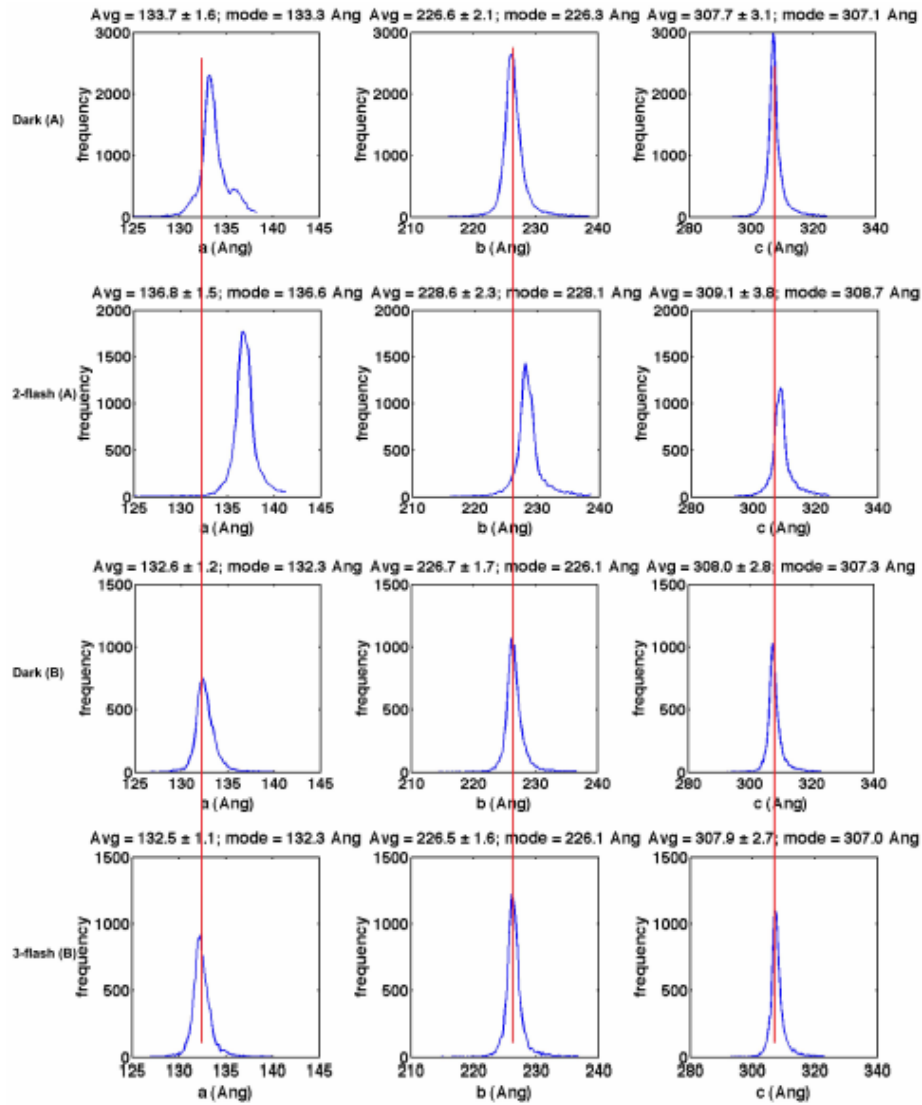
Figure 3.2: Distribution of Photosystem II Unit Cell Constants of 4 Different Femtosecond Crystallography Data Sets. Row 1 shows unit cell constants of the dark data set (S_1 state) collected at the CXI instrument in January 2012, (experiment (A)). Row 2 show unit cell constants of the double-flash data set (putative S_3 state) collected at the CXI instrument in January 2012 (experiment (A)). Row 3 shows unit cell constants of the dark data set (S_1 state) collected at the CXI instrument in June 2012 (experiment (B)) (quinone PQ_{decyl} was added to these crystals to allow replacement of the quinone for triple excitation). Row 4 shows unit cell constants of the triple-flash data set (putative S_4 state) collected at the CXI instrument in June 2012 (experiment (B)). The comparison of unit cell constants shows that significant changes in the unit cell constants are observed after double-flash excitation of photosystem II. These changes are fully reversed when photosystem II is excited by a three laser flashes. Although the number of indexed patterns currently available does not yet allow for the determination of an accurate structure of the PSII after triple excitation, the data allows extraction of information on the hit rates, indexing rates and unit cell constants, showing that the unit cell constants are identical for the dark S_1 and triple-flash state.

The data were phased by molecular replacement using a truncated version of the 1.9-Å structure (PDB code 3ARC) (Umena et al., 2011). Rigid body refinement (*phenix.refine*) (Afonine et al., 2012) was performed for both the dark and double-flash structures (see Methods section) for further details on molecular replacement and refinement). To reduce model bias, we calculated omit maps and simulated annealed maps (SA-omit maps) for the dark and double-flash data, omitting the coordinates of the Mn_4CaO_5 cluster from the model.

Fig. 3.3a-c shows the arrangement of protein subunits and cofactors of photosystem II, including the electron transport chain. The comparison of the electron density maps for the dark state (green) and the double-flash state (white) at a contour level of 1.5-σ is shown in Fig. 3.3d-f. Both maps show clear electron densities for the trans-membrane helices as well as loops and cofactors. Additional electron density maps for representative structural elements of PSII are shown in Fig. 3.4, and 3.5. Overall, the protein fits into the

electron densities for the dark and double-flash states and matches the high resolution structural model. However, differences appear in regions of the Mn_4CaO_5 cluster and the acceptor side, where the quinones and the non-heme iron are located. Determining the significance of these changes and their correlations is complicated at the low resolution of the data. Fig. 3.3g-i shows detailed views of the loops at the acceptor side of PSII. The quinones are not visible at the current resolution of 5 Å. The maps indicate differences between the electron densities of the dark and double-flash states in the loop regions and also in the position of the non-heme iron that is coordinated by the loops.
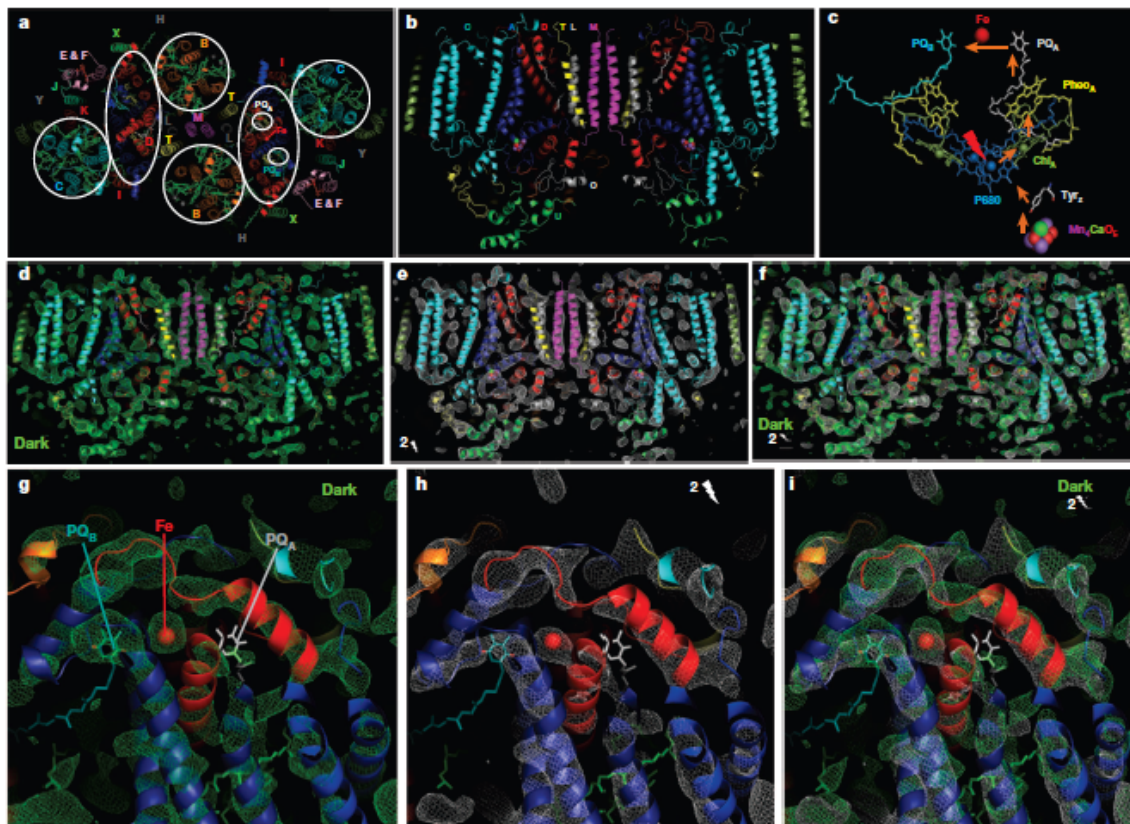
Figure 3.3: Overall Structure and Omit Map Electron Density of Photosystem II.
(a), Trans-membrane helices and cofactors in photosystem II (stromal view density map). The proteins are named according to their genes and labeled with colored letters. (b), Side view of PSII at its longest axis along the membrane plane. (c), Electron transport chain of PSII (P680 (blue), accessory chlorophylls (smudge-green), pheophytins (yellow) and plastoquinones PQ_A (white) and PQ_B (cyan)); atoms of the OEC are depicted as spheres (Mn purple, Ca green, O light red). d–f, Omit map electron densities (view as in b) at 1.5 σ for the dark state (S_1) (green) (d), double-flash state (putative S_3 state) (white) (e) and overlay of the two omit maps (f). g–i, Omit maps (1.5 σ) of the electron acceptor side of photosystem II for the dark (S_1) (green) (g), double-flash (putative S_3 state) (white) (h) and overlay of the two omit maps (i). Note that changes include a shift of the electron density of the non-heme iron.

Figure 3.4: Omit Map of the Dark and Double-Flash States of the Most Peripheral Photosystem II Membrane Integral Subunits and the Chlorophylls of the Primary Electron Donor P680. a–d, This picture features the peripheral subunits PsbZ (grey-green), PsbK (brown), PsaH (grey) and the core-antenna protein CP43 (PsbC) (cyan). The omit electron density map at the contour level of $1.5\,\sigma$ for the dark (S_1) state is depicted in green (a) and the double flash (putative S_3) state is depicted in white (b). c, The overlay of the two omit maps is shown at the contour level of $1.5\,\sigma$. The globular densities between PsbC and PsbK correspond to antenna chlorophylls. The figure shows that even the electron density for the two most peripheral helices that belong to subunit PsbZ are well defined. We also note the good match of the strongly kinked helix of PsbK

between the S_1 and S_3-state electron density maps. d, The subunits are labeled according to their genes in the view of the structural model. e–h, The figure features the surroundings of the two chlorophylls of P680 and the accessory chlorophyll of the active electron transfer branch of photosystem II (see Fig.3.3c). The omit electron density map at the contour level of $1.5\,\sigma$ for the dark (S_1) state is depicted in green (e) and the double flash (putative S_3) state is depicted in white (f). (g), The figure also shows the overlay of the two omit maps at the contour level of $1.5\,\sigma$. (h), Model of the chlorophylls of the primary electron donor P680 without electron density map.
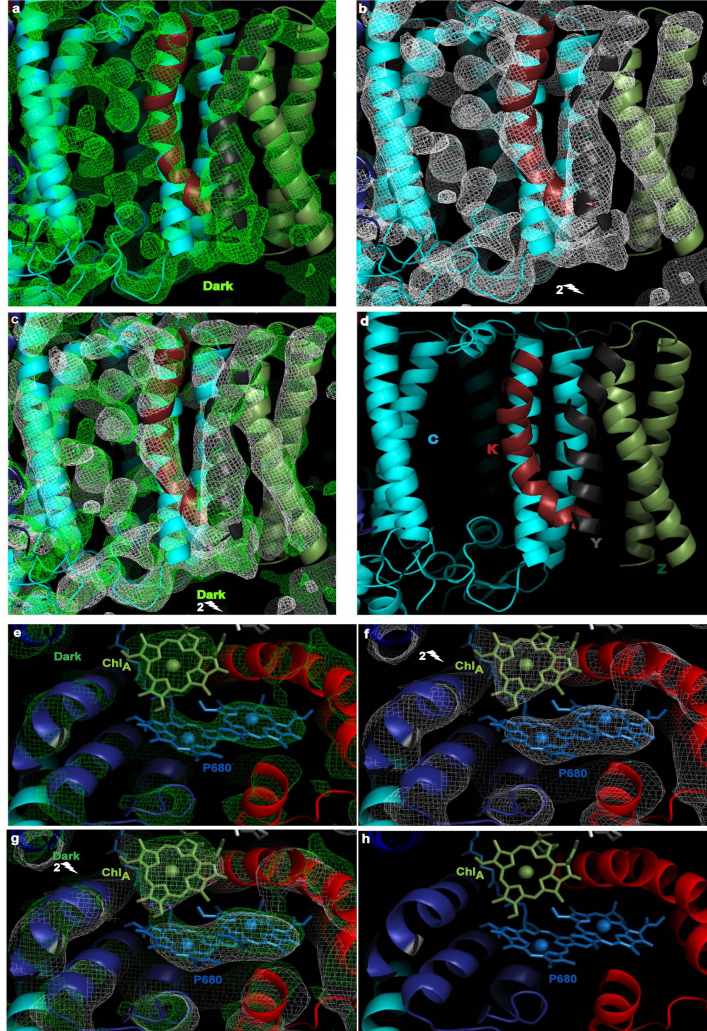
Figure 3.5: The Electron Acceptor Side of Photosystem II. Omit map electron density and structural model of the dark and double-flash state of photosystem II, the view from the stromal side onto the membrane plane. a–d, The loops that coordinate the non-hem iron and cover the quinone binding sites looking from the stromal side onto the membrane plane. The omit electron density map at the contour level of $1.5\,\sigma$ for the dark (S_1) state is depicted in green (a) and the double-flash (putative S_3) state is depicted in white (b). (c), The overlay of the two omit maps at $1.5\,\sigma$. (d), The structural model
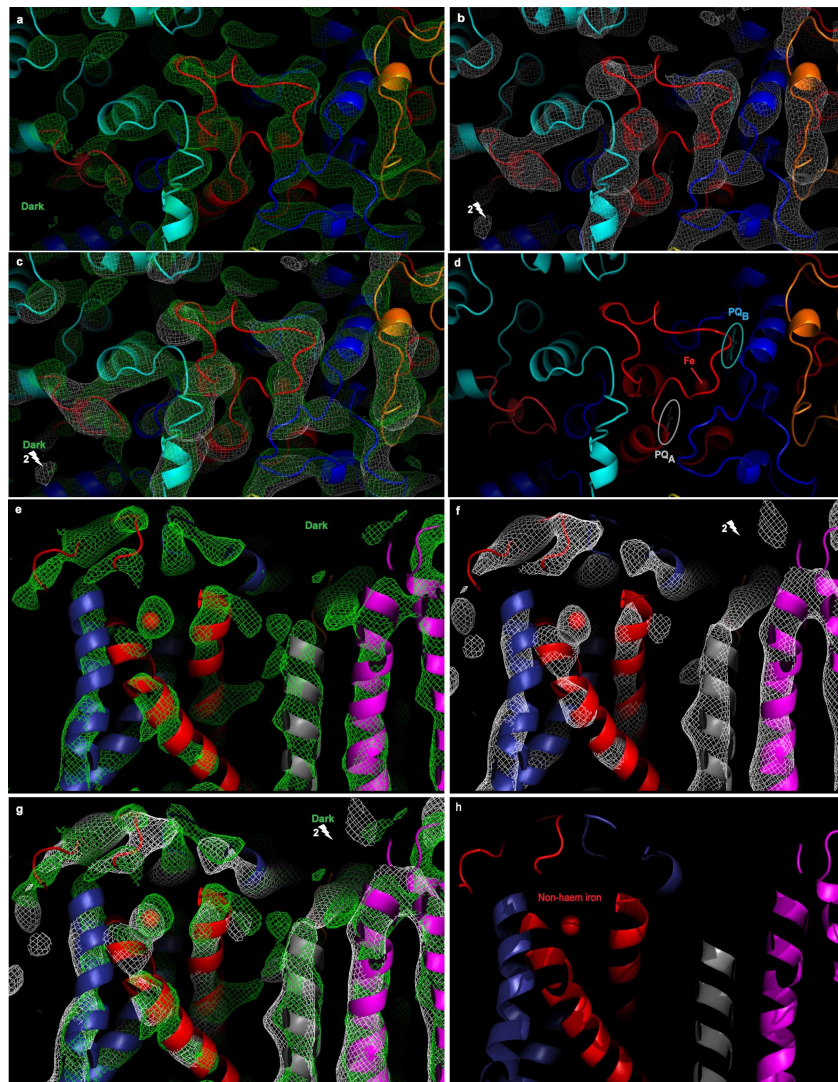
indicates the positions of PQ_A and PQ_B as well as the non-heme iron located below the loops. We note that the electron densities of the loop regions at the electron acceptor side show significant differences between the dark and the double flash states. The electron density of both states may suggest a conformation of the loops that could differ in their backbone trace from the model derived from the 1.9-Å structure from (Umena et al., 2011). e–h, The side view of the acceptor side of photosystem II along the plane in the membrane. The omit electron density map at the contour level of $1.5\,\sigma$ for the dark (S_1) state is depicted in green (e) and the double-flash (putative S_3) state is depicted in white (f). (g), The overlay of the two omit maps featuring the changes in the position of the non-heme iron and loop regions at the contour level of $1.5\,\sigma$. (h), Model of the electron acceptor side of photosystem II. The protein subunits are color coded as in Fig. 3.3a of the main text; the non-heme iron is depicted as a red sphere. The tightly bound plastoquinone PQ_A is shown in white, the mobile plastoquinone PQ_B is depicted in cyan.

In the following part, we focus on the structure in the undamaged dark S_1 state of the metal cluster in the OEC and the potential light-induced structural changes that may occur during the S-state transition. Figure 3.6 shows the SA-omit map of the OEC in the dark S_1 state for the Mn cluster in PSII with the 1.9-Å X-ray structure in Umena et al. (Umena et al., 2011). Interestingly, the electron-density map of the 'dangler' Mn atom from the 1.9 Å structure is located outside the dark S_1 state electron density, a feature also visible in the electron density map of (Kern et al., 2013). These structural observations are consistent with spectroscopic results, which indicate that the distance between the dangler Mn and the Mn_3O_xCa distorted cubane is indeed shorter in the dark S_1 state than in the 1.9-Å structure based on the synchrotron data, which might be influenced by X-ray induced reduction of the Mn ions in the metal cluster (Luber et al., 2011)(Davis et al., 2013). This shorter distance is in agreement with density function theory (DFT) studies (Luber et al., 2011)(Ames et al., 2011)(Isobe et al., 2012) based on the 1.9-Å structure of PSII (Umena et al., 2011), however, the current resolution limit of 5 Å does not allow a quantitative assessment.
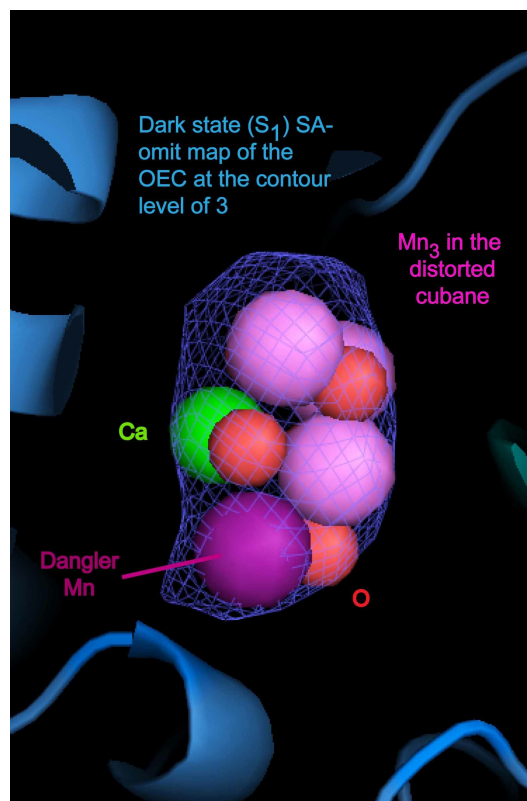
Figure 3.6: Simulated Annealed Omit Map of the Mn$_4$CaO$_5$ Cluster of Photosystem II. The electron density of the dark state of photosystem II. This figure shows the superimposed SA-omit maps for the dark (S$_1$) (blue) state of the Mn$_4$CaO$_5$ cluster. We use a different color scheme for the SA-omit maps and the 'regular 2F_o-F_c' omits maps to allow the reader a better orientation of the type of map shown in each figure. The electron density is shown at the contour level of 3.0 σ to ensure that it solely features the metal Mn$_4$CaO$_5$ cluster. The X-ray structure of the OEC at 1.9 Å from (Umena et al., 2011) is placed inside the SA-omit map for comparison. The nomenclature for the Mn atoms proposed in (Umena et al., 2011) is used for the color-coding of the individual Mn atoms in the cluster. The Mn ions that form the distorted Mn$_3$O$_x$Ca cubane (Mn1, Mn2 and Mn3) are depicted in light pink, while Mn4 (violet) (named the 'dangler' Mn) is located outside the cubane. This figure shows that the dangler Mn sticks out of the SA-omit map electron density, which indicates that this Mn atom may be located in closer proximity to the Mn$_3$CaO$_X$ cubane in the dark S$_1$ state that is not influenced by X-ray damage.

The mechanism for water splitting has been intensely debated and many models have been proposed. The recent 1.9 Å X-ray structure (Umena et al., 2011) formed the basis for more detailed theoretical studies of the process, yet the proposed mechanisms differ

(Ames et al., 2011)(Isobe et al., 2012)(Rivalta et al., 2012)(Siegbahn, 2012). Based on our TR-SFX structural data, we analyzed differences between the electron-density maps of the OEC, derived from the dark and the double-flash data sets. Fig. 3.7a and b show the SA-omit maps calculated for dark (blue) and double-flash state (yellow) and compared with the model of the metal cluster from the 1.9 Å structure (Umena et al., 2011) (Fig. 3.7c). The Mn_4CaO_5 cluster was omitted from the model for the calculation of the SA-omit map, which includes annealing at a virtual temperature of 5000 K to minimize phase bias. The SA-omit electron densities of the dark and double-flash states differ in the shape and position, as well as in the protein environment, of the Mn_4CaO_5 cluster. The dark state simulated-annealed (SA)-omit electron density for the OEC protein environment matches the model of the 1.9 Å structure (Umena et al., 2011), whereas the SA-omit map of the double-flash state differs significantly. Any interpretation of changes in the protein environment of the OEC is highly speculative at a resolution of 5-Å and heterogeneities in the S-state transitions. However, the SA-omit map of the double-flash state is suggestive of conformational changes, which may indicate of a movement of the CD loop (including the ligand D170 of the D1 protein) away from the OEC cluster. If confirmed at higher resolution, this could explain mutagenesis studies that questioned D170 as a ligand in the higher S-states (Debus et al., 2005). Furthermore, in the double flash state, the electron density of the metal cluster extends and shows a new connection to the AB loop at site where D61 of the D1 protein is located. Although D61 only serves as a second sphere ligand in the 1.9-Å crystal

95

structure (Umena et al., 2011) mutagenesis studies indicated an important role in the water oxidation process as the S_2 to S_3 transition is blocked in D61 mutants.
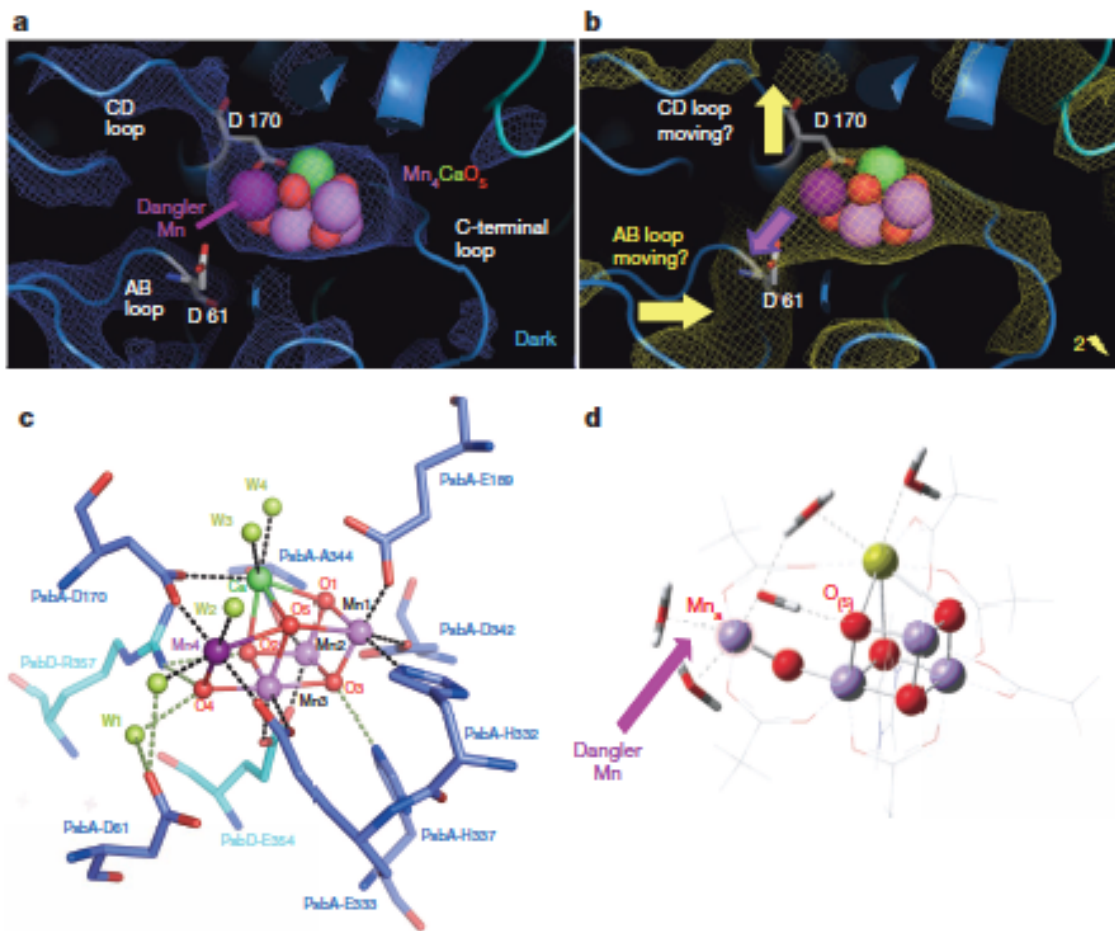
Figure 3.7: Comparison of the OEC Simulated Annealed Omit Maps of the Dark and Double-Flash States. a, b, At 1.5 σ for dark and double flash states of the Mn_4CaO_5 cluster of PSII for the dark S_1-state (blue) (a) and double-flash, putative S_3 state (b) with the 1.9 Å structural model (3ARC) from Umena et al., 2011. Mn in the distorted Mn_3O_xCa cubane (Mn-1 to Mn-3) (light-pink), dangler manganese (Mn-4) (violet), calcium (green) and oxygen (red). (c), 1.9-Å crystal structure of the Mn_4CaO_5 cluster with ligands from Umena et al., 2011 (PDB accession code: 3ARC). (d) Proposed model of the S_3 state based on DFT calculations by Isobe *et al.,* (Isobe et al., 2012) (reproduced with permission of The Royal Society of Chemistry) Larger diversions in the SA-omit map of the double-flash (putative S_3 state) include potential movement of the loop connecting trans-membrane helices C/D (CD loop) with D170 and AB loop (with D61), and an increase of the distance between the dangler Mn and the Mn_3O_xCa cubane (violet arrow).

The change in the electron-density of the OEC is suggestive of an increase in the distance between the cubane and the 'dangler' Mn and distortion in the cubane in the double-flash state. The observed electron densities (Fig. 3.7a, b) of the dark state and double flash state are consistent with conformational changes predicted in a recent DFT study of the S_3 state in Isobe et al., 2012 shown in Fig. 3.7d. The increased distance between the cubane and 'dangler' Mn could allow the second 'substrate' water molecule to bind between the Mn_3OxCa cubane and the dangler Mn in S_2 to S_3 state transition. It was shown by EXAFS (extended X-ray absorption fine structure) that the $Mn–Ca^{2+}$ distances in the Mn_3O_xCa cubane shrink in the S_3 state (Pushkar et al., 2008). Although the Jahn-Teller effect extends the distances between metals in the lower S-states of the OEC (Mn oxidation states +II, +III and +IV), a shrinking of the Mn_3O_xCa cubane is predicted in the S_3 state when all 4 Mn in the OEC have reached the oxidation state +IV. In the S_1 state, two Mn atoms are at +III and other two Mn atoms in +IV oxidation states. This comparison of the electron density in the dark and the double-flash states may indeed suggest an overall decrease in the dimension of the Mn_3O_xCa cubane in the double-flash state, which is in good agreement with the proposed S_3 state EXAFS and XES model (Dau et al., 2012). The consistency of spectroscopy and DFT studies with our observations may provide preliminary indications that a significant fraction of the OEC centers in our crystals have reached the S_3 state in the double flash experiment.

In light of new results on theoretical modeling of the OEC (Luber et al., 2011)(Siegbahn et al., 2012)(Isobe et al., 2012)(Perez Navarro et al., 2013)(Cox et al., 2013)(Rivalta et al., 2012) we further examined the SA-omit maps in the dark and

double-flash states for differences in the metal cluster that can be detected even at low resolution and discuss the results here in light of recent computational and spectroscopic studies on the metal cluster. The changes in the density of the Mn_4CaO_5 metal cluster are suggestive of an increase of the distance between the cubane and the 'dangler' Mn and a distortion of the cubane in the S_3-state. The observed electron densities are compared in Fig. 3.7a and b with the recent theoretical studies of Isobe and coworkers (Isobe et al., 2012), shown in Fig. 3.7d, who predicted a "breakage" of the dangler Mn from the cubane cluster in the S_3-state. Additionally, EXAFS data constrains the extent of the movement of the dangler Mn relative to the cubane (Grundmeier and Dau, 2012)(Yano and Yachandra, 2008). The increase in distance could allow for the binding of the second substrate water molecule between the dangler Mn and the Mn_3CaO_x cubane. The presence of a substrate water molecule between the dangler Mn and the distorted cubane in the higher S-states, has also been predicted to be essential for the catalytic mechanism in a recent DFT model of the full catalytic S-state cycle, including modeling of the substrate water exchange (Siegbahn, 2012)(Siegbahn, 2013).

In addition to the elongation, the overall dimensions of the Mn_4CaO_5 cluster appear to condense in the double-flash data set that may represent the putative S_3-state. This may include shrinking of the distance between the Ca^{2+} and the 3 Mn in the distorted cubane. EXAFS studies on PSII, where the Ca was substituted with Sr, showed significant changes in Mn-Mn or Mn-Sr distances in the S_3-state (Pushkar et al., 2008), which were interpreted to indicate the distance between Mn and Ca would shrink in the S_3-state. Our experimental findings suggest a shrinking of the Mn_4CaO_5 cluster in double-flash state,

which supports the hypothesis of a condensation of the Mn_3O_xCa cubane part of the Mn_4CaO_5 cluster in S_3 (Isobe et al., 2012). Models of Mn-oxygen cubane compounds show an increased distance between the Mn and O atoms in the cubane at lower oxidation states (+2 and +3) due to the Jahn-Teller (JT) effect (Rivalta et al., 2012)(Yamaguchi et al., 2013). Distances derived from a recently published model Mn-O and Mn_3O_xCa cubane structures (Kanady et al., 2011) indicate that Mn-O distances depend on the oxidation states of the Mn-ions: the average Mn^{+2}-O distance is 2.2-Å, the average Mn^{+3}-O distance is 2.0 Å and the average Mn^{4+}-O distance is 1.8-Å. Two models have been proposed on the basis of X-ray absorption and emission spectroscopy, one describes the S_3-state as Mn (+3 +4 +4 +4) and the other proposes Mn (+4 +4 +4 +4) (Dau et al., 2012)(Yano and Yachandra, 2007). In the model where all Mn ions have reached the Mn^{+4} oxidation state, a significant shrinking of the dimension of the cluster is expected due to the lack of the JT distortion with the average Mn-O distance being reduced to 1.8-Å (Yamaguchi et al., 2013). The shrinking of the overall dimensions of the metal cluster, supported by our maps of the double-flash state, appears to be in agreement with the studies on model compounds. This indicates that the JT distortion diminishes in the putative S_3-state during progression of the S-states cycle when all Mn reach their +4 oxidation states (Kanady et al., 2011).

The SA-omit maps of the dark (S_1) and the double-flash (putative S_3) states may be also indicative of changes in the protein environment of the Mn_4CaO_5 cluster. While the electron density map in the dark S_1-state overall follows the protein backbone of the 1.9 Å structure, larger perturbations of the protein environment of the cluster are visible in

99

the double-flash state. The double-flash state electron density map may suggest a movement of the loop which connects the trans-membrane helices C and D at the lumenal site (the CD loop, including D170) away from the metal cluster and a movement of the AB loop (connecting the trans-membrane helices A and B) into closer vicinity to the cluster, which may allow D61 to be become part of the ligand sphere of the metal cluster. While this interpretation of changes in the protein environment of the cluster is highly speculative at the given resolution, it could explain the results of mutagenesis studies on PSII. Although the mutation of D170 (which coordinates the dangler Mn and the Ca in the 1.9 Å structure of PSII) into an alanine has no strong effects on the oxygen evolution function (Debus et al., 2005)(Chu et al., 1995) less than 15% of the oxygen evolution function remains in the Asp61Ala mutation (Hundelt et al., 1998)(Dilbeck et al., 2012). This mutagenesis result was difficult to rationalize because D61 is found only in the second ligand sphere of the OEC in the 1.9-Å structure (Umena et al., 2011). However, our SA-Omit electron density map of the metal cluster in the double-flash state shows a connection to the protein electron density in close vicinity to Asp61 (see Fig. 3.7b). This finding may provide a first indication that Asp61 may serve as a ligand to the dangler manganese in the higher S-states. While details of the conformational changes cannot be unraveled at the current resolution of 5 Å, the comparison of the dark and double-flash state SA omit maps provide an indication that the protein ligand sphere of the Mn_4CaO_5 cluster may undergo significant changes when the OEC reaches the double-flash (putative S_3) state.

Our time-resolved SFX study captures the image of PSII after it has been excited by 2 saturating flashes and provides experimental evidence for structural changes occurring in the putative S_3 state of the OEC, accompanied by structural changes at the acceptor side of PSII. As the resolution is limited to 5 Å, the interpretation of the changes observed is preliminary. This work is a proof-of-principle that time-resolved SFX can unravel conformational changes at moderate resolution, and lays the foundation for the high resolution analysis of PSII at all stages of the water oxidation process in the future. To unlock the secrets of the water-splitting mechanism by TR-SFX at atomic detail, the resolution must be further improved and structures must be determined from all the S-states with multiple time delays.

3.2. Discussion of Methods

In this section and following sub-sections, all the methods and technical details that form the basis for the success of these challenging experiments will be thoroughly discussed.

3.2.1. CXI Instrumental Set-up and Sample Delivery for TR-SFX Experiment on PSII Crystals in the Dark and Double Flash State

The microsecond time-resolved SFX (TR-SFX) experiment on micro crystals of PSII was performed at the Coherent X-ray Imaging (CXI) instrument (Boutet and Williams, 2010) at the Linac Coherent Light Source (LCLS), located in SLAC (Stanford Linear Accelerator center). A schematic of the pump-probe experimental set up for this study is illustrated in Fig. 3.1b and 3.1c. A stream of PSII crystals (size of ~2-μm) was delivered

to the interaction region of 1-μm diameter of FEL beam as a suspension of crystals with coaxial gas focusing liquid jet injector (see (Weierstall et al., 2012)(Weierstall et al., 2008)(DePonte et al., 2008). The PSII crystals were illuminated with two subsequent visible laser pulses ($\lambda = 527$-nm) at a frequency of 60-Hz, followed by probing with 50-fs X-ray pulse at a frequency of 120-Hz. Thus, the experimental strategy of pump and probe allows the collection of diffraction data from "dark" and "light" states alternatively. The injection process was supported with an anti-settling device (Lomb et al., 2012) to keep those crystals as suspended in the mother liquor and permit a temperature control of the sample. The stainless steel syringe containing the crystals was mounted on a rotating holder (i.e., anti-settling device), which has a peltier cooler to keep the temperature at 10°C. This helped to keep the PSII crystals at the growth temperature until they were injected to X-ray FEL beam. The glass made nozzle tips were polished for visible pump laser excitation. The upstream part of the nozzle tip was black coated to prevent any pre-illumination of the crystal suspension, upstream of the nozzle tip. The gas focusing liquid injector provided a liquid jet of crystal suspension of 4-μm. The velocity of the liquid jet was approx.12-μl/min. The 4-μm diameter liquid jet was intersected with X-ray focal area of 2-μm^2 of full-width half maximum (FWHM). For the entire experiment, X-ray photon energy of 6.0 keV ($\lambda = 2.05$-Å) and X-ray pulse duration of 50-fs were used. The X-ray diffracted intensities were recorded at the Cornell-SLAC Pixel Array Detector (CS-PAD), which is made of 64 panels of 194 x185 pixels, totaling of 1752 x 1752 pixels including gaps between the tiles (Hart et al., 2012).

3.2.2. Design and Set-up Pump-Laser Instrument and Excitation Scheme for PSII Crystals to Double-Excited State

In this TR-SFX experiment, PSII crystals were excited twice with two optical pump laser pulses, which were generated by a frequency doubled Nd:YLF laser. This was a nano-second pulsed laser emitting visible light at the wavelength of 527-nm. The laser was fiber-coupled and set up on a table outside the experimental chamber. It was channeled through an optical fiber into the chamber and onto the head of the liquid injector. This wavelength provided a reasonable compromise between transmission (~20% estimated) and absorption of light in the PSII crystals. A calculation was initially done to optimize the laser power onto 2-5-μm PSII crystals. Thus, the homogeneous excitation of the PSII crystal was ensured with the "saturating" laser pulses. The optical double pulse was produced by an active Q-switch with 'on-times' chosen such that pulse energies of both pulses match. The desired timing of the optical laser pulse with respect to X-ray pulse was achieved through the LCLS/CXI (Emma et al., 2010) event code generator or reader (EVG/EVR), with a precision in timing signal less than 1-μs before the X-ray pulse arrived. There was also another photodiode, namely Acqiris digitizer, which also recorded timing of optical laser pulses. This resulted in one pump laser pulse of 90-ns and second one of 150-ns. This ensured the number of photons incident onto the crystal is same in both pulses. A schematic of laser excitation is provided in Fig. 3.1c. The laser was focused at the tip of the nozzle (which is 100-μm upstream of X-ray interaction point) with a spot diameter of 400-μm with an incident pulse energy of 6-μJ at the sample, i.e., three times more than the required pulse energy to optimally excite 2-μm

PSII crystal (Han et al., 2012). The laser had a top-hat profile. The laser beam diameter and the aim point was chosen based on desired pump-probe time delays, sample flow profile, flow speed inside the capillary (average speed of 85-mm/sec in 50-μm inner diameter capillary) and in liquid jet (12-m/sec in 4-μm diameter jet). This ensured the illumination at the tip of the nozzle and at the liquid jet. It also allows double excitation of crystals before they were probed by X-ray pulse. The time delays between pump-laser pulses were chosen based on the 3 times of the half-life of the excited states of PSII in the Kok cycle to establish the excited states with optimal population in that state. This half-life of the excited states of PSII was estimated based on measurement made in 1980s on PSII membrane fragments, not crystals. The time delay between 1st pump pulse and 2nd one was set to 210-μs, which is the 3 times of the half-life of S_2 state (i.e., $t_{1/2}$ = 70-μs (Dekker et al., 1984)). Similarly the delay between 2nd pump pulse and X-ray probe pulse was set to 570-μs, which is 3 times of the half-life of S_3 state (i.e., $t_{1/2}$ = 190-μs (Dekker et al., 1984)) (See Fig. 3.1c). The optical pump laser works at 60-Hz frequency, while the X-ray FEL pulse is triggered at 120-Hz frequency. This led to the fact that every alternate diffraction pattern was collected from the "light" state. Since, no pumping scheme is 100% efficient in advancing the reaction, thereby not all PSII crystals can be ensured to reach S_3 excited state purely. This may lead to heterogeneity of different states. Thereby, the state, reached after double excitation, was coined as **double-flash state or putative S_3 state**. However, an independent measurement (Kupitz and Basu et al., 2014) for population of the S_3 state using EPR showed the approx. 72% of PSII molecules had advanced into S_3 state with two-lasers excitation. This population estimate is also

104

supported by another independent study by Styring and co-workers, (Han et al., 2012) which shows the transition rate of 75% from an EPR study on spinach under optimized condition. The molar extinction coefficient of chlorophyll ($\varepsilon_{527nm} = 8022$ L x mol^{-1} x cm^{-1}) was determined from dissolved PSII crystals at 527-nm. This was then used to calculate the optimal pulse energy of 2.3-mJ required to fully pump a 1-2-μm PSII crystal. We ensured from our "saturating" pumping scheme that there is zero probability of 0 photon absorption incident. It is noted that each PSII molecule contains 35 chlorophylls, 10 carotenoids. With so many pigments and use of "saturating" pulses, it is very unlikely to occur zero-photon absorption event. Given the molar extinction coefficient of chlorophylls ($\varepsilon_{527nm} = 8022$ L x mol^{-1} x cm^{-1}), the absorbance of PSII molecule at 527-nm wavelength is 0.12. It was calculated that the optical laser pulse at 6 μJ, 90 ns pulse duration at 527 nm contained 1.6×10^{13} photons. The volume of the stream (nozzle and jet) that was irradiated by the pulse contained 1.3×10^{12} molecules of P680. With the ~ 10-fold excess of photons we estimate the probability of zero-photon absorption is less than 2%.

Excitation energy transfer in PSII is limited by transfer to the trap (P680) (Raszewsky and Renger, 2008), therefore when one of the excitons in the antenna system reaches P680, P680* is formed and charge separation takes place. The kinetics of the initial charge separation events have been determined by femtosecond transient absorption spectroscopy combined with kinetic compartment modeling (Holzwarth et al., 2006). The initial charge separation takes place in 1.5-ps, the electron is then transferred via the accessory Chl$_{D1}$ to Pheophytin$_{D1}$, forming the charge separated state P680$^+$ Pheo$^-$ in 5.5-

ps. From Pheo$^-$ the electron is transferred in 200-ps to plastoquinone Q_A forming Q_A^- and from Q_A^- in 200-400 μs (de Wijn and van Gorkom, 2001) to the plastoquinone Q_B forming the semiquinone Q_B^- in the Q_B binding site.

At the donor site P680$^+$ is reduced to P680 in 200-ns (Renger and Schlodder, 2011) by the redox active Try$_Z$ forming Try$_Z^+$, which subsequently is reduced by one electron from the OEC, which changes its redox state thereby progressing the OEC in the S-state cycle. When two of the antenna pigments are excited in one saturating flash, the second exciton will reach P680$^+$ before it is reduced to P680. The excitation energy is then quenched by P680$^+$ or Q_A^- (Shinkarev and Govindjee, 1993) and dissipated as heat. Although, P680$^+$ to P680 conversion is a relatively fast process (i.e., 200-ns), still the reduction of Try$_Z^+$ by extracting an electron from OEC is a slower process (i.e., 70-μs during S_1->S_2 conversion). It means even if it is possible that P680 will become available to absorb the second exciton, still the OEC cannot be oxidized for a second time until Try$_Z^+$ has been reduced to Try$_Z$ by the OEC, which occurs with a time constant of 70-μs. So, it is clear that a second exciton will be dissipated as heat. Thereby, even multiple photon absorption by the PSII antenna system leads in most cases only to one charge separation event.

There is always a small chance that the excitation energy is lost by heat or fluorescence before it reaches P680 leading to "misses", or that the excitation of two pigments leads to two subsequent charge separations (which can happen when the second exciton reaches the reaction center after P680$^+$ has already been reduced) leading to small numbers of so called double hits. Thereby, accounting for such small fraction multiple

photon absorption events, the state reached after double excitation was called "putative S_3" state or double-flash state.

3.2.3. Processing and Data Evaluation of Time-Resolved SFX Data from PSII Crystals

A total of 5,528,071 diffraction frames (combining "dark" (S_1) and "light" (putative-S_3)) was recorded on the CS-PAD detector during the experiment. All diffraction frames were pre-processed with "hitfinding" algorithm, implemented in Cheetah (Barty et al., 2014) package.

The pre-processing and data reduction step involved subtraction of dark current (or, detector electronic noise), masking of dead, hot, cold, and saturated pixels on the detector. The detector panel edges were also masked. The local background correction was applied to each and every pixel on the detector during the pre-processing step in Cheetah (Barty et al., 2014). An example PSII diffraction raw-frame before background correction is provided in Fig. 3.8 (Left-side). In order to identify a diffraction pattern as crystal hit, specific criteria were optimized along with local background correction. The same pattern shown in Fig. 3.8 (Left-side) as well as in Fig. 3.8 (Right-side) to represent before and after background correction respectively. First, diffraction pattern must contain at least 25 peaks or more to be considered for analysis. The criterion to be a peak was to have intensity above the threshold value of 400 analog-to-digital unit (ADU). Then, a signal-to-noise (SNR) value of 25 was also applied as threshold. Once, a peak satisfies the criteria of threshold intensities, a rectangular box of neighboring 8 pixels was

considered to estimate local background noise. Then the average background intensity, estimated within the box, was then subtracted from the actual peak intensity. Then, a variance of intensity was also estimated for the noise level, caused by Poisson shot-noise and background scattering noise. The signal-to-noise value is the ratio of background corrected peak intensity to the variance of background intensities.
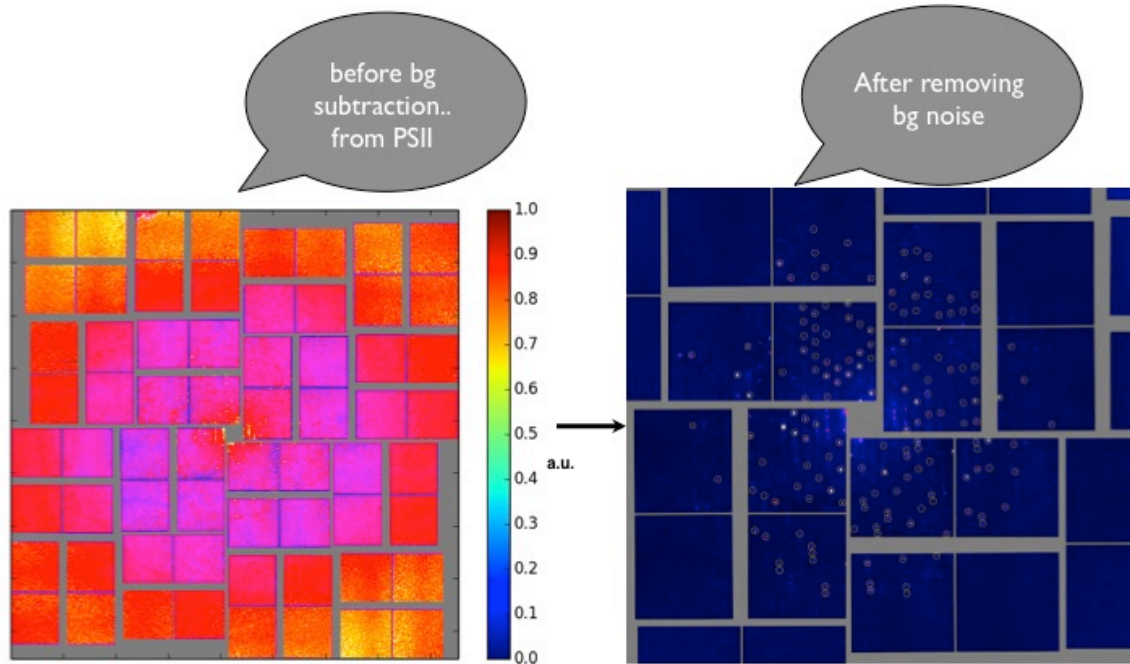
Figure 3.8: Pre-processing of PSII Diffraction Patterns. Left-side image shows the diffraction raw-frame before background (bg) correction with a color-scale bar in arbitrary unit (a.u). Right-side image shows the same diffraction pattern after background correction and after identifying diffraction spots from PSII crystals.

Thus, after pre-processing every raw-diffraction frame, the entire data set combining alternating frames from dark/light states, was sorted into two sets – dark (S_1) state and double flash (putative S_3) state based on the signals from photodiode and from the video camera inside the experimental chamber, which had been encoded to each frame. Thus, a data set of 71,628 patterns of potential "crystal hits" of PSII crystals in dark (S_1) state and

a data set of 63,363 patterns from potential PSII crystals in double flash (putative S_3) state were obtained. It is noted that dark state patterns from the alternating dark/light runs and from runs, where data were collected from only dark state, were merged together as there was no difference in unit cell parameters (See Table 3.1b). Example diffraction patterns from the dark (S_1) state and from the double-flash state (putative-S_3) are provided in Fig. 3.9(top) and 3.9(bottom), showing the diffraction limit of dark data set to 5.0-Å and that for double flash state to 5.5-Å resolutions.
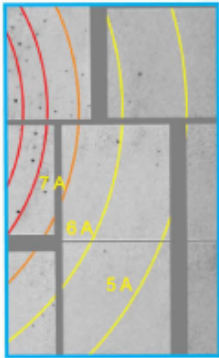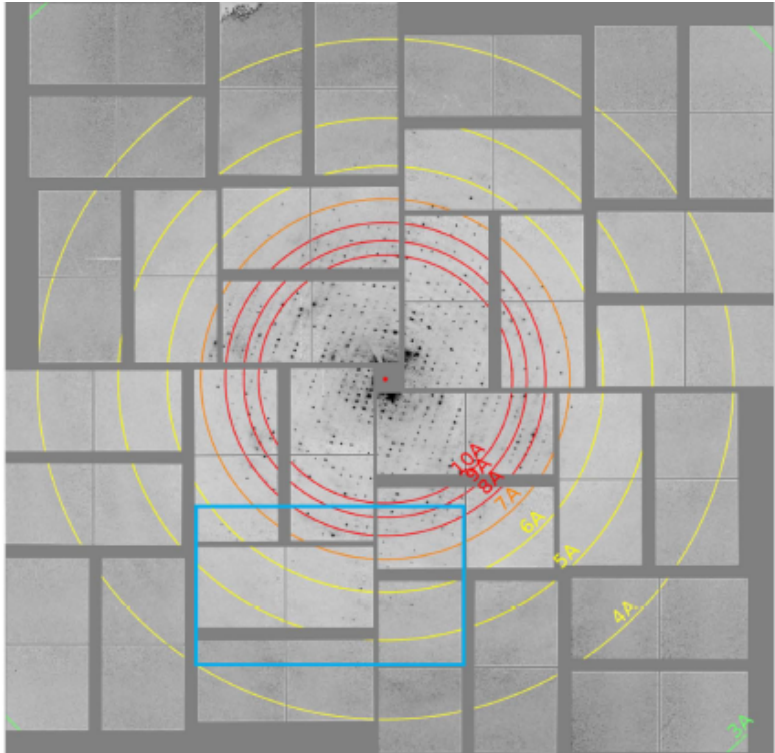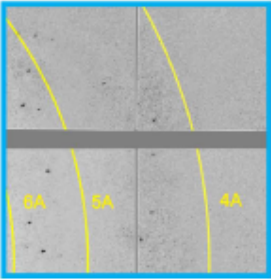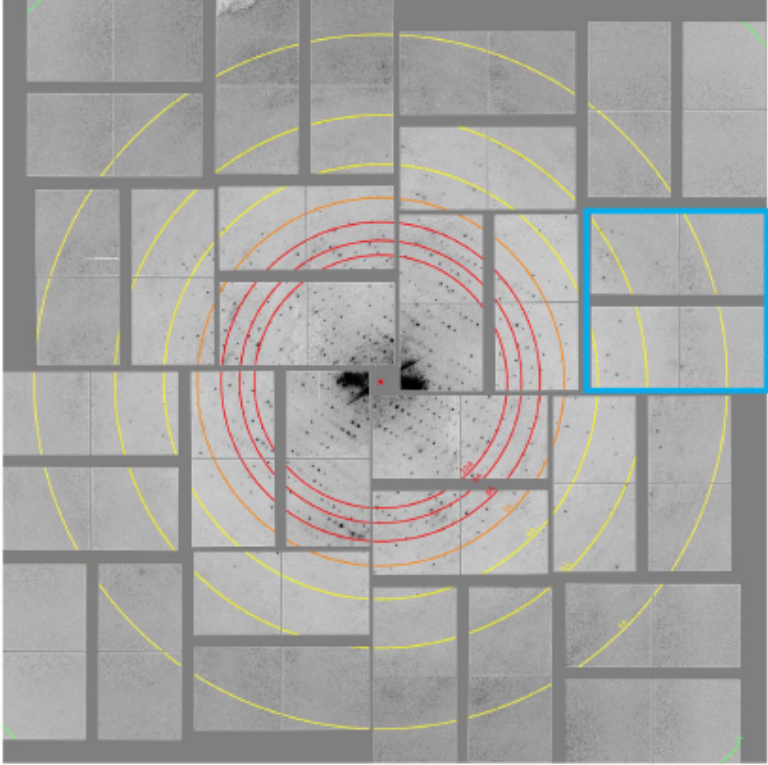
Figure 3.9: Local Background Corrected and Pre-processed Diffraction Patterns from PSII Mircocrystals. (top) diffraction pattern from the dark (S_1) state at 5-Å, (bottom) diffraction pattern from double flash (putative S_3) state showing Bragg peaks up to 5.5-Å.

All crystal hits from both data sets were passed as separate sets to CrystFEL software suite (White et al., 2012) for auto-indexing with MOSFLM (Steller et al., 1997)(Powell, 1999) using orthorhombic unit cell dimension of PSII from *Thermosynechococcus elongatus* (PDB code: 1FE1) within a chosen tolerance limit of 6%, 5%, 5% for reciprocal axes *a*, b*, c** respectively for dark (S_1) state. Similarly, crystal hits from double flash data set were also passed to CrystFEL suite to auto-index with MOSFLM using the same orthorhombic unit cell dimension within a chosen tolerance limit of 8%, 5%, 5% for reciprocal axes *a*, b*, c** respectively. From the successfully indexed patterns, crystal orientations and unit cell dimensions were obtained. Thereby, 34,554 indexed patterns from the dark state with an "indexing rate" (percentage of indexable hits out of total crystal hits obtained after pre-processing) of 48% were obtained. Similarly, 18,772 patterns from the double flash state with an indexing rate of 29% were obtained. Fig. 3.10 shows the quality of indexing of PSII diffraction pattern.
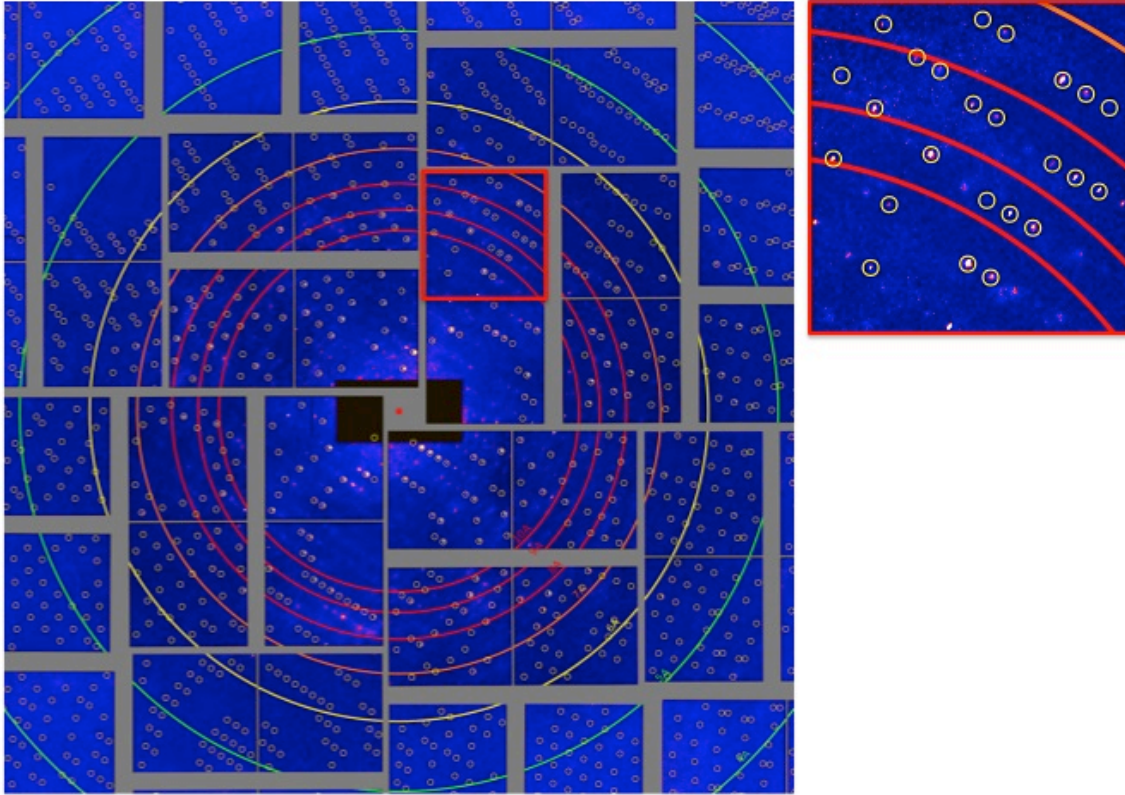
Figure 3.10: An Indexed PSII Diffraction Pattern. It shows the quality of indexing along with resolution rings and a zoom-in section (marked with red box) is also shown to notice the diffraction spots and indexing quality.

The unit cell dimensions for the dark (S_1) state were a=133.3-Å, b=226.3-Å, and c=307.1-Å in orthorhombic Bravais lattice type and that for the double flash (putative S_3) state was a=136.6-Å, b=228.1-Å, and c=308.7-Å in orthorhombic Bravais lattice type (See Fig. 3.2 and Table 3.1 for error margins and unit cell distribution). The unit cell is orthorhombic and shows very significant changes in the unit cell dimensions between the dark S_1 and double-flash data sets (A) state (see Table 3.1 and Fig. 3.2). The most pronounced change is in the dimension of the a axis which increases by 3.3-Å. This change in unit cell constants is accompanied by the slight decrease in diffraction quality (5 vs. 5.5-Å resolution) (see Tables 3.1, 3.2) and significant lowering of indexing rates.

112

This change in the unit cell dimension is fully reversed to the unit cell constants observed in the dark S_1-state when PSII crystals are excited by three laser flashes, eventually reaching the putative transient S_4-state (See Fig. 3.2 and Section 3.1 for detail interpretation).

Table 3.1: Data Statistic Comparison of Hit and Indexing Rates as well as the Unit Cell Constants from 4 Different Data Sets Collected on Photosystem II Crystals.

a

	Dark (A)	2-flash(A)	Dark (B)	3-flash(B)
Crystal hits	71,628	63,363	33,373	32,190
Indexed images	34,554	18,772	11,664	12,543
Indexing rate (%)	48	29	35	38
Unit cell constant a (Å)	133.3 ± 1.6	136.6 ± 1.5	132.3 ± 1.2	132.3 ± 1.1
Unit cell constant b (Å)	226.3 ± 2.1	228.1 ± 2.3	226.1 ± 1.7	226.1 ± 1.6
Unit cell constant c (Å)	307.1 ± 3.1	308.7 ± 3.8	307.3 ± 2.8	307.0 ± 2.7

b

	Dark (A) (only dark)	Dark (A) (dark alternating with illumination)
Crystal hits	12,087	59,541
Indexed images	5,170	29,384
Indexing rate (%)	42.77	49.35
Unit cell constant a (Å)	133.2 ± 1.9	133.3 ± 1.5
Unit cell constant b (Å)	225.8 ± 2.2	226.3 ± 2.0
Unit cell constant c (Å)	307.6 ± 3.2	307.1 ± 3.1

(a) The data sets dark (A) and double-flash (A) were collected at the CXI instrument in January 2012 and may represent the dark S_1 state and putative S_3 state of photosystem II for which data evaluation and structural changes are discussed in this work. For comparison, the statistics are shown for data sets collected on the dark state S_1 and the transient triple-flash state (that is, putative S_4 state) at the CXI instrument in June 2012. **(b)** Data statistics for dark (A) separated into runs where the laser was switched off (only dark state) and dark state images from runs where alternate dark and light states patterns were recorded. **(c)** Data statistics from this work and from Kern et al. (Kern et al., 2013). Comparison of the data evaluation statistics of the dark S_1 state and double-flash (putative S_3) state data from this work evaluated by the CrystFEL software suite (White et al., 2012) along with data from (Kern et al., 2013) on the dark S_1 state and the single excited S_2 state evaluated with the software suite cctbx.xfel (Sauter et al., 2013). The numbers in brackets refer to

c

	Kern et. al (2013)[18]		This work	
	S_1	S_2	Dark (S_1)	2-flash (putative S_3)
Resolution range (Å)	82.9 - 5.7 (5.9 - 5.7)	83.0 - 5.9 (6.1 - 5.9)	100.6 - 5.0 (5.11 - 5.0)	102.3 - 5.5 (5.66 - 5.5)
Total reflections	1,475,630 (8,036)	564,722 (8,662)	28,679,554 (1,679,683)	12,476,013 (1,018,721)
Unique reflections	27,220 (2,290)	24,671 (2,143)	40,946 (2,710)	32,066 (2,651)
Multiplicity	54.2 (3.5)	22.9 (4.0)	700.35 (618.0)	388.55 (381.1)
$CC_{1/2}$	0.802 (0.343)	0.661 (0.376)	0.914 (0.740)	0.877 (0.635)
R_{work}	0.276 (0.360)	0.284 (0.306)	0.260 (0.350)	0.280 (0.382)
R_{free}	0.315 (0.350)	0.317 (0.387)	0.262 (0.343)	0.290 (0.347)

Each diffraction pattern represents very thin slice, cutting through the Ewald sphere, leading to 'partial' Bragg reflection (White et al., 2014)(Kirian et al., 2011). Hence, a very high multiplicity (i.e., each reflection should be measured enough number of times) is needed in order to obtain 'full' Bragg reflection and complete structure factor. The dark (S_1) and double-flash (putative S_3) data sets that were used for structure factor determination consist of 34,554 and 18,772 indexed patterns, respectively. Our data sets were merged separately in three dimensions and the structure factors were extracted separately from dark and double-flash data sets of PSII protein using the Monte Carlo method (Kirian et al., 2011)(White et al., 2012) which integrates the snapshots partial reflections from randomly oriented crystals of varying size and shape. Average multiplicities are 684 and 373 in all resolution shells from 19.20-Å to 4.03-Å of the dark (S_1) and double-flash states, respectively (see Tables 3.2a, b).

Table 3.2: Statistics of the Femtosecond Crystallography X-ray Diffraction Data Sets. Statistics of the femtosecond crystallography X-ray diffraction data set of (a), the dark (S_1) state by resolution bins and (b), the double-flash (putative S_3) state by resolution bins.

a

Resolution (Å)	observed unique reflections	Measurements	Completeness (%)	Multiplicity	$I/\sigma(I)$	$CC_{1/2}$	Reflections (used for $CC_{1/2}$)
19.20	4,255	3,081,618	99.63	724.2	23.12	0.990	4,236
9.60	4,038	3,048,418	99.48	754.9	16.69	0.991	4,017
8.03	3,980	3,064,098	99.20	769.9	12.26	0.989	3,926
7.17	3,928	2,847,009	98.35	724.8	8.04	0.977	3,783
6.58	3,826	2,608,679	96.28	681.8	5.67	0.957	3,552
6.16	3,808	2,497,532	95.70	655.9	4.31	0.936	3,446
5.82	3,682	2,552,117	92.91	693.1	3.59	0.903	3,200
5.55	3,663	2,548,799	92.73	695.8	3.21	0.890	3,105
5.32	3,543	2,489,203	90.04	702.6	2.75	0.853	2,822
5.13	3,426	2,241,076	86.95	654.1	2.41	0.826	2,664
4.96	3,555	2,193,921	90.46	617.1	2.29	0.740	2,767
4.81	3,518	2,241,551	89.84	637.2	2.10	0.695	2,655
4.68	3,448	2,246,609	88.23	651.6	1.89	0.392	2,549
4.56	3,439	2,201,217	86.95	640.1	1.77	0.463	2,442
4.45	3,316	2,176,673	84.85	656.4	1.69	0.445	2,327
4.36	3,166	2,097,792	81.87	662.6	1.52	0.216	2,078
4.27	3,116	2,111,675	78.86	677.7	1.47	0.135	1,963
4.18	2,995	2,097,219	77.03	700.2	1.40	0.070	1,841
4.11	2,850	2,009,233	73.00	705.0	1.27	0.062	1,607
4.03	2,930	2,032,330	74.99	693.6	1.32	0.068	1,721
Overall	70,482	48,386,769	88.85	684.93	5.42	0.994	56,691

b

Resolution (Å)	observed unique reflections	Measurements	Completeness (%)	Multiplicity	$I/\sigma(I)$	$CC_{1/2}$	Reflections (used for $CC_{1/2}$)
19.20	4,395	1,724,037	99.30	392.3	16.30	0.974	4,361
9.60	4,157	1,728,443	98.79	415.8	11.84	0.983	4.103
8.03	4,085	1,718,604	97.80	420.7	8.71	0.977	3,960
7.17	3,916	1,546,207	94.73	394.8	5.72	0.963	3,628
6.58	3,747	1,389,049	90.73	370.7	3.93	0.926	3,242
6.16	3,671	1,308,516	88.97	356.4	2.91	0.860	3,001
5.82	3,463	1,306,601	84.28	377.3	2.35	0.789	2,626
5.55	3,365	1,280,059	82.27	380.4	2.06	0.785	2,439
5.32	3,193	1,221,532	77.90	382.6	1.79	0.635	2,202
5.13	2,946	1,046,614	72.05	355.3	1.60	0.474	1,893
4.96	3,107	1,040,022	76.28	334.7	1.51	0.450	1,935
4.81	3,046	1,056,445	74.84	346.8	1.45	0.308	1,847
4.68	3,078	1,087,015	75.20	353.2	1.40	0.167	1,890
4.56	2,968	1,034,930	73.16	348.7	1.34	0.159	1,764
4.45	2,859	1,021,562	70.26	357.3	1.30	0.086	1,649
4.36	2,771	1,003,272	68.20	362.1	1.22	0.007	1,510
4.27	2,652	979,252	65.37	369.3	1.17	0.042	1,394
4.18	2,628	1,001,005	65.15	380.9	1.18	0.072	1,386
4.11	2,447	937,143	59.95	383.0	1.13	0.010	1,294
4.03	2,579	976,398	64.00	378.6	1.14	0.037	1,350
Overall	65,073	24,496,706	78.96	373.04	4.09	0.988	47,474

A comparison of the data statistics of our work with that of Kern *et al.* (Kern et al.,

2013) is shown in Table 3.1c. Our data sets show significantly higher multiplicity of the data and better correlation coefficients ($CC_{1/2}$) when compared to Kern *et al*., which are indicative of the quality of the merged reflections. From Table 3.1c, it is noticed that Kern et al., had a very low multiplicity of 3.5 in the highest resolution shell (5.9 to 5.7-Å), whereas ours data has a very high multiplicity of 618 in the highest resolution shell (5.1 -5.0-Å). A multiplicity of 3.5 ((Kern et al., 2013), Table 3.1c) is considered as 'unacceptable' poor data quality in the field of SFX. Kern et al., data also had a very low $CC_{1/2}$ value of 0.343 in the highest resolution shell (5.9 to 5.7-Å), whereas our data shows a high $CC_{1/2}$ value of 0.740 even in the highest resolution shell (5.1 to 5.0-Å). On contrary to the poor multiplicity and $CC_{1/2}$ value (SFX data quality indicators), the R-factors of Kern et al., work are reasonably good. This can be caused due to high-resolution model bias introduced during molecular replacement step, which can artificially reduce the R-factors, resulting in a decent model.

Several matrices have been calculated to evaluate the PSII data quality. R_{split} is a metric, estimating internal consistency of the data, defined in White et al., 2012, as

$$R_{split} = \sqrt{2} \frac{\sum_{hkl} |I_{even} - I_{odd}|}{\sum_{hkl} |I_{even} + I_{odd}|},$$ .... (Eq. 3.2.3.1)

To calculate R_{split}, the each data set is split into two random halves, followed by merging each half separately in Monte Carlo approach. Then the intensity values of every reflection will be compared between those two halves, leading to R_{split} calculation. In an ideal case, $R_{split} = 0$, implies that all reflections measured in that data set from various crystals are measured perfectly. This means that the same *hkl* reflection in each data set have identical intensities. In reality, this ideal case does not occur in real crystals.

117

Thereby, R_{split} estimates the overall error in intensities, measured for each reflection. Since, SFX method measures mostly 'partial' Bragg reflections, where the fraction or 'partiality' is unknown. If each reflection is measured many number of times (i.e., high multiplicity), R_{split} value should be very low. This would imply that error in measuring Bragg intensities is very low and thereby the data quality is good. R_{split} was calculated for the dark and double-pumped data sets separately. The overall R_{split} values over all resolution shells were 0.07 for the dark and 0.09 for the double-pumped data sets, which show that the multiplicity of our data sets are sufficient and thereby of good quality. Fig. 3.11a shows the change in R_{split} values for the dark and the double-flash states data as a function of resolution bins. R_{split} values were also calculated with varying number of indexed patterns, and plotted as a function of resolution for the dark (Fig. 3.11b) and the double-flash (Fig. 3.11c) data sets respectively. It is observed that R_{split} value decreases with increasing number of indexed patterns, implying the increase in multiplicity of the data, i.e., high Monte-Carlo convergence of the data. These plots were useful in determining the resolution for the data set. The protocol is to decide the resolution at the point, where R_{split} value raises sharply, implying increasing error in measured intensities. In case of PSII data, R_{split} value changes very gently without indicating a sharp rise. Thereby, a feedback from other metrics (such as $CC_{1/2}$) and finally refinement statistics (Table 3.3) were used to decide on resolution.
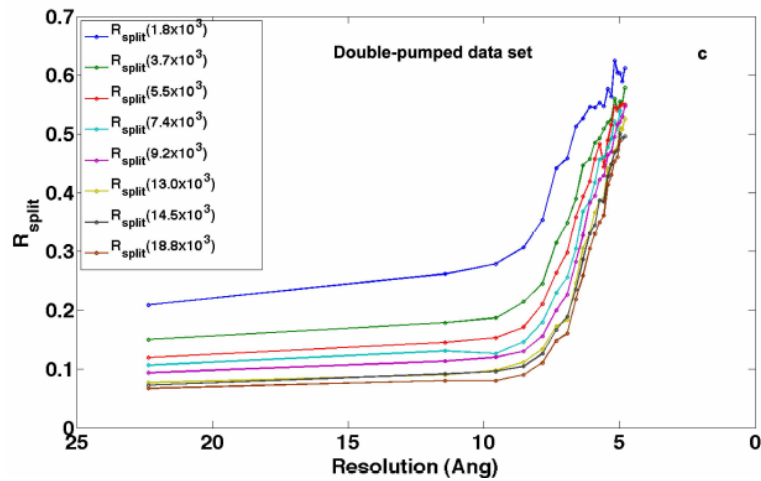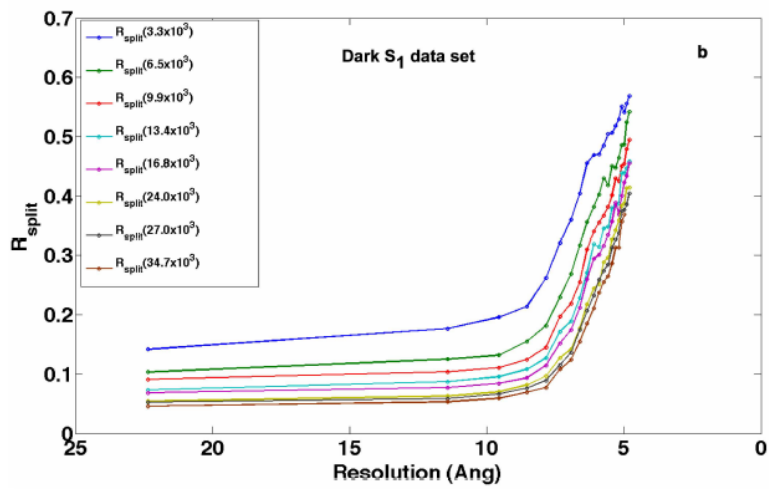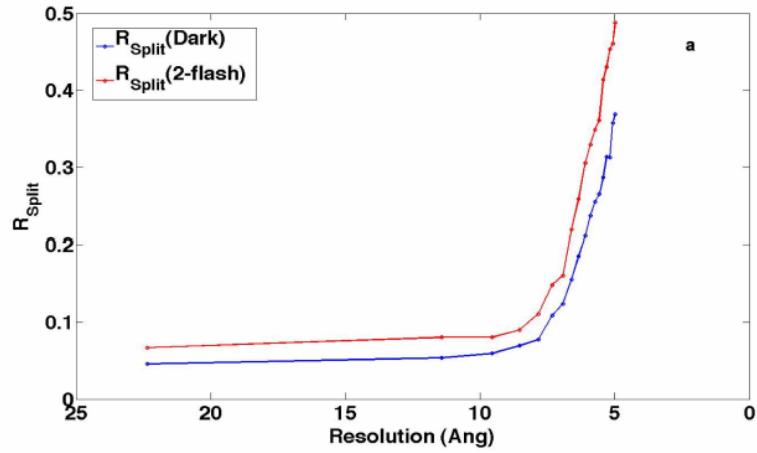
Figure 3.11: R_{split} as a Function of Resolution Bins and Number of Indexed Patterns. (a), R_{split} as a function of the resolution shell (in total 20 bins) for dark state data (blue) and double-pumped state data (red). (b), R_{split} as a function of resolution bins for dark S1 state, R_{split} decreases indicating better data quality with increase in number of indexed patterns from 3,300 to 34,000 images. (c), R_{split} as a function of resolution bins for the dark and double-flash states, the R_{split} decreases indicating better quality with increase in number of indexed patterns from 1,800 to 18,800 images.

There is another quality metric that provides useful insights, called $CC_{1/2}$ or CC^* are based on Pearson correlation term and thoroughly described in (Karplus and Diederichs, 2012).

$$CC_{1/2} = \frac{1}{(N-1)} \frac{\sum_{hkl}(I_{even} - <I_{even}>).(I_{odd} - <I_{odd}>)}{\sigma(I_{even}).\sigma(I_{odd})},$$ .... (Eq. 3.2.3.2)

$C_{1/2}$ measures the Pearson cross-correlation between the intensity values of the same *hkl* reflection from two different diffraction patterns. Higher the $CC_{1/2}$ value, higher is the consistency between different measurements of the same reflection. In an ideal case, $CC_{1/2}$ value would be 1.0, implying there was no systematic error while measuring the same *hkl* reflection from different crystals in varying orientations. $CC_{1/2}$ metric is considered to be more robust and often found to provide better estimation of resolution cut-off for the data set.

The $CC_{1/2}$ value was calculated for each HKL reflection to estimate the correlation of intensities between two halves of each data set. Figure 3.12 shows changes in $CC_{1/2}$ as a function of resolution shells for the dark and double-flash data sets. Here, in case of PSII data, overall $CC_{1/2}$ of 0.994 for dark (S_1) state and that of 0.988 for double-flash (putative S_3) state. Although the $CC_{1/2}$ value at the highest resolution shell (5.13-4.96 Å) for dark (S_1) state is quite high (0.745) (See Table 3.2a,b) and that for the double-flash state is

120

0.635 (highest resolution shell 5.55 – 5.32 Å), we set the resolution at 5.0 and 5.5-Å respectively as the final refinement R-factor did not improve.
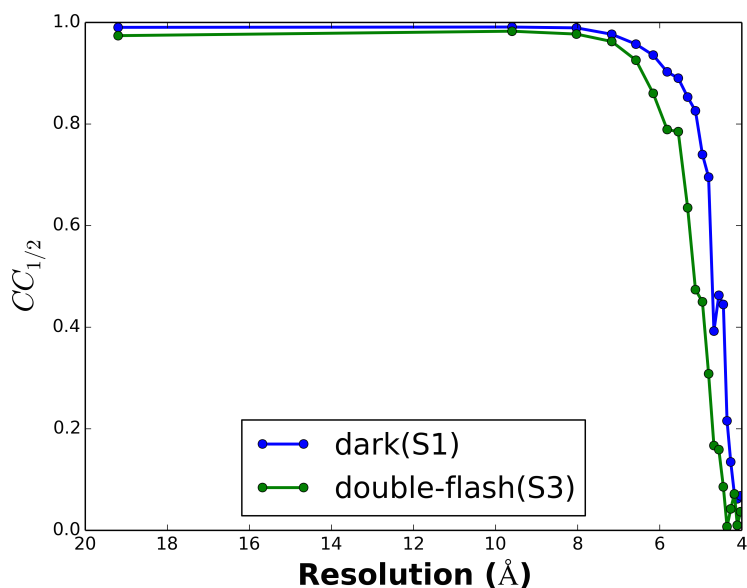
Figure 3.12: $CC_{1/2}$ as a Function of Resolution Bins. $CC_{1/2}$ plots are for the dark (blue curve) and the double-flash (green curve) data sets.

3.2.4. Molecular Replacement

Molecular Replacement (Read, 2003) was carried out by the Phaser program, which is a part of the CCP4 suite (McCoy, 2007) using the PSII X-ray structure at 1.9 Å resolution of Umena and coworkers (PDBID: 3ARC) (Umena et al., 2011) as the search model, which was modified by removing waters, detergents, lipids, and alternative conformers of the amino acids. (Galstyan et al., 2012) reported that one monomer (denoted by capital letters) in PSII dimer is better defined and more complete than the 2nd monomer (denoted by small letters). In addition, there is also a non-crystallographic symmetry between these two monomers, making them structurally different. So, we used monomer 1 of the PSII dimer (where the monomer 1 subunits are labeled with capital

letters and small letters are used for monomer 2 in 3ARC model) as search model for molecular replacement to solve the phase problem by using the program Phaser (version 2.5.3). After we found monomer 1, we repeated the search for monomer 2.

3.2.5. Structure Refinement

The dark (S_1) and double-flash (putative S_3) states structures were refined by *phenix.refine* (Afonine et al., 2012) using segmented rigid body refinement, where each C-alpha chain of each protein subunit was considered as a rigid body. The cofactors were also considered as rigid bodies. During the rigid body refinement, we considered only translational refinement and not rotational refinement of the rigid entities. So, the RMS bond angle was not refined. We used the original B factors from the 3ARC model because B-factor refinement is not useful at the given resolution of 5.0 Å. After three refinement cycles, R-factors for the dark state at 5.0-Å of R_{work} of 0.260 (R_{free} = 0.262) were observed. The R-factors for the double-flash state at 5.5-Å are R_{work} of 0.280 (R_{free} = 0.290). Refinement statistics are shown in Table 3.3.

Table 3.3: The Refinement Statistics for the Dark (S_1) and the Double-Flash (putative S_3) states. The data were collected during January 2012 TR-SFX experiments at CXI, LCLS.

	Dark data set	Double-flash data set
Wavelength (Å)	2.05	2.05
Resolution range (Å)	100.6 - 5.0	102.3 - 5.5
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit Cell length (Å)	133.3 ± 1.6, 226.3 ± 2.1, 307.1 ± 3.1	136.6 ± 1.5, 228.1 ± 2.3, 308.7 ± 3.8
Total reflections	28,679,554 (1,679,683)	12,476,013 (1,018,721)
Unique reflections	40,946 (2,710)	32,066 (2,651)
Multiplicity	700.35 (618.0)	388.55 (381.1)
Completeness (%)	99.98 (100)	99.88 (100)
Mean $I/\sigma(I)$	10.65 (2.1)	8.03 (1.75)
$CC_{1/2}$[‡]	0.914 (0.740)	0.877 (0.635)
R_{split}	0.07 (0.37)	0.09 (0.49)
R_{work}	0.260 (0.3502)	0.280 (0.3820)
R_{free}	0.262 (0.3434)	0.290 (0.3477)
RMS* (bonds) Å	0.039	0.039
RMS* (angles) deg	3.029	3.029
Number of atoms	49,817	49,817
Protein residues	5,214	5,214
Ramachandran[*] favored (%)	97.7	97.7
Ramachandran[*] outliers (%)	0.2	0.2
Clashscore (Molprobity)	5.5	5.8
Mean B-factor[#] ($Å^2$)	33.7	33.7

3.2.6. Calculation of Simulated Annealing Omit Maps

The composite omit maps, defined by Bhat (Bhat, 1988), were calculated using 'Omit' program (Vellieux and Dijkstra, 1997) (CCP4 package), where the Mn-cluster was removed entirely. It is also noted that for the calculation of omit maps, the model

used was right after MR step, i.e., prior to any type of refinement to avoid model bias. The principle of omit map calculation is to calculate first a map which is based on the model without the 'omitted' region (i.e., $F_{(model - omit)}$). Then, the structure factor amplitudes from the data (i.e., F_{obs}) is used along with $F_{(model - omit)}$ to generate an electron density map for the omitted region. If F_{obs} contains structure factors information from that particular 'omitted' region, the electron density for that region should be found. It should be noted that the omit map represents an electron density map, which minimizes the effect of model bias. Thus, the omit maps are generally more noisy than general $2F_o - F_c$ maps. There is a research article by Thomas Terwilliger (Terwilliger et al., 2008), where the proper uses of omit maps are described. For this work, the unit cell dimensions of the dark (S_1) state and double-flash (putative S_3) state are different, leading to a classic problem of non-isomorphism. Therefore, ($F_{light} - F_{dark}$) difference electron density maps could not be used or calculated to determine the conformational differences between the dark and double-flash states. The reason behind this is that the sampling of HKL reflections within the unit cell would be different for two states due to difference in unit cell sizes, requiring scaling of the structure factors in reciprocal space, which is a non-trivial problem. Therefore, along with conventional omit maps (done by CCP4 program), more robust simulated annealing composite omit maps (SA-omit map) were calculated separately for both states. Then they were superimposed to compare the differences between the two states. Superimposition was also non-trivial under the condition of non-isomorphism.

124

For the superimposition of two omit maps (calculated by CCP4 program) from the dark and double-flash states, the following steps were carried out. First, using the omit maps as inputs, a new set coordinate files (PDB files) were generated from the MR solutions of dark and double-flash states separately, so that each of those two omit maps fits the model, using the Molrep program in CCP4 suite. Second, the new set of coordinate files, generated by Molrep program, for the dark and double-flash states, were opened and superimposed in Coot (Emsley et al., 2010). Later, the superimposed coordinate file was saved. During superimposition, the double-flash state coordinate file was considered as the moving object and the dark state coordinate file as the fixed object. During this process, Coot provided Euler angles and translational coordinates (x,y,z values) as a result of the superimposition. Third, using these Euler angles and translational coordinates as rotational operator with opposite sign, the double-flash state omit map wa rotated using MAPMASK program in the CCP4 package. The electron density maps were in two different coordinate frames due to non-isomorphism. Thereby, during the overlaying of maps, the rotated double-flash omit map (output of MAPMASK program) was moved over the superimposed coordinate file. The same procedure was also followed for the dark state omit map, overlaying onto the superimposed coordinate file, using MAPMASK program in the CCP4 package. Example 'omit' electron density maps are shown in Fig. 3.3, 3,4, and 3.5.

In addition to this omit map calculation, a much more robust feature, called simulated annealing omit maps (SA-omit map) were calculated. In case of SA-omit map calculation, the calculation of omit map is in principle similar to the one described earlier

125

in this section. The difference lies in the 'simulated annealing' approach. It means during the process of SA omit maps calculation, the model was virtually heated up to 5000K temperature, at which atoms would be displaced so much that almost no model bias can be present. Then, the model would be slowly cooled down at a step of 100K to room temperature, leading to re-build the model along with the omit map calculation. Similar to the composite omit map calculation, SA omit map was also calculated using the MR model without the OEC or Mn-cluster before applying any type of refinement. (Terwilliger et al., 2008) mentioned that it is important to calculate Omit maps only on MR solutions of structures to avoid model bias. So, SA-omit maps for both dark and double-flash states were calculated with a start temperature of 5000K using *phenix.autobuild* program (Terwilliger et al., 2008) and later superimposed them using *phenix.superimpose_maps* program. Fig. 3.6 and 3.7a-b represent the SA-omit map for OEC region from the dark and double-flash states respectively.

3.2.7. Control Tests Based on SA-Omit Maps to Cross-Validate the Results

It is noted that the electron density and its shape changes can be assigned with more confidence for the heavy metal cluster than the electron density changes in the protein environment at the moderate resolution of 5.0-Å. Low-resolution model bias is a classic problem in crystallography. In this crystallographic work, therefore, this model-bias issue was addressed critically. So, it was important to ensure that the 'low-resolution' model-bias was completely removed from the SA omit maps and not trapped in the local minima

126

of the conformational space. Thereby, several control tests were also performed. They are explained below.

Test 1: Both data sets – dark (i.e., S_1) and double-flash (i.e., putative S_3) states were randomly split into two halves. Then, SA omit maps were calculated from each of those data sets. It means that we produced four SA-omit maps – S_1(split1), S_1(split2), and S_3(split1), S_3(split2). Then, SA-omit maps from S_1(split1) and S_1(split2) were compared using superimposition to investigate if those maps show similar shape changes as shown in Fig 3.7 for S_1 and putative S_3 SA-omit maps. The same comparison was also performed on S_3(split1), S_3(split2). The comparison, shown in Fig. 3.13, concludes that SA-omit maps from S_1(split1) and S_1(split2) are identical in shape. The same is true for S_3(split1), S_3(split2). Thereby, this test proves two facts – one, the SA omit electron densities between two random halves of each data set are internally consistent. Second, the difference in the SA-omit maps shapes, shown in Fig. 3.7 could not be due to low-resolution noise.
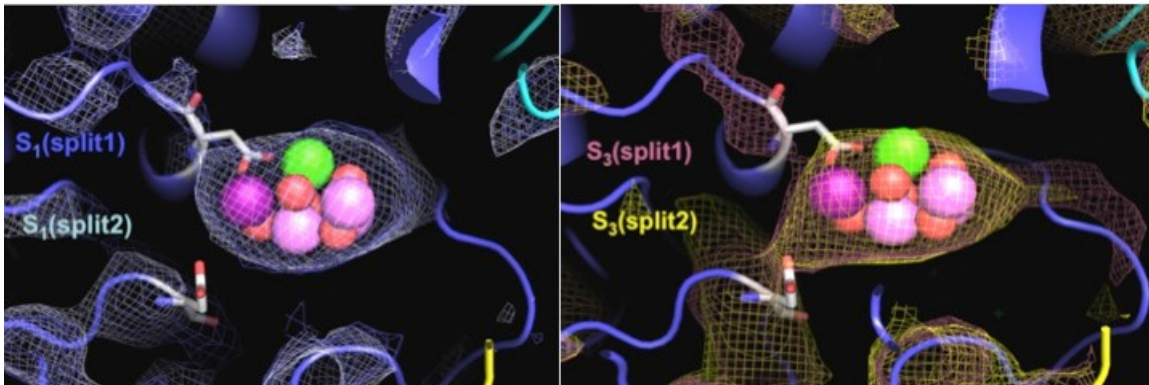
Figure 3.13: Comparison of SA-Omit Maps Between Two Randomly Split Halves at 1.5-σ. (a) shows the comparison of two splits from the S_1 state. S_1(split1) is in marine-blue and S_1(split2) is colored in light-blue. (b) shows the comparison of two splits from the putative S_3 state. S_3(split1) is colored in salmon-red and S_3(split2) is in yellow.

Test 2: After receiving positive feedback from test 1, another robust test was performed by using a different start temperature for simulated annealing step during SA omit map calculation. For this purpose, a different start temperature of 2000-K (instead of 5000-K) for simulated annealing step was chosen. 2500-K as start temperature was also tested. SA omit maps for both data sets – dark (S_1) and double-flash (putative-S_3) were calculated using 2000-K as start temperature. Then, those SA omit maps were superimposed for comparing with those from 5000-K-start temperature for the dark and double-flash data sets respectively. The results are shown in Fig. 3.14a,b. In Fig. 3.14c, dark state SA omit map (at 2000-K) was superimposed over double-flash state SA omit map (at 2000-K). In Fig 3.14d, the same superimposition was done but using SA omit maps at 5000-K. All four figures in Fig. 3.14 show that the movement and shapes of SA omit map electron densities were all consistent and the differences in SA omit map shapes, observed in Fig 3.7 between dark and double-flash states is quite robust with

respect to varying SA omit map temperature. This test also consolidates that the observations made in Fig. 3.7 are not suffering from model bias.
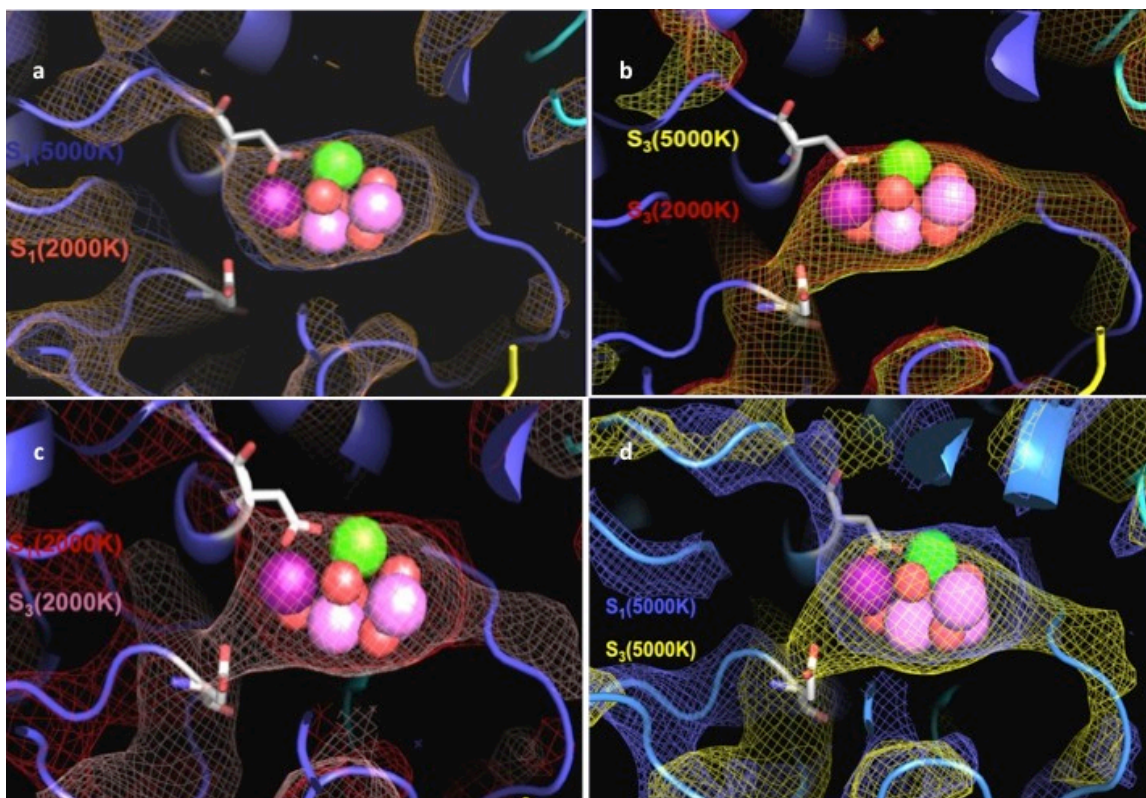
Figure 3.14: Comparison of SA-Omit Maps, Calculated using Different Start Temperatures at 1.5-σ. (a) shows the superimposition of SA omit maps from S_1 state with 2000-K start temperature (in red) and with 5000-K start temperature (in blue). (b) shows the superimposition of SA omit maps from putative S_3 state with 2000-K start temperature (in dense red) and with 5000-K start temperature (in yellow). (c) shows the superimposition of SA omit maps at 2000-K start temperature from S_1 (in red) and putative S_3 (in salmon-red) states. (d) shows the superimposition of SA omit maps at 5000-K start temperature from S_1 (in blue) and putative S_3 (in yellow) states.

Test 3: The previously mentioned two control tests (page no. 119-122) established the fact that the difference in the shapes of SA omit electron density maps around the OEC and neighboring protein environment are neither suffering from low-resolution noise nor model-bias, thereby very robust to those two factors. However, there was

another factor to consider and that was non-isomorphism. This PSII work was a classic problem of non-isomorphism, where the unit cell dimensions of the dark and putative S_3 states are different, preventing to use ($F_{light} - F_{dark}$) difference electron density maps. Since, PSII data was not isomorphous as well as two data sets had slightly different resolution cut-offs (5-Å vs. 5.5-Å), so these two factors could lead to difference in sampling HKL reflections in reciprocal space. Particularly, when the electron density around heavy metal cluster is considered at such low-resolution, different sampling of HKL reflections in two data sets can cause error due to Fourier series truncation. Thereby, it was important to compare the intensities or the structure factor amplitudes from the exactly identical set of HKL reflections. In order to address the issue of non-isomorphism, a new control test of calculating SA omit maps from both data sets using exactly identical HKL reflections was considered. Thereby, a python script was written to match and extract only those HKL reflections, which were identical in both data sets. Then, SA omit maps for both data sets were calculated using those identical set of HKL reflections. The superimposition of such SA-omit maps is shown in Fig 3.15. The comparison of the SA omit maps calculated with the identical set of structure factors supports and reveals the same electron density shape changes. Thereby, the observation or conformational change seen in Fig 3.7 could be confirmed.

In Fig 3.15 (Top), a table, detailing the data statistics for both data sets with identical set of HKL reflections, is shown. It is noted that use of identical set of HKL reflections correspond to a slightly different resolution cut-off for the two data sets because the unit cell dimensions of the dark state data set were smaller than that of the double-flash data

set. The resolution range was defined by the HKL reflection of lowest order (HKL of (0,0,4) => 76.7-Å for the dark data set and 77.2-Å for the double-flash data set) and the reflection of highest order (HKL of (24,10,5) => 5.37-Å and 5.5-Å for the dark and double-flash datasets respectively). The resolution cut-off is thereby higher for the dark data set (5.37-Å) than for the double flash data set (5.5-Å), while the number of unique reflections is identical, leading to a lower overall completeness for the dark dataset (96.9%) compared to the double flash data set (99.9 %). In Fig 3.15(Bottom), the SA omit maps with identical set of HKL reflections for both data sets show the same electron density changes that were observed in Fig 3.7.

Table 3.4: Statistics at 1.5-σ using Identical Sets of HKL Reflections from Dark and Double-Flash Data Sets.

	Dark data-set	Double-flash data-set
Wavelength (Å)	2.05	2.05
Lowest resolution reflection	(0,0,4) reflection => 76.7-Å	(0,0,4) reflection => 77.2-Å
Highest resolution reflection	(24,10,5) reflection => 5.37-Å	(24,10,5) reflection => 5.50-Å
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell dimensions (Å)	133.3 ± 1.6, 226.3 ± 2.1, 307.1 ± 3.1	136.6 ± 1.5, 228.1 ± 2.3, 308.7 ± 3.8
Unique Reflections	32,104	32,104
Completeness (%)	96.9	99.9
Mean I/σ(I)	9.92 (1.96)	6.12 (1.8)
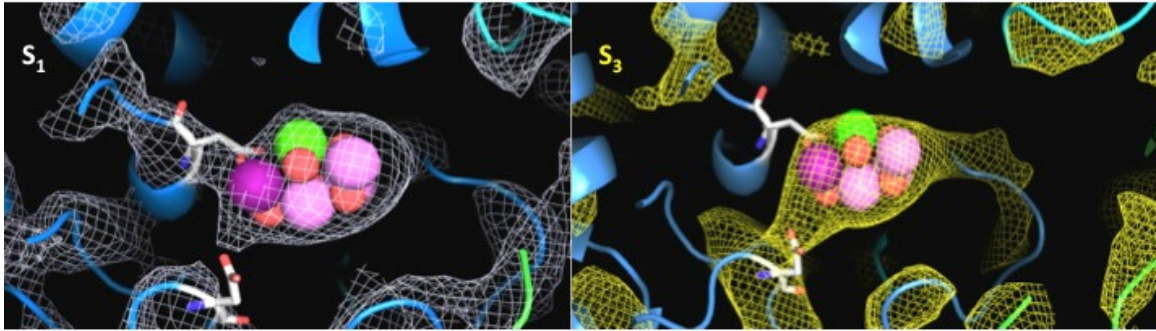
Figure 3.15: SA-Omit Maps at 1.5-σ using Identical Sets of HKL Reflections. It shows the dark state (in gray left) and double-flash state (in yellow, right).

4. TIME-RESOLVED SFX ON THE TRIPLE-FLASH (S$_4$)-STATE OF PSII CRYSTALS

This chapter describes the time-resolved SFX experiment on Photosystem II (PSII) crystals to investigate the transient higher S-state (i.e., S$_4$ state) in the Kok cycle (See Fig. 3.1a) using triple-flash excitation. This work was performed on June 2012 at the CXI instrument of the LCLS at SLAC Nat'l laboratory. This is a continuation of the work discussed in the previous chapter. In previous chapter, the time-resolved SFX experiment on the double excited putative S$_3$ state of PSII have been discussed along with its structural insights and observed conformational changes. The previously discussed Kok cycle (See Chapter 3) involves four charge-separation events, of which the event leading to transient S$_4$ state would be investigated in this work, using time-resolved SFX method to obtain the structural insights as well as visualizing the relevant conformational changes, important to elucidate the water splitting mechanism. This chapter will describe the experimental changes applied to the scheme used in previously discussed double-flash experiment and following results with discussion of the structural interpretation.

4.1. CXI Instrument and Experimental Set-up for the Study of Triple-Flash State of PSII Crystals

Before discussing the experimental details, it is important to revisit the S-states kinetics briefly based on the Kok cycle. PSII upon absorption of photon from the sunlight undergoes a complicated cycle of multiple charge-separation events. As it progress through the cycle, PSII catalyzes the oxidation of water to oxygen via extracting four

electrons from the two water molecules through Oxygen Evolving Complex (OEC). This part is extensively discussed in Chapter 3. The time constants or the half-life time of the different S-states are in micro to millisecond time range (S_1->S_2 70-μs; S_2->S_3 190-μs; S_3->S_4->S_0 1.5-ms; S_0 -> S_1 30-μs) (Dekker et al., 1984)(Siegbahn, 2008)(Yano and Yachandra, 2008). From the scheme of time constants, it is evident that the S_4 state is transient, i.e., there is no study showing the trapping of S_4 state and related S_4 state kinetics. During the June 2012 time-resolved SFX experiment, we aimed at probing S_4 state of PSII. We originally proposed a three lasers excitation scheme (See, Fig. 4.1) to excite the PSII crystals to the transient S_4 state. Fig. 4.1 shows the three different laser-schemes, where the time-delay between 3rd and X-ray 'probe' pulse would be varied with three different delays (10-μs, 250-μs, and 1.5-ms). The first scheme denoted by 'a', shows the time-delay between 1st and 2nd flashes as 210-μs (3 x half-life $_{(S1 -> S2)}$ of 70-μs) to establish the S_2 state (See Chapter 3), time-delay between 2nd and 3rd flashes as 570-μs (3 x half-life $_{(S1 -> S2)}$ of 190-μs) to establish the S_3 state, and the time-delay between 3rd flash and X-ray 'probe' pulse is 10-μs. So, all three schemes have identical time-delays for 1st and 2nd laser excitations but differ in the time-delay between the 3rd laser excitation and X-ray probe pulse. Since, the kinetics and time-constant for the S_4 state is unknown, we wanted to explore several possible time-delays to broadly study the S_4 state kinetics.10-μs time-delay would have explained if there is any very fast conformational changes occur after the 3rd flash. 250-μs time-delay would probe a state where the oxidation of the OEC to S_4 has occurred, which precedes the O_2 evolution. At the 1.5-ms time-delay, the S_0 state is reached. However, due to limited beamtime, we could only

134

explore one of the proposed time-points and we chose laser excitation scheme 'b' in Fig. 4.1. This means, the proposed time-delay between the 3rd laser excitation and X-ray probe pulse of 250-μs was used for the experiment because it may have the highest population of the S_4 state. Since, S_0 state is relatively more stable and there are relevant studies of S_0 state (Pal et al., 2013), third laser-scheme was expected to ensure the advancement of the reaction up to S_0 state. We are recently awarded a new beamtime in October 2015, when we will explore more time-points between the 3rd flash (S_3->S_0) and X-ray pulse. Thus we can investigate the steps, involved during O-O bond formation. Unfortunately, during June, 2012 beamtime, due to experimental limitations with sample delivery and hardware at the beamline, the data collection and probing of higher S-state were performed only based on 2nd laser-scheme (i.e., scheme 'b' in Fig. 4.1).
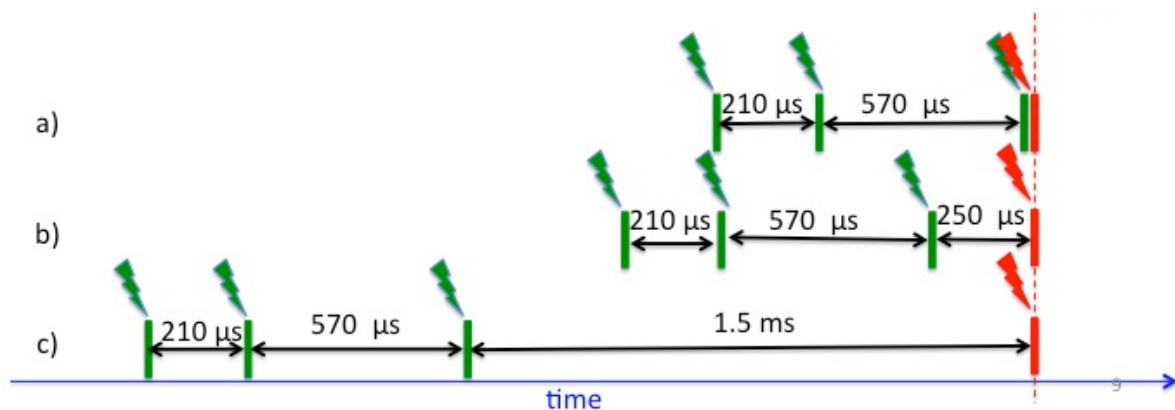
Figure 4.1: Pump-Laser Excitation Scheme for the Time-Resolved Experiment to Probe S$_4$ State of PSII. Green arrows denote 'pump' or excitation visible laser pulses (λ=527-nm) and red arrows denote the 'probe' X-ray laser pulse. Scheme 'a' probes the excited state after a few μs delay of 3rd pump laser excitation. There were three different excitation schemes proposed. Scheme 'b' probes the excited state after 250-μs of 3rd pump laser excitation. Scheme 'c' probes the excited state after 1.5-ms of 3rd pump laser excitation.

The three lasers excitation scheme requires two optical pump lasers (λ=527-nm). Optical pump laser 1 provides 'pumping' or excitation pulses one and two to advance the PSII crystals to S_3 excited state. This pump laser 1 is triggered by X-ray pulse "1". Optical pump laser 2 provides excitation pulse three to advance the PSII crystals from S_3 to higher S-states i.e., S_4 state. This 2nd pump laser is triggered by the next incoming X-ray pulse "-1". The diffraction data was collected after 250-μs time-delay to establish the higher S-state (i.e., putative S_4 state). The pump lasers operated at a frequency of 60-Hz (same as the double-flash experiment described in Chapter 3). The X-ray probe pulses were provided at a frequency of 120-Hz. Thereby, similar to the set up described in Chapter 3 and shown in Fig. 3.1b, every alternate diffraction pattern was from the 'light' or excited state (putative S_4 state) in dark/light mode. It is noted that no excitation scheme is 100% efficient. Therefore, the excited state might be a mixture of multiple states, resulting in heterogeneity. Therefore, it is called putative S_4 state or triple-flash state. The time-resolved experimental set up including the illumination volume of the optical lasers is same as described for the double-flash experiment in Chapter 3. So, readers are redirected to the relevant sections in Chapter 3 for the details.

There is another important experimental change, regarding the PSII crystal preparation. In order to reach S_4 state and 3rd charge separation event to occur, changes, regarding the quinone were necessary. As shown in Fig. 3.1a of the S-state transition scheme, each charge separation event leads to one electron to be transferred from P680* (i.e., Chlorophyll-a) to the mobile plastoquinone PQ_B in the Q_B binding pocket. After double laser excitation (i.e., once PSII reaches the putative S_3 state), two electrons will be

136

accumulated at the plastoquinone site B, converting PQ_B to PQ_B^{2-}, which will uptake two protons and leave the Q_B binding pocket as PQH_2. In native membrane, this empty binding pocket is replenished by plastoquinone from the PQ pool located at the membrane to continue the reaction to advance to the higher S-states. Therefore, during the experiments, it was necessary to provide an "artificial PQ pool" using a quinone derivative. Dr. Christopher Kupitz from our lab and Dr. Jesse Bergkamp tested different quinone derivatives for their abilities to maintain O2 evolution in PSII crystals in a steady state experiment. The best quninone derivative was PQ_{10} (Plastoquinone head group with an aliphatic tail of 10 carbon atoms) based on the steady-state oxygen evolution rate. Dr. Jesse Bergkamp from Drs. Thomas and Ana Moore's group synthesized PQ_{10} molecule and the other PQ derivatives tested. In the crystallization process, Christopher Kupitz introduced PQ_{10} as "an artificial quinone pool" such that PSII crystals could progress the reaction to reach higher S-state upon triple flash excitation. The exact crystallization protocol was mentioned elsewhere (Kupitz and Basu, 2014).

4.2. Data Processing and Evaluation of Time-Resolved SFX Data from the Triple-Flash State of PSII Crystals

The time-resolved SFX data from the triple-flash state of PSII micro-crystals were collected in alternate dark/light fashion, producing two data sets – from the dark or S_1 state and from the triple-flash or putative S_4 state. Data were collected on the CSPAD detector at the X-ray photon energy of 6-keV (λ=2.05-Å). The diffraction data was evaluated up to the resolution of 4-Å.

The raw diffraction frames (total of 50 terabytes data) were pre-processed using *Cheetah* software package (Barty et al., 2014). During this step, electronic noise of the CSPAD detector, bad, hot, dead pixels on the detector were masked. The scattering at a very low scattering angle from the liquid jet (i.e., cylindrical Fourier transform of the jet) was also masked. Based on criteria - threshold intensity, minimum permissible signal-to-noise ratio, minimum number of peaks per pattern, each of those raw diffraction frames was processed to identify them either as "crystal hit" or to reject frames which were empty or had very weakly diffracted peaks from the crystals. Thus, 33,373 diffraction patterns were identified as crystal hits from the dark or S_1 data set and 32,190 such diffraction patterns from the triple-flash or putative S_4 data set. An example diffraction pattern from the triple-flash state is shown in Fig. 4.2. It is noted that diffraction patterns were sorted into two groups – dark or S_1 and triple-flash or S_4 states based on laser information recorded by event-code reader (EVR), provided by the LCLS. Separate codes were written to perform the sorting of diffraction frames into groups based on information from the EVR and the photodiode of the video camera.
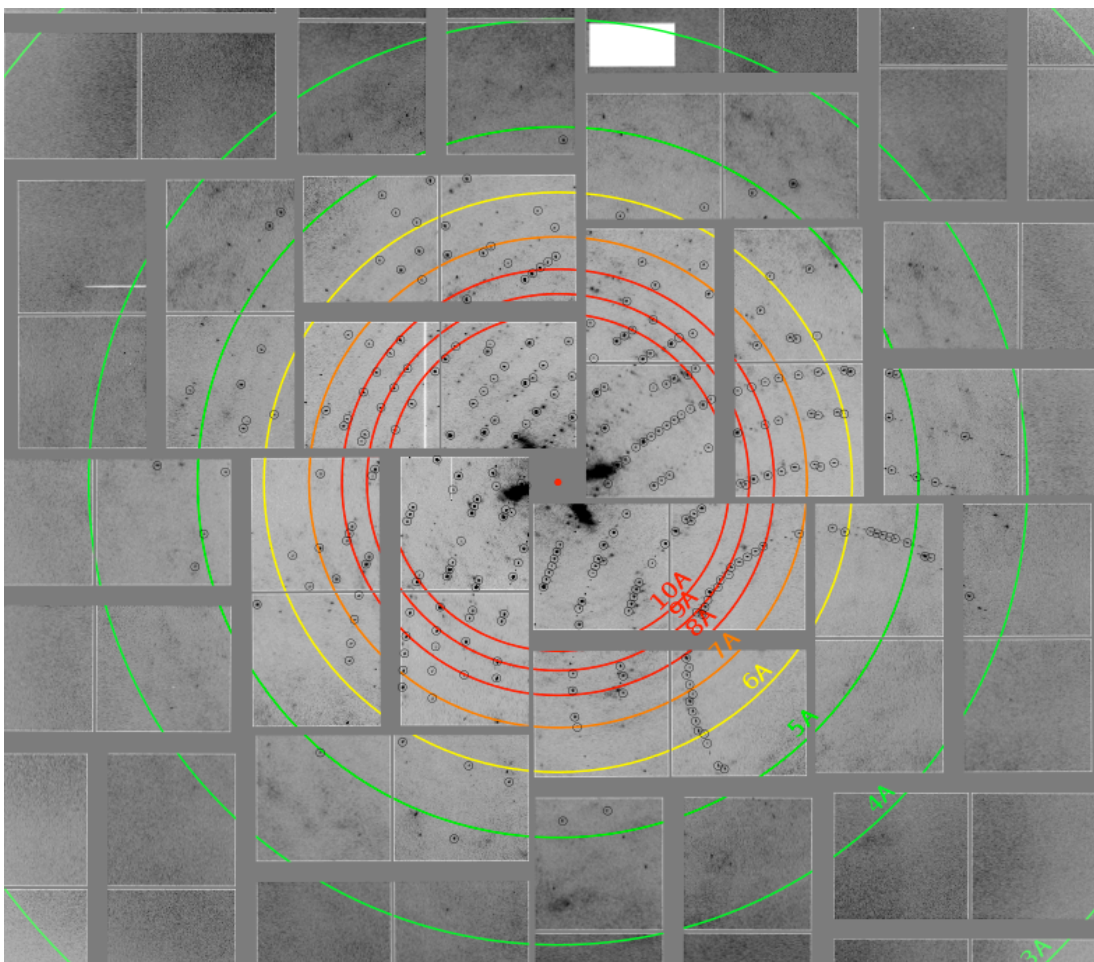
Figure 4.2: Example Diffraction Pattern from the Triple-Flash or Putative S_4 State.
It shows Bragg peak up to 4-Å.

After these pre-processing steps, these crystal hits were passed to the CrystFEL software suite (White et al., 2012) separately for auto-indexing them using MOSFLM, DiraX, and XDS programs. 11,664 and 12,543 diffraction patterns were indexed separately from the dark or S_1 and the triple-flash or putative S_4 states separately at the indexing yield of 34.95% and 38.96% respectively. All patterns were indexed in orthorhombic lattice type with unit cell dimensions of a=132.3, b=226.1, and c=307.3-Å for the dark or S_1 data set. The unit cell dimensions for the triple-flash or putative S_4 state

were identical with those of S_1 state, i.e., a=132.3, b=226.1, and c=307.0-Å. In Fig. 3.3 (See Chapter 3), the unit cell dimensions were compared among all the data sets, i.e., the S_1, putative S_3 and putative S_4 states. The unit cell dimensions increased from the S_1 to the putative S_3 states and reversed to the original values (i.e., dimensions in the S_1 state) in the putative S_4 state. The reason has extensively discussed in previous Chapter 3. Here, in Fig. 4.3, we have provided the unit cell dimensions distribution from June 2012 data sets – S_1 and putative S_4 states.
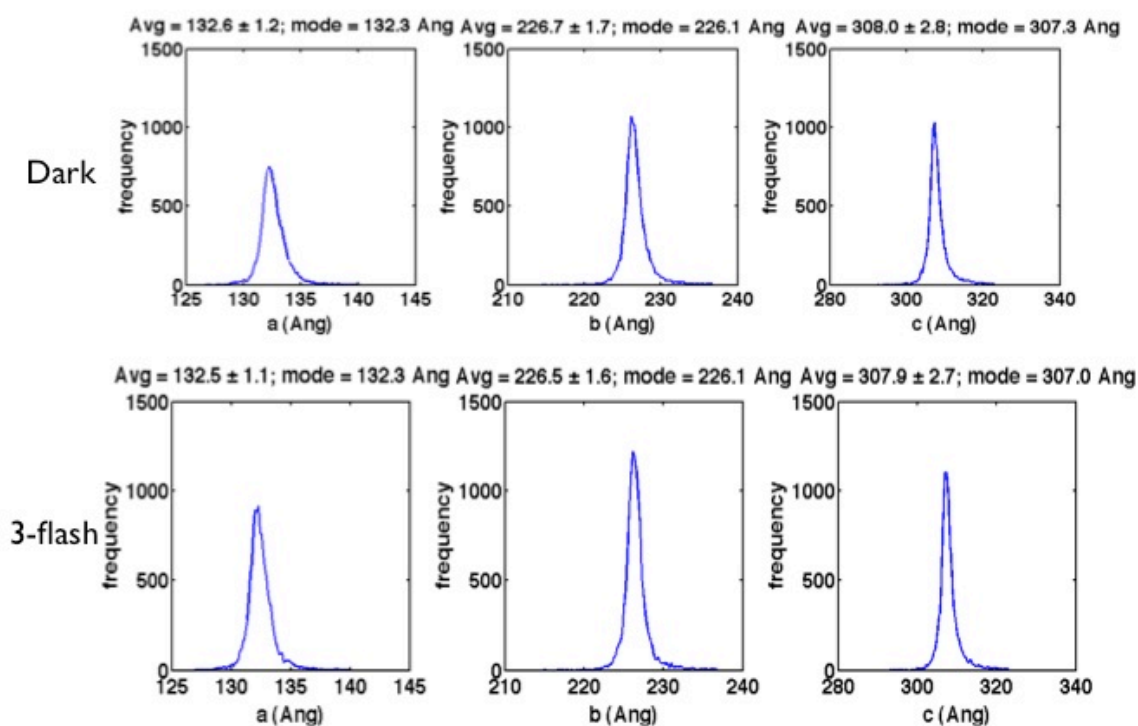
Figure 4.3: Unit Cell Parameters Distribution from the Dark or S_1 and the 3-Flash or Putative S_4 States.

After indexing, the extracted HKL reflections from the indexed patterns were integrated and merged using the Monte Carlo method (Kirian et al., 2011), implemented in CrystFEL suite. Two separate lists of Bragg reflections along with their intensities for the dark and the triple-flash states respectively were obtained. Table 4.1 and 4.2 show the

data statistics for the dark and the triple-flash states respectively. The overall average multiplicities of each reflection measurements were 63.83 and 68.68 for the dark and the triple-flash states respectively. Although the $CC_{1/2}$ values decreased significantly beyond 4.5-Å for both data sets, still both structures could be refined up to 4.0-Å. On the contrary Rsplit values were reasonable even at the 3.0-Å resolution (See Table 4.1 and 4.2). From both tables, the trends of R-split values were consistent until 4.0-Å and started fluctuating beyond 4.0-Å.

Table 4.1: Overall Data Statistics of the Dark or S_1 State. The data is shown in the resolution range of 14.11-3.06-Å.

Resolution (Å)	#Unique reflection	#Measurement	Multiplicity	$I/\sigma(I)$	$CC_{1/2}$	R_{split}	# reflections (for $CC_{1/2}$)
14.11	9,466	1,463,746	154.6	12.09	0.9903	0.0786	9,462
7.27	9,113	976,019	107.1	8.85	0.9556	0.1320	9,113
6.08	9,053	787,312	87.0	8.46	0.8788	0.1520	9,053
5.43	9,024	623,162	69.1	7.85	0.7795	0.1635	9,024
4.99	8,963	658,893	73.5	8.22	0.7240	0.1521	8,963
4.66	8,945	624,460	69.8	8.05	0.5731	0.1551	8,945
4.41	8,934	561,005	62.8	7.54	0.2809	0.1691	8,934
4.20	8,937	588,125	65.8	7.42	0.0738	0.1745	8,937
4.03	8,929	547,169	61.3	6.98	0.0766	0.1900	8,929
3.88	8,915	581,813	65.3	7.33	-0.007	0.1728	8,915
3.76	8,885	619,407	69.7	7.67	-0.005	0.1609	8,885
3.64	8,847	626,312	70.8	7.62	0.0052	0.1581	8,847
3.54	8,922	552,751	62.0	7.11	0.0017	0.1718	8,922
3.45	8,837	480,359	54.4	6.67	0.0177	0.1822	8,836
3.37	8,921	448,686	50.3	6.39	0.0125	0.1918	8,921
3.30	8,853	365,651	41.3	5.85	0.006	0.2162	8,852
3.23	8,836	300,530	34.0	5.06	0.029	0.2782	8,832
3.17	8,865	261,856	29.5	4.15	-0.0153	0.4231	8,859
3.11	8,836	228,429	25.9	4.65	0.013	0.3357	8,821
3.06	8,901	199,680	22.4	4.58	-0.0022	0.2989	8,868
Overall	**178,982**	**114,953,65**	**63.83**	**7.15**	**0.9895**	**0.1036**	**236,752**

Table 4.2: Overall Data Statistics of the Triple-Flash or S₄ State. The data is shown in the resolution range of 14.11-3.06-Å.

Resolution (Å)	#Unique reflection	#Measurement	Multiplicity	$I/\sigma(I)$	$CC_{1/2}$	R_{split}	# reflections (for $CC_{1/2}$)
14.11	9,454	1,574,696	166.6	12.59	0.9903	0.0767	9,449
7.27	9,114	1,056,817	116.0	9.25	0.9529	0.1304	9,114
6.08	9,036	848,503	93.9	8.84	0.8938	0.1450	9,036
5.43	9,027	674,121	74.7	8.16	0.8003	0.1569	9,027
4.99	8,949	708,707	79.2	8.55	0.7287	0.1464	8,949
4.66	8,934	670,770	75.1	8.41	0.6319	0.1451	8,934
4.41	8,919	602,439	67.5	7.85	0.2852	0.1607	8,918
4.20	8,942	632,487	70.7	7.79	0.0830	0.1630	8,942
4.03	8,912	583,340	65.5	7.31	0.0269	0.1831	8,912
3.88	8,895	623,338	70.1	7.65	0.0293	0.1646	8,895
3.76	8,876	666,291	75.1	8.04	0.0394	0.1511	8,876
3.64	8,854	674,729	76.2	8.02	0.0057	0.1494	8,854
3.54	8,905	591,463	66.4	7.49	0.0053	0.1617	8,905
3.45	8,845	516,225	58.4	7.02	0.0052	0.1742	8,845
3.37	8,872	480,320	54.1	6.75	0.0056	0.1813	8,872
3.30	8,880	392,946	44.3	6.14	0.0352	0.2022	8,880
3.23	8,826	322,225	36.5	5.29	0.0182	0.2679	8,824
3.17	8,860	281,287	31.7	4.24	-0.004	0.4056	8,855
3.11	8,816	243,293	27.6	4.81	0.0275	0.3298	8,807
3.06	8,875	212,868	24.0	4.77	-0.0015	0.2820	8,855
Overall	**178,791**	**123,568,65**	**68.68**	**7.47**	**0.9895**	**0.1441**	**237,375**

4.3. Structural Model of Triple-Flash Transient S₄ State

In order to obtain structural solution and final electron densities, each of the two reflection lists was phased with 'Phaser' program (Read, 2003) from CCP4 suite (McCoy, 2007). During phasing step with Molecular Replacement (MR) method, the 1.9-Å structure (PDB code 3ARC; (Umena et al., 2011)) was used as initial search model. This step is same as the one mentioned in previous chapter 3 for the double-flash data set. The 3ARC search model was modified by removing lipids, detergents, and water molecules. In addition, for the triple-flash or putative S₄ state data set, the original

142

plastoquinone model in the Q_B binding pocket was replaced by the artificial plastoquinone, i.e., PQ_{10}. Thus the dark and the triple-flash data sets were phased using the Phaser program (Read, 2003).

The MR solutions obtained from the Phaser program were separately refined with 3 cycles of segmented rigid-body and occupancy refinements. In this case, each subunit was considered as separate rigid-body. During the refinement process, Non-Crystallographic Symmetry (NCS) restraints and torsion angle restraints were maintained. The required restraints regarding bond-angles and dihedral angles were provided through *cif* files. After refinement, the final R-work/R-free for the triple-flash state structure was 0.309/0.314. In Fig. 4.4 and 4.5, 2Fo-Fc electron density maps from the triple-flash state have been shown. In Fig. 4.5, the map looks already superior to the 5-Å structure (See Chapter 3) with many features, being well resolved. The Pheophytin cofactors, the protein surrounding of the non-heme iron, and the mobile quinone (PQ_{10}) in the Q_B binding site were clearly visible in the electron density map at 1.0-σ contour level. The unit cell dimensions were same between the dark and the triple-flash states, thereby the data sets were isomorphous unlike double-flash data set (See Chapter 3). Therefore, instead of simulated annealing omit maps, we made attempts to calculate F_{light} − F_{dark} maps but did not reveal any structural changes between two states − S_1 and putative S_4 states. In addition, 5000 (approx.) different DEN (Deformable Elastic Network; (Schröder et al., 2007)) refinements with different combinations of DEN parameters settings were performed. It implies that the data sets might not have enough completeness and multiplicities to interpret changes.

Therefore, it could be interpreted in several factors. (a) S_4 state being extremely transient could not be reached and the electron density shows a mixture of different states where changes are not interpretable. (b) Even if the PSII micro-crystals advanced to S_4 state, the population in the S_4 state might be too low to see subtle conformational changes at the given resolution of 4.0-Å.
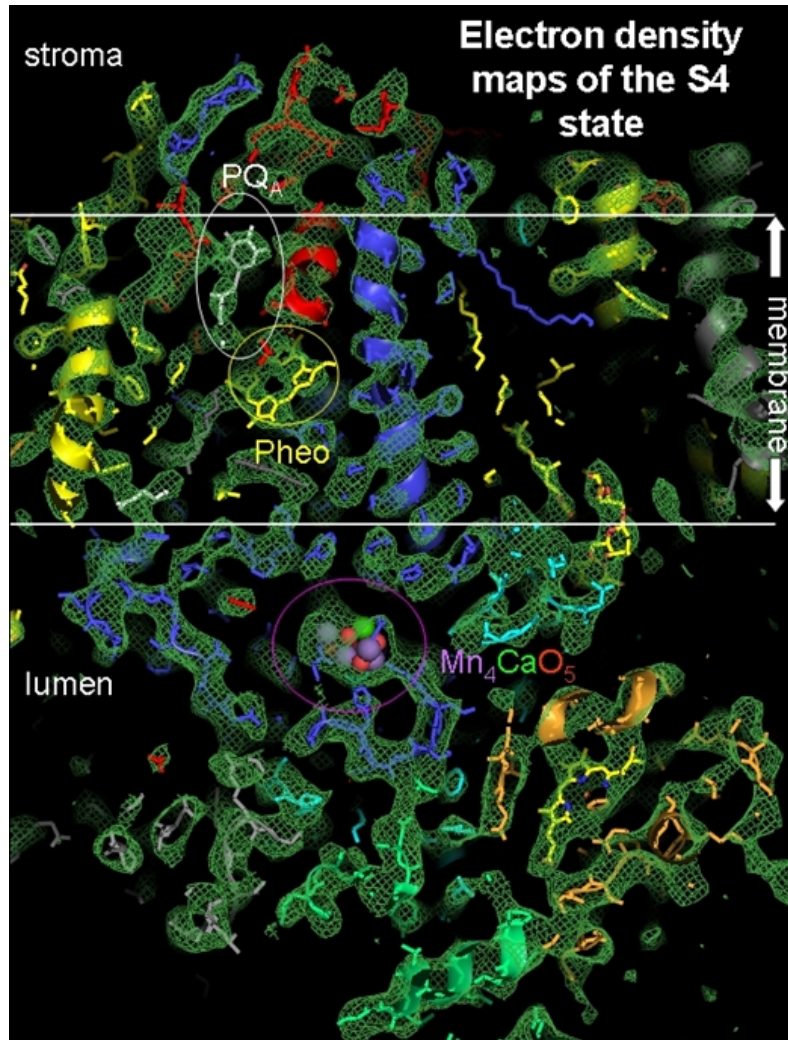
Figure 4.4: $2F_o - F_c$ Electron Density Map (Green) of the S_4 State at 1.5-σ. This shows the Mn-cluster, protein environment from the side view of the cell membrane.
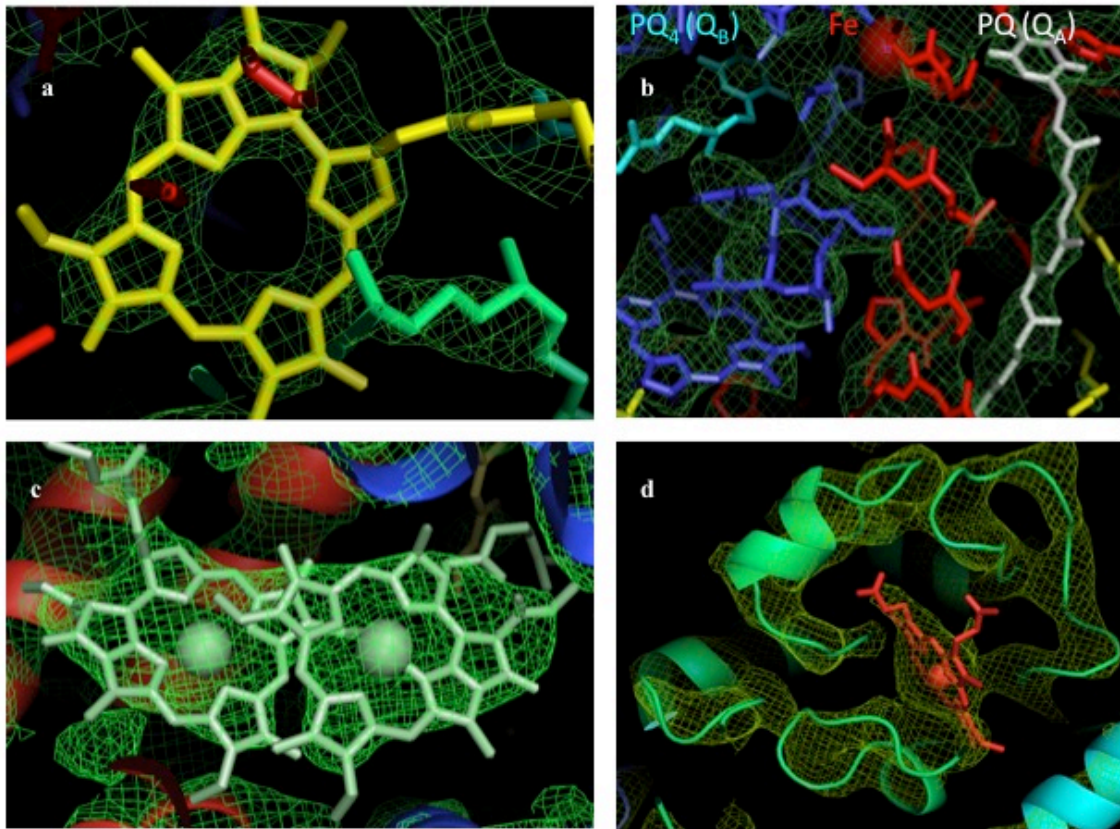
Figure 4.5: $2F_o - F_c$ Electron Density Map (green) of the S_4 State, Showing Various Important Residues Separately. (a) It shows the electron density map for the Pheophytine residue (yellow) in donought shape at 1.0-σ level. (b) It shows the density map for the non-heme Fe atom (red), flanked by PQ_A (white) and mobile artificial PQ_B (cyan) at 1.0-σ level. (c) It displays the P680 chlorophylls (i.e., core of the reaction center) in pale green color and density is contoured at 1.5-σ level. (d) It displays the heme iron residue (red) along with the F chain loop in green color. The density map is contoured at 2.0-σ level.

In conclusion, the time-resolved experiments on the triple-flash state of PSII crystals were partly successful. The artificial PQ pool was successfully introduced in the PSII crystals. The observation of PQ10 cofactor in the electron density of putative S_4 state indicated that native PQ_n has been replaced by artificial PQ derivative, i.e., PQ_{10}. It could

145

be possible that S_4 state, being very short-lived, contains only very subtle conformational changes in the protein environment, surrounding the Mn-cluster. Under such circumstances, the failure to observe conformational change could be accounted for limited resolution. Perhaps, such subtle changes could only be seen beyond 2.5-Å resolution, which is not permitted from the current diffraction data quality. However, we hope that in the future, beamtimes will be awarded to perform more experiments with better crystal quality. The structure at the triple-flash state may be determined at higher resolution and thereby the subtle structural changes could be visible.

5. HIGH SPATIAL AND TEMPORAL RESOLUTION OF PHOTOACTIVE YELLOW PROTEIN CRYSTALS, USING SFX

This chapter will describe the time-resolved SFX experiment, performed on Photoactive Yellow Protein (PYP), a very well-studied model system, at the LCLS facility using Serial Femtosecond X-ray crystallography (SFX) method. This time-resolved work demonstrates that SFX can produce high time and spatial resolution from time-resolved studies of protein crystals, as well as enabling to quantify the structural changes between various reaction intermediates thoroughly. This work was first published in *Science* journal in December 2014 (Tenboer, Basu et al*.,* 2014).

5.1. Motivation

In Chapters 3 and 4, two early time-resolved experiments and related structural findings have been described. Both of those two experiments were carried out in 2012, which was right after 3 years of the LCLS being operational. Thereby, those early attempts in time-resolved studies proved that nominal resolution from large membrane protein complexes could be adequate to qualitatively interpret structural changes between ground and excited states. However it was very important to do more time-resolved SFX experiments to establish that SFX and the XFEL techniques could achieve high temporal and spatial resolution, enabling to visualize real-time molecular movies. Early successes in time-resolved works on Photosystem II (Kupitz and Basu et al., 2014) motivated to continue the effort to establish SFX as method for making molecular movies by enabling to reach higher time and spatial resolution. A classic work by Boutet et al., (Boutet et al.,

2012), demonstrates that SFX can achieve high spatial resolution (i.e., 1.9-Å) along with damage free structure from the lysozyme crystals, a very small molecule compared to PSII. However, it was already realized from the time-resolved Laue experiences that in order to visualize and quantify very subtle conformational changes (i.e., movement of certain residue by ~1-Å), less than 2-Å resolution is still not sufficient. Therefore, it was the high time to execute TR-SFX experiments with a model system (such as PYP), well studied in time-resolved Laue method (Schotte et al., 2004)(Schotte et al., 2012) such that a proof-of-concept experiment could be done to establish TR-SFX as a method, capable of producing high time and spatial resolution, given the crystal quality is sufficiently good.

5.2. Aim at High Time and Spatial Resolution – Proof-of-Principle Experiment

Aim of this time-resolved work was to achieve high temporal and spatial resolution using a model system, namely photoactive yellow protein (PYP), well studied by the time-resolved Laue method (Ihee et al., 2005)(Schotte et al., 2012). The goal was to establish TR-SFX as the technique towards making of molecular movies at atomic resolution. Laue time-resolved studies have some limitations: a) radiation damage may hinder the interpretation of structural changes, b) Time-resolution is limited to ≤100-ps, c) large crystals (≥200-μm to 3mm) are required, which is not easy to excite and due to large absorption path length, only a few molecules undergo photo-reaction (in case of PYP, it is <10%, (Schmidt et al., 2013)). All these limitations can be circumvented in the TR-SFX method.

The TR-SFX experiment on PYP was performed in June 2014 at the CXI instrument of the LCLS. PYP is purple-bacterial photoreceptor. It absorbs blue-light and undergoes a photocycle (See Fig. 5.1) consisting of several short-lived intermediates – I_T, I_{CT}, PR_1, PR_2, PB_1, and PB_2. The time-constants for the photo-conversion steps range from femtosecond to second. PYP has a chromophore, called *p*-coumeric acid, covalently bonded to a Cys residue. Relatively long-lived (i.e., lifetime > 100-ps) intermediates have already been studied by the Laue method (Schotte, 2003)(Schotte et al., 2004)(Ihee et al., 2005).

PYP is a small soluble protein with 125 residues. It can be crystallized in very high quality, similar to those of lysozyme. Thereby, PYP could serve as a very robust model system for a proof-of-principle experiment for TR-SFX at near-atomic resolution. In this work, two data sets from two different excited states, at the time-delays of 10-ns and 1-μs were collected and evaluated to 1.5-Å resolution. This is the very first TR-SFX work (Tenboer et al., 2014), which could achieve temporal resolution of 10-ns with a biological system. It is noted that in the original Science paper (Tenboer et al., 2014), the resolution of presented data was 1.6-Å. Later further data processing improved the resolution to 1.5-Å, presented in this dissertation.
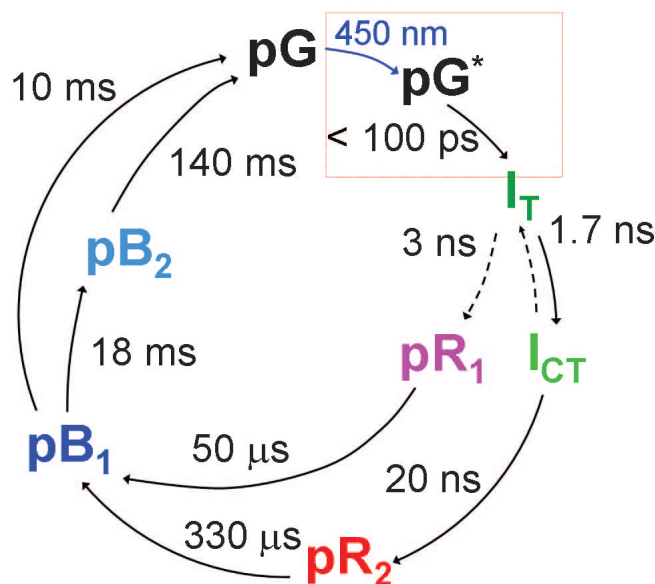
Figure 5.1: Photocycle of Photoactive Yellow Protein (PYP). This depicts various short-lived intermediates of PYP along with their lifetimes.

5.3. Experimental Methods for TR-SFX Experiment of PYP

The time-resolved experiment on PYP micro crystals (size ~3 to 5-μm) was performed at the CXI (Coherent X-ray Imaging) instrument at the LCLS facility in June 2014. The experiment was carried out at 8.995 keV X-ray energy in pump-probe manner (described in (Kupitz and Basu et al., 2014)) with time-delays of 10-ns and 1-μs between pump pulse and X-ray probe pulse. The time-delay of 1-μs between pump and X-ray pulses produces a mixture of two reaction intermediates - pR_1 and pR_2 states in the photocycle and the time-delay of 10-ns produces a mixture of three reaction intermediates − I_{CT}, pR_1, and pR_2 (See Fig. 5.1). The PYP micro crystals were injected at the X-ray interaction region using the Gas Dynamic Virtual Nozzle (GDVN) liquid injector (Weierstall et al., 2012). The liquid jet velocity ranges from 12-μl/min to 25-μl/min and X-ray FEL beam operated at 120-Hz frequency. The photochemical reaction was initiated

with two 20-Hz nanosecond lasers (λ = 450-nm), which were used in series such that pumping frequency became 40-Hz. The Xray pulse frequency was 120-Hz. This means, every third crystal diffraction pattern would be from the 'pumped' or excited state. In Chapter 3, a schematic of pump-probe TR-SFX experiment on PSII has been shown (Fig. 3.1b). The experimental set up for the PYP TR-SFX experiment was very similar, except that two pump lasers operated together at 40-Hz at λ = 450-nm with an illumination spot size of 150-nm diameter. This means that instead of collecting data in an alternating dark-light mode, during the PYP experiment, data were collected in a light-dark-dark-light mode, producing two times more data from the ground (i.e., dark) state than that of the excited (i.e., light) state.

Two Opetek nano-second lasers from UW-Milwaukee and BioCARS, University of Chicago, were brought to the LCLS for this experiment. The output of these two lasers was connected to a single 400-μm fiber and channeled into the sample chamber in vacuum. The laser was focused at the jet such that 90% of the intensity was focused at the 150-μm diameter illumination spot. The laser power was set at 800-μJ/mm^2 or 800 x 10^{-12}-J/μm^2 at a wavelength of 450-nm. In order to optimally pump a 2-μm cuboid shaped PYP crystal, we estimated that the total number of photons at 450-nm wavelength in one 'pump pulse' should be 7.3 x 10^9. The concentration of PYP molecules in crystal was 63-mMol/Litre, thereby number of PYP molecules in a 2-μm cuboid shaped crystal was 3.79 x 10^8. Given that each molecule needs one photon at 450-nm wavelength, approximately, 24-photons/molecule was added to ensure a high reaction initiation of ~50%. It was necessary because the primary quantum yield of PYP molecule is 10%. To

synchronize the pump laser pulses with the X-ray pulses, two event code reader signals (EVR) were generated by the LCLS at 20-Hz. There were also two Aquiris photodiodes, each of which connected to each pump laser to cross-validate the actions of the EVR channels, i.e., to ensure the time-stamp of the pump laser is properly recorded in each diffraction frame, recorded at the CSPAD (Cornell-SLAC Pixel Array Detector). There was an additional fast diode placed inside the vacuum chamber to verify that the laser light (i.e., visible light) was delivered to the crystal in jet. It also served as timing diagnostics between pump laser and X-ray. To determine time T_0, the x-ray and laser signals were measured on the same fast diode in the vacuum chamber. The X-ray arrival time was referenced, which provided temporal information about the delay between laser and X-ray on a shot by shot basis with nanosecond resolution. The delays on the two event codes were adjusted to overlap the X-ray and laser signals and therefore determine T_0.

5.4. Study of Reaction Kinetics and Reaction Initiation of PYP Crystals

These time-resolved experiments were performed in 'pump-probe' manner, explained in the previous section. No pumping or excitation process is 100% efficient. As a result, a mixture of various excited states, including some fraction from the ground state would be obtained. The degree of this heterogeneity would depend on efficacy of the 'excitation process'. In order to extract structural information, purely related to only one excited state, the structural factors need to be deconvoluted using singular value decomposition method ((Schmidt et al., 2003); see Section 1.3.5). However, to design an optimal

excitation scheme, it is important to understand the reaction kinetics of the molecules in solution and crystal phases.

In this work, there are two excited states being studied - one excited state with lifetime of 1-μs time and second one with 10-ns lifetime. From Fig. 5.1, it is clear that there would be two prominent intermediates in the mixture after 1-μs of excitation. In case of kinetics measurements in the solution phase of PYP molecules, concentration of various intermediates (at λ=450-nm) as a function of reaction time could be plotted (Fig. 5.2a). The plot shows two solid vertical lines, indicating the time-points being studied in this work. From this kinetics study, it is evident that at the 1-μs time-point, there will be a mixture of pR_1 and pR_2 intermediates. At the 10-ns time-point, intermediate I_{CT} will exist along with I_T, and pR_1 intermediates. The population distribution of various intermediates at 10-ns time-point is nicely shown in a kinetics plot (Fig. 5.2b) of wild-type PYP molecules in a recent study from Ihee and coworkers (Jung et al., 2013).
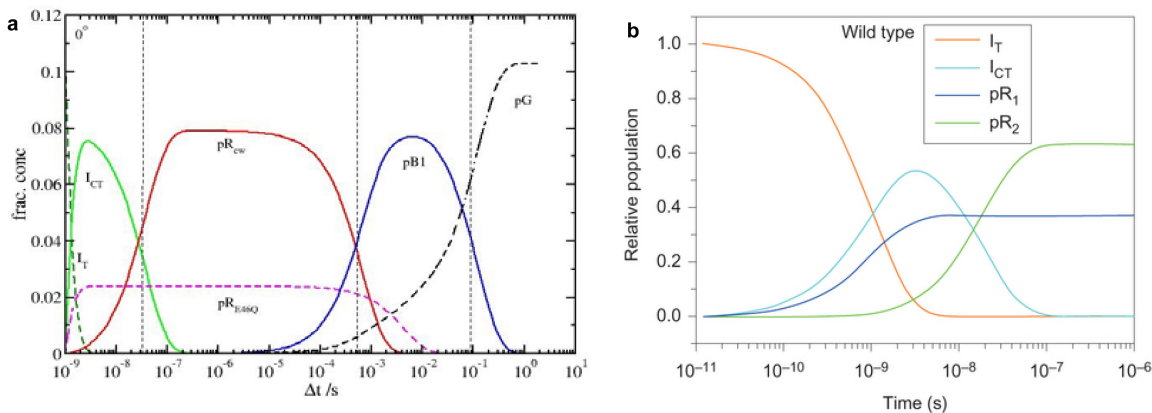
Figure 5.2: Reaction Kinetics Study of PYP in Solution Phase. (a) It shows the plot of concentration change of various PYP intermediates as a function of reaction time progress. (b) It shows a plot from (Jung et al., 2013), showing relative population of various intermediates as a function of the time-progression of photo-cycle.

153

To calculate the rates and yield of excitation and reaction initiation in PYP crystals, there are several factors that need to be considered. In this work, a visible laser at λ=450-nm, providing 5-ns top-hat profiled laser pulse was used for the excitation process. The laser fluence used was 800-μJ/mm^2. So, given the size of the crystal be 2-μm, laser energy incident at a cross-section of 2 x 2-μm^2 would be 3.2 x 10^{-9}-J. This corresponds to the amount of energy received by the 2-μm cuboid shaped PYP crystal. Energy of one blue photon at 450-nm is 0.044 x 10^{-17}-J. Thereby number of photons incident on a 2-μm sized crystal was (3.2 x 10^{-9} J/0.044 x 10^{-17}-J =) 7.3 x 10^9. The protein concentration in a PYP crystal was 63-mM, resulting in a concentration of molecules in the PYP crystal of (63 x 10^{-3}-mole L^{-1} x 6.023 x 10^{23}-mole^{-1} /10^{15}-μm^{3}L^{-1} =) 3.79 x 10^7 molecules/μm^3. One PYP crystal of the size 2 x 2 x 2-μm with a volume of 8-μm^3 contains 3 x 10^8 molecules (number of unit cells in one 2-μm cuboid PYP crystal = 4.38 x 10^7). It implies that when 7.3 x 10^9 number of blue-photons incident on the 3 x 10^8 molecules per 2-μm sized PYP crystal, ~24-blue-photons/molecule will be excess. This means, all molecules received at least one blue-photon. But, excitation process also depends on quantum efficiency, molar extinction co-efficient (45,500-cm^2/mol) and the penetration depth of the medium (i.e., longest dimension of the crystal). The primary quantum yield of PYP molecules is 10% at λ=450-nm. This means, upon receiving blue-photon, pG converts into pG* (See, Fig. 5.1) and only 10% of pG actually advances further in the photo-cycle and 90% returns from pG* to the ground state, i.e., pG. It is noted that the lifetime of pG* is 500-fs (Lincoln et al., 2012). When a 5-ns optical pump laser will be used for the excitation, pG -> pG* transition will occur multiple times and each time, after excitation, 10% of

154

remaining pG* will advance further in the photocycle, which in turn leads to the estimation of ~55% of reaction progression from pG* to further steps in the photo-cycle. Fig. 5.3 shows that ~90% molecules in a 2-μm thick crystal will reach pG* photo-excited state by the end of a laser pulse duration of 5-ns. But because the primary quantum yield of PYP molecule is 10%, and in 5-ns pulse duration, pG -> pG* conversion can occur multiple times (pG* life time is very short ~500-fs), eventually an estimated amount of ~55% of pG (i.e., ground state) will further advance in the photo-cycle after the end of 5-ns pulse duration. It is evident that the advantage of using micro-crystals in the time-resolved study is larger reaction initiation because of shorter penetration depth compared to large macro-crystals used in Laue crystallography.
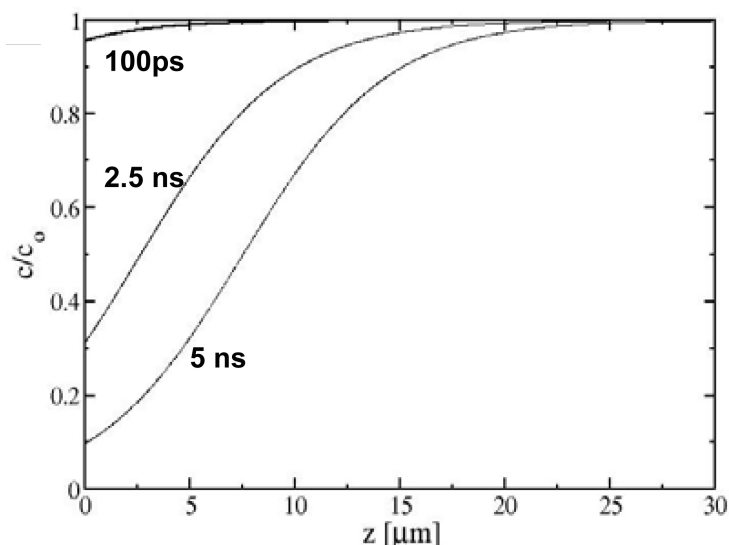
Figure 5.3: Quantification of Laser Excitation of PYP Crystals. This plot shows the fraction of concentration left in the 'unexcited' or ground state as a function of varying penetration depth or crystal size. From the plot, it is noted that ~90% molecules are excited to pG*state by the end of 5-ns pulse duration for a 2-μm sized crystal. It is noted that primary quantum yield of PYP molecule at λ=450-nm is 10%. It means even though ~90% molecules reach pG* state, only 10% of them advance further in the photo-cycle. But because of very short lifetime of pG* (~500-fs), pG->pG* conversion can happen multiple times during one 5-ns pulse duration, advancing estimated amount of ~55% to the final products.

155

5.5. Data Processing – Strategies to Achieve High-Spatially Resolved Data

This time-resolved experiment on PYP micro-crystals was carried out at 8.995-keV (i.e., λ=1.33-Å) X-ray energy with a ~0.3% band width and 1-mrad angular divergence. ~3.7 million diffraction patterns in total were collected from the stream of PYP crystals in the dark (or ground) and the light (or excited) states. The CSPAD detector was placed at 80-mm distance from the crystal interaction region such that the resolution at the edge of the detector was 1.45-Å. It is true that the quality of the data and resolution range mostly depends on the quality of the crystals and experimental limitation of the detector. Quality of the PYP crystals was high and crystallization protocols were well established for the PYP protein (Tenboer et al., 2014). On the other hand, time-resolved information, i.e., the subtle structural changes between two time-dependent states requires significantly high resolution (i.e., <2-Å). Therefore, a new strategy for data extraction from the CSPAD detector was implemented using two different pixel gain settings. Low resolution, i.e., small scattering angle diffraction spots are always very high in intensity, saturating the detector pixels electronics and the high resolution, i.e., wide scattering angle diffraction spots are always weak and with low signal-to-noise ratio, those are almost hidden in the background noise because scattering intensity decreases with increasing scattering angle. High intensity diffraction at small scattering angle, saturating the low-resolution diffraction spots, results in erroneous peak integration. At high scattering angle, the challenges are to extract the weak intensity signals of high-resolution diffraction spots from the background noise, which are the key to information on the structural changes at the atomic level. Thereby, during the experiments, the pixels on the

156

CSPAD detector up to 4-Å were set at a 6.8 times lower pixels-gain setting and pixels outside the 4-Å boundary (i.e., the high-resolution pixels) were set at 6.8 times higher pixels-gain setting, assuming the detector sensitivity being flat-field. Thus, the low-resolution diffraction spots were made weaker, reducing the number of saturated pixels in that resolution range and allowing for data to be collected at higher photon flux (i.e., less attenuation). Fig. 5.6 shows background corrected diffraction pattern of PYP crystal with two-different pixel-gain setting. This strategy was very important and led to the visualization of significant structural changes between two time-dependent states. The incident X-ray intensity was increased up to 30% transmission safely to strengthen the high-spatial resolution data without compromising the low-resolution data by saturation.

In this experiment, two different time-points – 10-ns and 1-µs in the photocycle were studied. In each of these two pump-probe experiments, every third pattern was from the excited state, while the other two patterns were from the ground state. Thereby, 2,049,294 frames from the dark (or ground) and 1,024,684 patterns from the corresponding 1-µs time point light (or excited) states were collected (See Table 5.1). For the 10-ns time-point data collection, 414,326 frames from the dark (or ground) and 207,168 frames from the light (or excited) states were collected.

All of these raw diffraction frames (in XTC format) were pre-processed using the "Cheetah" program (Barty et al., 2014). During this step, diffraction patterns pertaining to PYP crystals were extracted based on several steps, including detector electronic noise subtraction, masking of bad, hot, dead pixels, background scattering correction with a signal-to-noise ratio of 7, threshold peak intensity of 200, and minimum number of peaks

of 25 criteria. Fig. 5.4 shows an example diffraction raw-frame without background correction. Fig. 5.5 shows the same pattern after background correction (i.e., after pre-processing step).
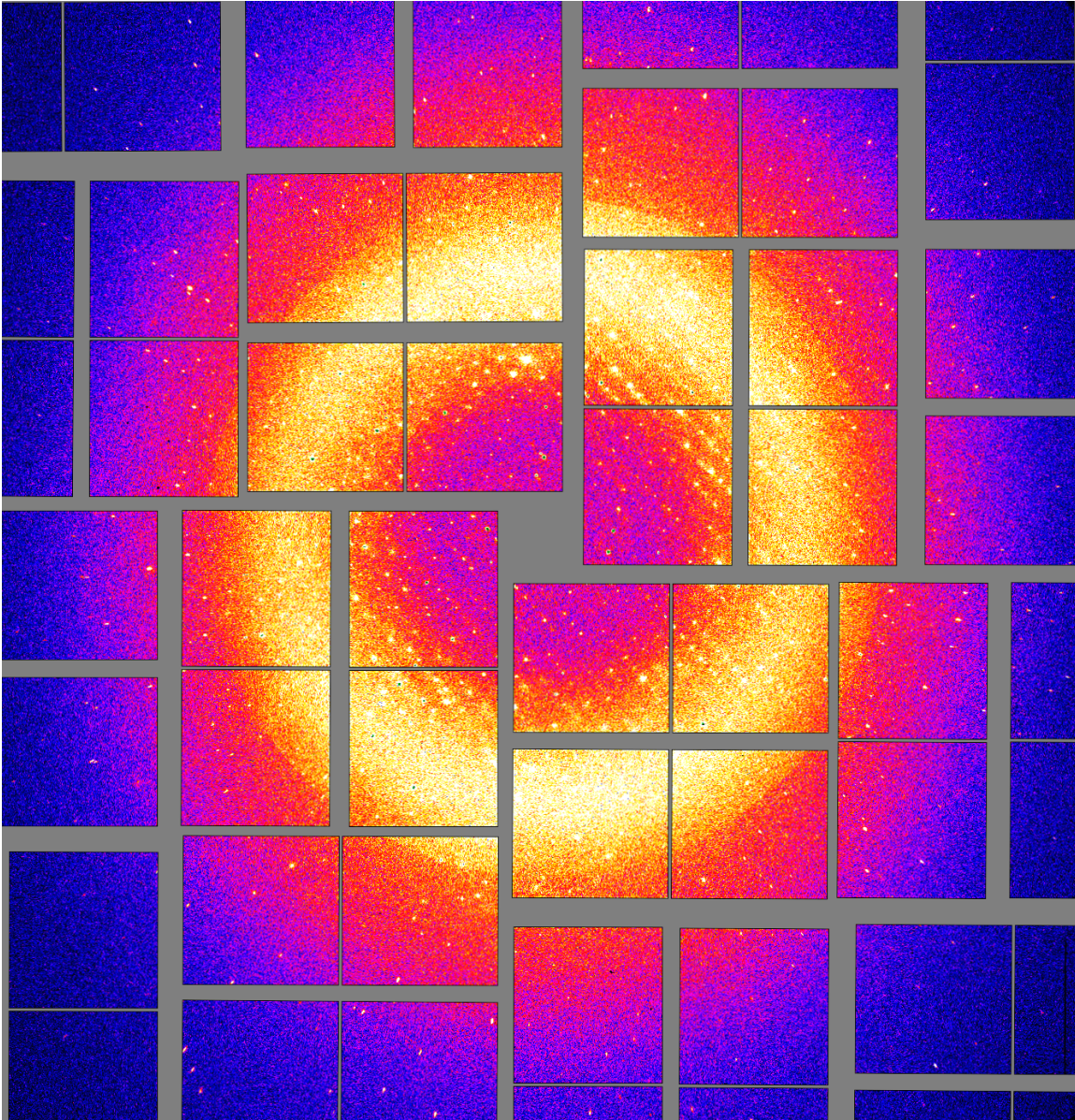
Figure 5.4: An Example Raw Diffraction Frame from PYP Crystal before Pre-Processing Step.
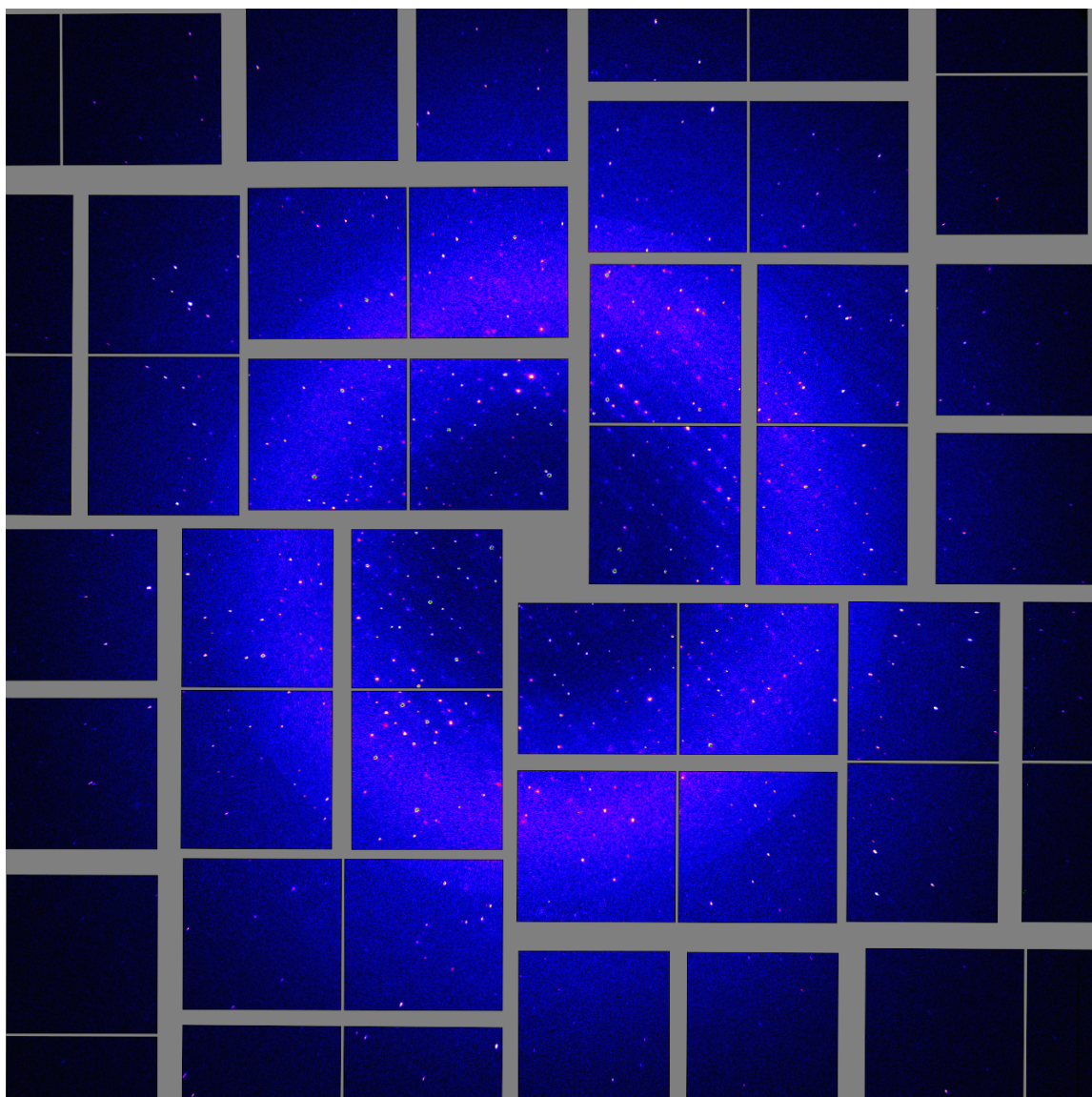
Figure 5.5: The Same Diffraction Pattern (as Shown in Fig. 5.4) after Background Correction.

Since the detector was put very close to the interaction region, edges and corners of the detector were partly shadowed due to the shroud of the nozzle, the crystal delivery

instrument. Those 'shadowed' regions on the detector were also masked. Fig. 5.6 shows background corrected diffraction frame with two different pixels-gain-setting. It is noted the correction of the pixel gain factors and accordingly scale the intensities from two different regions (i.e., high and low- resolutions) of the detector were done (shown in Fig. 5.7) using a script that is now added to the main Cheetah package. This script identifies pixels locations with lower gain setting, multiplies those pixel values with the suitable gain-factor (in PYP experiments, the factor was 6.8) and saves those pixels information as a "gain map". This gain map, containing the identity of pixels, set to lower and higher gain modes, was provided for the intensity scaling. Fig. 5.7 shows such diffraction frame after intensity scaling based on a gain map.
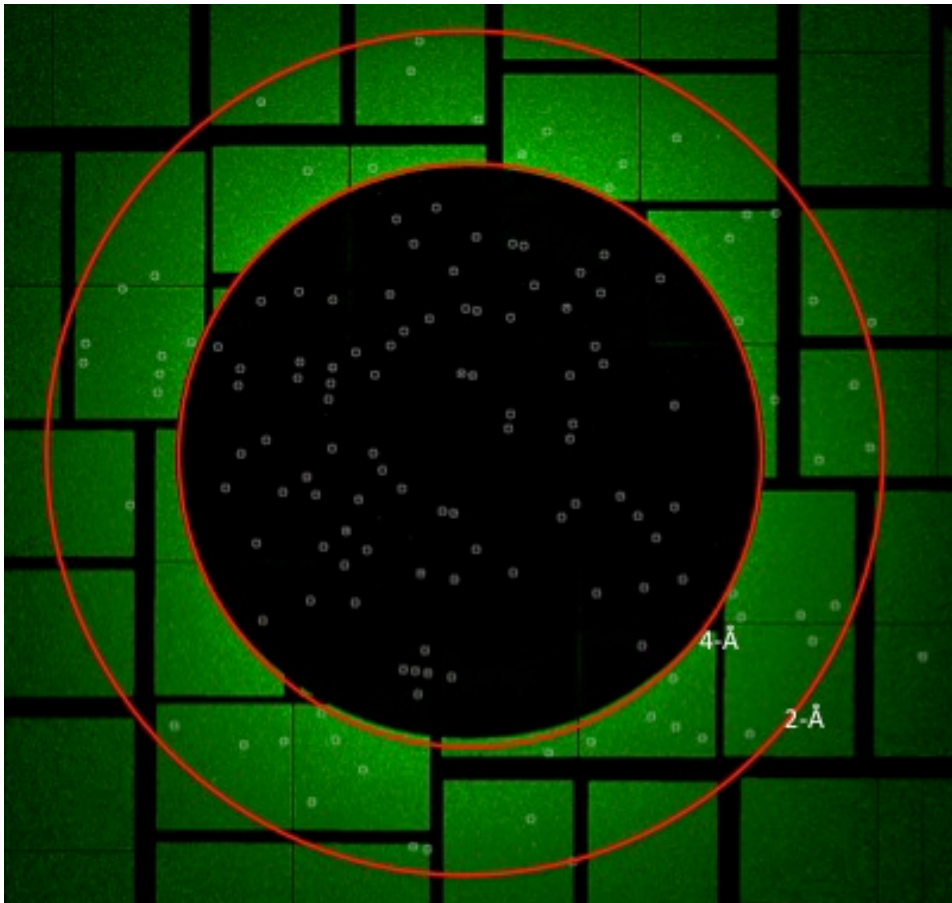
Figure 5.6: An Example of Background Corrected Diffraction Pattern with Pixels in Two Different Gain-Settings.
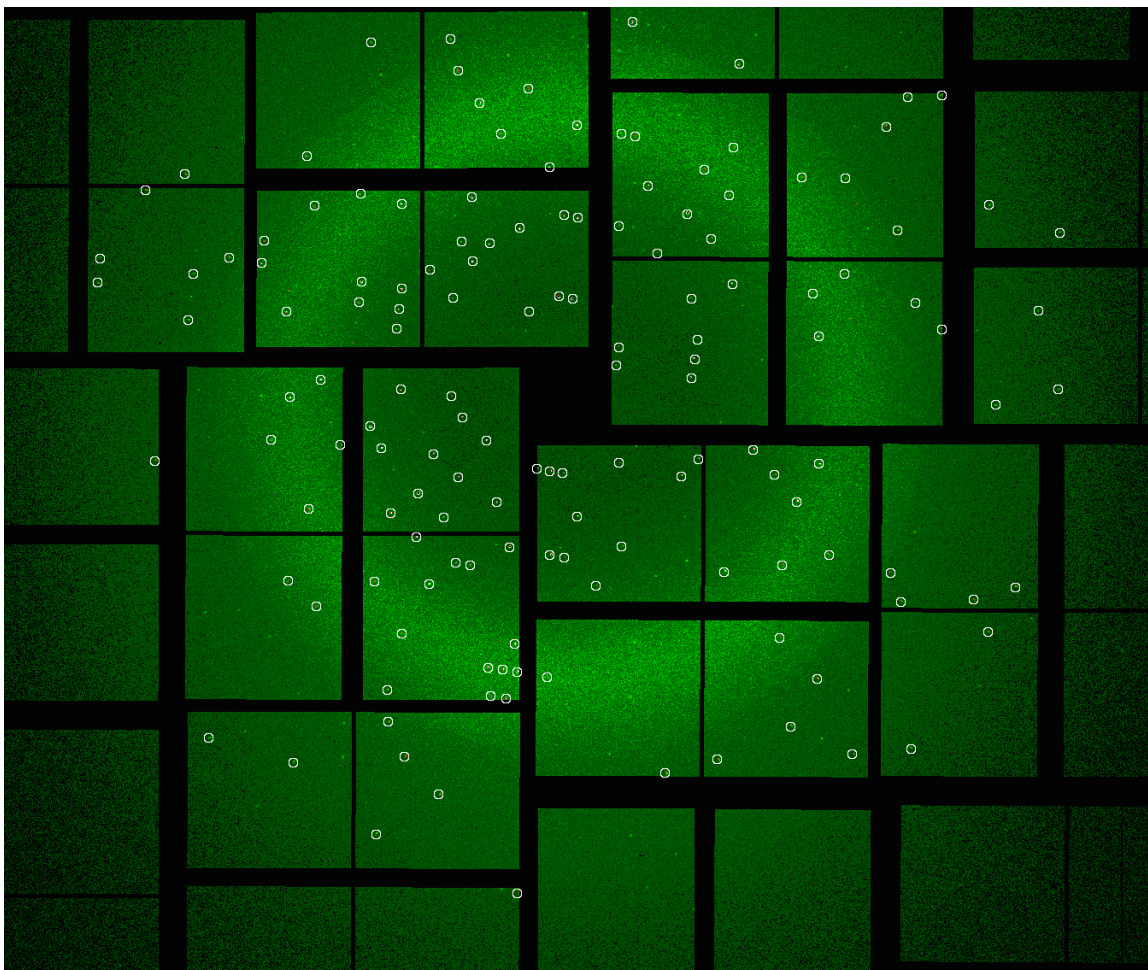
Figure 5.7: An Example of Background and Pixels-Gain Setting Corrected Diffraction Frame. The pixel-gain correction is performed based on a gain map. Note: this is the same diffraction frame as the one shown in Fig. 5.6.

The pre-processing steps are essential to extract accurate diffraction intensity data from the diffraction pattern and to reduce the large amount of collected raw data. It eliminates frames with no diffraction spots from the crystals, i.e., frames caused by the "misses" from X-ray and crystal interaction. It also rejects frames with very few and weak diffracting spots, as they do not contribute to the data evaluation process. An example diffraction pattern from PYP crystal is shown in Fig. 5.8.

162

After the pre-processing step, 95,154 crystal "hits" from the dark and 45,995 crystal "hits" from the excited states were obtained from 1-µs time-point data set. For the 10-ns time-point data set, 75,757 and 36,632 crystal "hits" from the dark and the corresponding 10-ns excited states were obtained respectively. It is noted that the sorting of time-resolved data set into diffraction patterns from the dark and the light states was done with a code, written during the experiment to correctly separate the data sets in those two sets at the earliest stage of processing and treat them separately onwards.
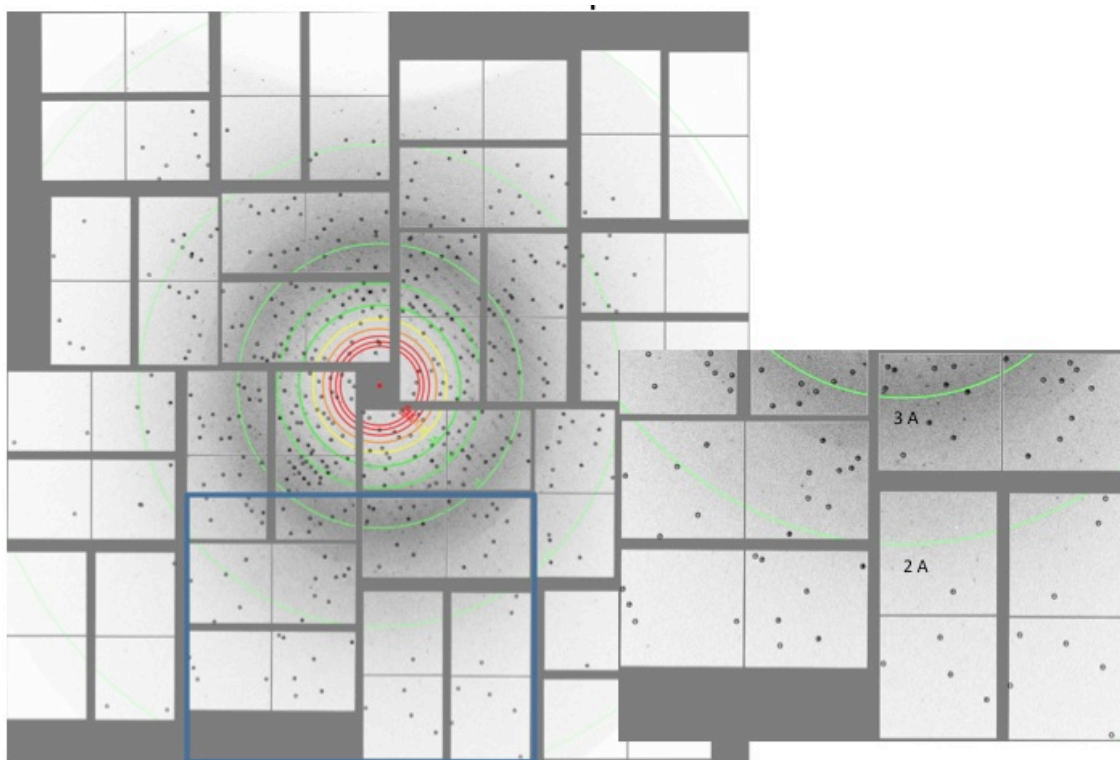
Figure 5.8: Example of a Typical PYP Crystal Diffraction Pattern. This shows the diffracted spots up to the edge of the detector (i.e., beyond 2-Å).

Later, these patterns from PYP crystal "hits" were passed to CrystFEL suite (White et al., 2012). Indexing of each of these patterns were performed separately using MOSFLM, DiraX, and XDS algorithm, which are implemented in CrystFEL for the evaluation of

SFX data. As a result, crystal orientations and corresponding lattice parameters were determined as $a=b= 66.9 \pm 0.22$ and $c= 40.8 \pm 0.19$-Å in the hexagonal lattice type. From the 1-μs data set, 71,577 patterns out of 95,154 crystal hits from the dark state were indexed with a very high indexing yield of 75.22%. 34,701 patterns out of 45,995 crystal hits (i.e., indexing yield of 75.44%) were indexed from the light (or 1-μs) excited state. From 10-ns data set, 51,394 patterns out of 75,757 crystal hits (i.e., indexing yield=71.62%) were indexed from the dark state and 24,940 patterns out of 36,632 crystal hits (i.e., indexing yield = 68.1%) were indexed from the 10-ns lifetime excited state. The unit cell parameters mentioned above (See Table 5.1) are identical for the ground and for the excited states within a very narrow error range. It is noted that PYP crystals, being in the hexagonal space group P6$_3$, possess pseudo-merohedral twinning (i.e., not physical twinning), resulting in indexing ambiguity. The indexing ambiguity was solved using the "ambigator" program, implemented in CrystFEL based on (Brehm and Diedrichs et al., 2014). After solving the indexing ambiguity of the diffraction data, the reflection data set of very high multiplicity (See multiplicity in Table 5.1), containing 'partial' intensities of Bragg reflections were integrated and merged using Monte Carlo (Kirian et al., 2011) method, implemented in CrystFEL. Thus, total of four reflection-lists – from the dark and the light or excited states at Δt time-point of 10-ns and 1-μs.

Table 5.1: Data Statistics from the Both Time-Resolved Data Sets.

	10-ns		1-µs	
	Dark	**Light**	**Dark**	**Light**
Resolution (Å)	1.5	1.5	1.5	1.5
X-ray energy (keV)	8.995	8.995	8.995	8.995
CSPAD to jet distance	75-mm	75-mm	75-mm	75-mm
Pump fluence	800-µJ/mm^2	800-µJ/mm^2	800-µJ/mm^2	800-µJ/mm^2
Pump wavelength (nm)	450	450	450	450
Frames	414,326	207,168	2,049,294	1,024,684
Hits	75,757	36,632	95,154	45,995
Indexed	51,394 (71.62%)	24,940 (68.1%)	71,577 (75.22%)	34,701 (75.44%)
Pixel attenuation[to]	Yes [4-Å]	Yes [4-Å]	Yes [4-Å]	Yes [4-Å]
Cell constants	66.9±0.19, 66.9±0.19, 40.8±0.15	66.9±0.19, 66.9±0.20, 40.8±0.16	66.9±0.22, 66.9±0.24, 40.8±0.18	66.9±0.22, 66.9±0.22, 40.8±0.19
Space group	P6$_3$	P6$_3$	P6$_3$	P6$_3$
Completeness	98.62	96.62	99.45	98.82
Multiplicity	2,270.67	1,114.63	3,117.76	1,526.50
I/σ(I)	6.0 (2.3)	4.3 (1.7)	12.4 (4.7)	8.7 (3.0)
R$_{split}$ (%)	12.20 (54.33)	16.78 (60.24)	5.76 (41.78)	8.00 (50.72)
R$_{work}$ (%)	16.7	--	19.8	--
R$_{free}$ (%)	19.5	--	22.1	--
<|ΔF|>/<σ$_{ΔF}$>	8.1/8.2		6.3/5.07	
Reaction initiation	--	--		40%; 18% pR$_1$, 22% pR$_2$

Difference Electron Density Map Calculation:

Using these reflection-lists, difference electron density (DED) maps ($F_{light} - F_{dark}$) were calculated. For this, the knowledge of accurate phase information from the dark (or ground) state was mandatory. This phase information was derived from the dark state X-ray structure (PDB code: 2PHY) (Borgstahl et al., 1995). The resolution (30-Å to 1.5-Å) and quality of the observed dark state structure factors was so good that 2PHY model fits into the SFX diffraction data very nicely at the MR stage without doing any refinement. However, a routine refinement of 10 cycles was performed on the model generated based on the observed structure factors from the dark data set using Refmac program (Murshudov et al., 1997) from the CCP4 suite. As a result, R-work of 19.8% and R-free of 22.1% (See Table 5.1) were achieved from the dark data of 1-μs time-point data set. Then the observed dark structure factors were scaled on an absolute scale to those from the calculated dark PYP model. The excited state structure factors were also scaled against the observed 'scaled' dark state structure factors using Scaleit program in the CCP4 suite. Weighted difference structure factor amplitudes were calculated (See Table 5.1) as explained in (Ren et al., 2001)(Schmidt et al., 2003). Using these difference amplitudes and the phases from the dark state model, weighted difference electron density maps were calculated (See Fig. 5.9).

Calculation of the Reaction Initiation Yield:

After calculating the difference electron density (DED) map, i.e., ($F_{light} - F_{dark}$), it is important to quantify the extent of the reaction progression at Δt time used for the study.

166

For this purpose, a comparison between 'calculated' DED map with 1-μs DED map from observed 1-μs data was performed. The fitting parameters provided the information about the reaction extent. Based on kinetics study (See Section 5.4), at the 1-μs time-point, a mixture of pR_1 and pR_2 intermediates would be obtained. There is already a pdb entry 1TS7 (Ihee et al., 2005), containing two intermediates (i.e., pR_1 and pR_2). Therefore, their structures and phase information are already known. The phase and the dark state structure are known from 2PHY (mentioned previously). Two difference electron density maps (i.e., (F_{pR1} - F_{dark}) and (F_{pR2} - F_{dark})) were calculated by subtracting structure factors of the dark model from the structure factors of pR_1 and pR_2 separately (See Fig. 5.9). These two DED maps are called $\Delta\rho(pR_1)$ and $\Delta\rho(pR_2)$ respectively. Since both of these intermediates were present in 1-μs time-point data, the linear combination of these two DED maps were fitted with the DED map, obtained from the observed 1-μs data set, collected during this work. The fitting parameters would be the scaling factors, (i.e., linear coefficients), which will determine the extent or occupancy of each intermediates in the map. That in turn provides the information about the extent (See Table 5.1) of reaction progression, using the following equation: $\{\Delta\rho^{obs} - [Sc_1. \Delta\rho(pR_1) + Sc_2 . \Delta\rho(pR_2)]\}^2_{min}$. Based on the occupancy calculation of each intermediates in the map, the yield of pR_1 and pR_2 are 18% and 22% respectively (i.e., total of 40%). This means 40% of ground state PYP molecules reached the reaction intermediate states at 1-μs time-point.

167
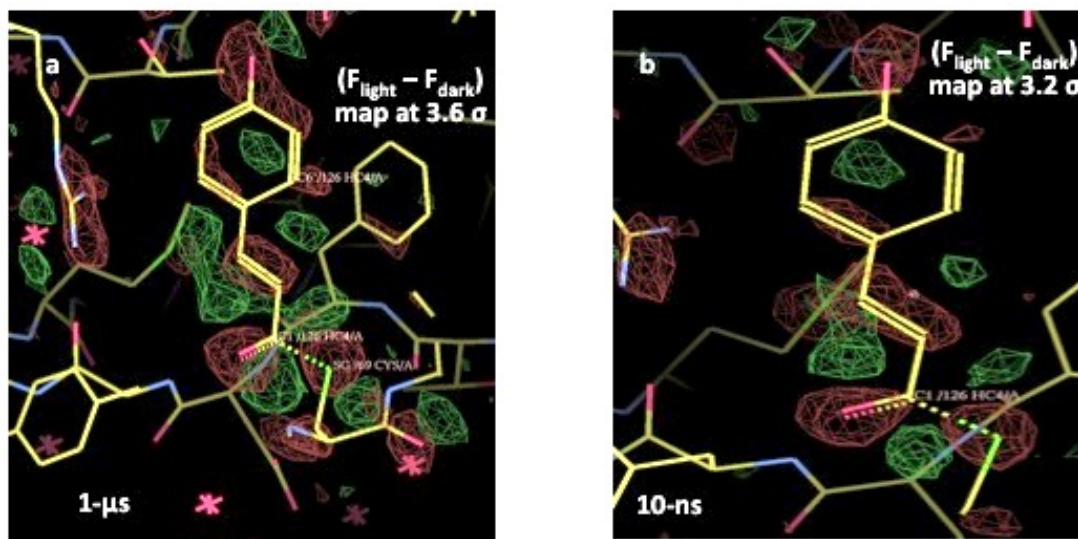
Figure 5.9: Difference Electron Density Maps ($F_{light} - F_{dark}$). These maps are from 1-μs time-point, shown in (a) and 10-ns time-point, shown in (b). The green mesh represents the positive difference density features and the red mesh stands for negative difference density features.

5.6. Effects of Indexing Ambiguities on Structural Findings

PYP crystals belong to the hexagonal lattice type in P6$_3$ space group. In this space group, pseudo-merohedral twinning, leading to indexing ambiguity can occur. Twinning is a phenomenon, where two different but very similar crystals of the same species are joined together in different orientation. However, PYP crystals are not physically twinned. The merohedral twinning in PYP crystals leads to twofold indexing ambiguities because of length of $a*$ and $b*$ being same. The indexing ambiguity is caused by two possible indices assignments to the HKL reflections – (h,k,l) and (k,h,l). Because of same lengths of $a*$ and $b*$, h and k in (h,k,l) operator can flip to result in (k,h,l). The indexing ambiguities lead to two solutions for the indexing of each pattern: convention 1 (i.e., h,k,l) and convention 2 (i.e., k,h,l). As mentioned in the data processing section 5.5, the

'ambigator' program, implemented in CrystFEL, solved the indexing ambiguity problem. It calculates correlation metric between two indexed intensities of a reflection, indexed in two different conventions and group those indexed reflections into (*h,k,l*) and (*k,h,l*) conventions. It means that for a reflection H measured in the dark state, i.e., the ground state, the structure factor amplitude denoted by D_H can be written as,

$$D_H = \langle D_H^{1,2} \rangle = \frac{\langle D_H^1 \rangle + \langle D_H^2 \rangle}{2}.$$ .... (Eq. 5.1)

The bracket operator represents the average value of the structure factor amplitude. The superscripts 1 and 2 represent convention 1 (i.e., *h,k,l*) and convention 2 (i.e., *k,h,l*) for indexing. When the indexing ambiguity is not solved, the measured amplitude for a reflection H is $\langle D_H^{1,2} \rangle$, i.e., an equal mixture of average amplitudes from both conventions. Similar to Eq. 5.1, another expression can be written for reflection H measured in light or excited state data set as,

$$L_H = \langle L_H^{1,2} \rangle = \frac{\langle L_H^1 \rangle + \langle L_H^2 \rangle}{2}.$$ .... (Eq. 5.2)

Since, this time-resolved work aimed to extract structure factor differences between ground and excited states, therefore, $\Delta F = L - D$ needed to be extracted. So, for the twinned data set, based on Eq. 5.1 and 5.2, ΔF can be written as,

$$\Delta F_H^{1,2} = \langle L_H^{1,2} \rangle - \langle D_H^{1,2} \rangle = \frac{\langle L_H^1 \rangle + \langle L_H^2 \rangle}{2} - \frac{\langle D_H^1 \rangle + \langle D_H^2 \rangle}{2} = \frac{\langle L_H^1 \rangle - \langle D_H^1 \rangle}{2} + \frac{\langle L_H^2 \rangle - \langle D_H^2 \rangle}{2}$$

$$= \frac{\Delta F_H^1}{2} + \frac{\Delta F_H^2}{2}.$$ .... (Eq. 5.3)

From the above equation, for the pseudo-merohedral twinned data set, the measured structure factor difference will be an equal mixture of two amplitude differences in two conventions. In order to calculate the difference electron density from this pseudo

merohedral twinned data based on the difference structure factor amplitudes, obtained from Eq. 5.3 and the phase information φ_H^1, obtained from the dark state model, indexed in convention 1, would be,

$$\Delta\rho(\vec{r}) =$$

$$\frac{1}{V_{cell}}\sum_{hkl}\Delta F_H^{1,2}.e^{i\varphi_H^1}.e^{2\pi\vec{q}.r} = \frac{1}{V_{cell}}\sum_{hkl}\frac{\Delta F_H^1}{2}.e^{i\varphi_H^1}.e^{2\pi\vec{q}.r} + \frac{1}{V_{cell}}\sum_{hkl}\frac{\Delta F_H^2}{2}.e^{i\varphi_H^1}.e^{2\pi\vec{q}.r} \quad \text{(Eq. 5.4)}$$

The above equation is obtained using Eq. 5.3. The 1st term in Eq. 5.4, consisting of difference structure factor amplitude and phase information, calculated based on convention 1 would contribute to the difference electron density but in the scale of half. The 2nd term, consisting of difference structure factor amplitude, calculated based on convention 2 and phase information from convention 1, would produce empty difference electron density (Fig. 5.10). Therefore, it is important to solve the indexing ambiguity so that the mixture of two different indexing conventions could be deconvoluted and difference electron density, instead of calculating in half-scale, could be calculated in 'full' scale, resulting in higher difference peaks.
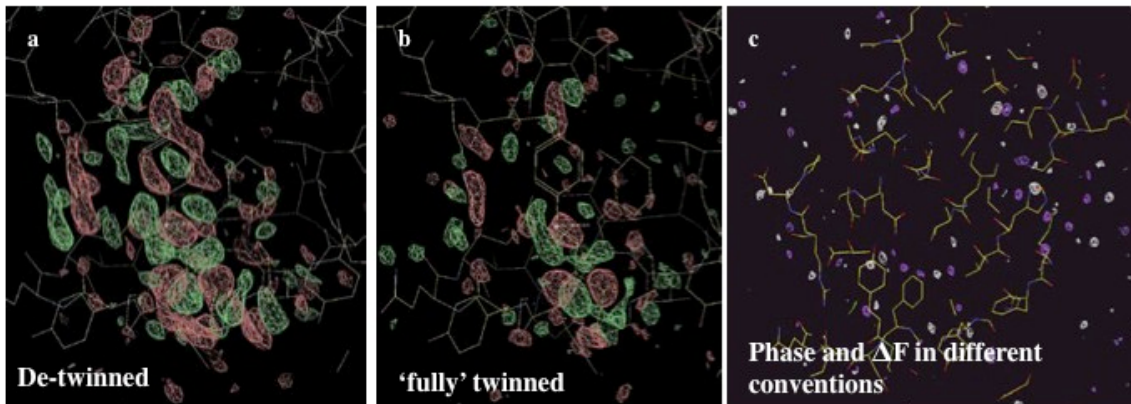
Figure 5.10: Difference Electron Density Maps, showing the Effects of Indexing Ambiguity. (a) It shows the DED map from de-twinned 1-μs data set. (b) It shows the DED map from the 'fully' twinned 1-μs data set, revealing the difference features at the half scale, representing 1st term in Eq. 5.4. (c) It shows the DED map, calculated using the difference amplitudes in convention 2 and phase information in convention 1, representing the 2nd term in Eq. 5.4.

5.7. Conformational Changes Observed between the Excited and Ground States

This section will summarize the results and observations, made from the PYP work. We were able to determine structures from the ground and the two excited states at very high spatial resolution of 1.5-Å. From the 1-μs and 10-ns – both data sets, significant conformational differences between the ground and excited states were observed (See Fig. 5.9 and Fig. 5.11). The *trans* to *cis* conformational change along with the movement of S atom of Cys residue was clearly visible from the ($F_{light} - F_{dark}$) difference electron density maps at both time-points (1-μs and 10-ns) at a contour level above 3.0-σ. The reaction progression was possible to quantify from 1-μs time point data set by calculating the occupancies of pR$_1$ and pR$_2$ intermediates (See previous section). The reaction initiation was 40% which is much higher than that achieved in Laue method. Because of the use of large macro-crystals, the laser penetration depth became too high to initiate the

171

reaction by more than 10% in time-resolved Laue Method (Schmidt et al., 2013), whereas in TR-SFX method, the use of submicron crystals with very short penetration depth (See section 5.4) excelled the reaction progression almost up to 40%. That was the key to produce high quality difference density maps that clearly unravel the structural features at the excited state. Fig. 5.11 shows the comparison between DED maps from this work (i.e., TR-SFX work) (Tenboer et al., 2014) and DED maps from the TR-Laue work (Schmidt et al., 2013) at the APS synchrotron facility. TR-SFX DED maps depict more pronounced and much better spatially connected features than those from TR-Laue DED maps. As explained before, this is due to large (40%) reaction initiation. In Fig. 5.11, all negative features of the DED maps are shown as red peaks, while all positive ones are shown as blue peaks. The yellow model represents the dark or reference structure. The pR_1 structure is shown in pink and pR_2 is shown in red. The highest positive peak at +18-σ and negative peak at -22-σ were located around sulfur atom of the Cys residue, covalently bonded to the chromophore. This implied the largest structural changes took place near the sulfur atoms.

For 10-ns data set, in Fig. 5.11, only the I_{CT} intermediate is shown in green even though 10-ns time-point represents a mixture of three different intermediates – I_T, I_{CT}, and pR_1.
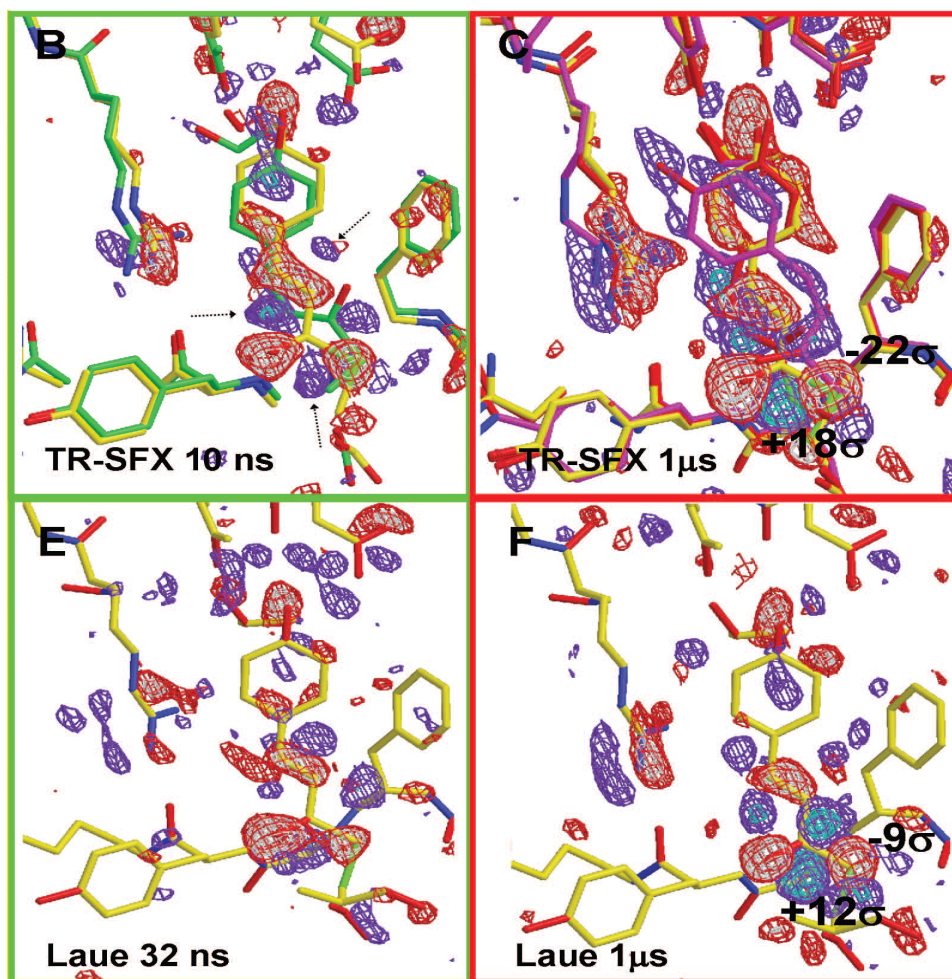
Figure 5.11: Difference Electron Density Maps (F$_{light}$ − F$_{dark}$). (a) It shows the DED map for 10-ns time-point, green model for I$_{CT}$ intermediate and yellow model from 2PHY dark model, two arrows are provided to show the direction of conformational difference features. (b) It shows the DED map for 1-μs time-point. Pink model represents pR$_1$ and red model is for pR$_2$ intermediates. The negative peak (red) around the sulfur atom at -22σ and positive peak (cyan) at +18σ are also shown. (c) and (d) show the DED maps from the TR-Laue data for comparison.

A systematic study of the minimum number of diffraction patterns and resolution required to resolve the small time-resolved structural changes in PYP was performed. The resolution of the data was truncated artificially to 3-Å and 5-Å. The DED maps were calculated based on those resolutions. The resulting DED maps (Fig. 5.12) prove that

DED features were almost faint away at 3-Å and almost disappear at 5-Å. It must be noted that the resolution required to produce a conformational change depends on the extent of the conformational change and the atoms involved. This means large conformational changes of the protein and changes in the position of heavy metals can be obtained at nominal resolution of 5-Å (Kupitz and Basu et al., 2014).
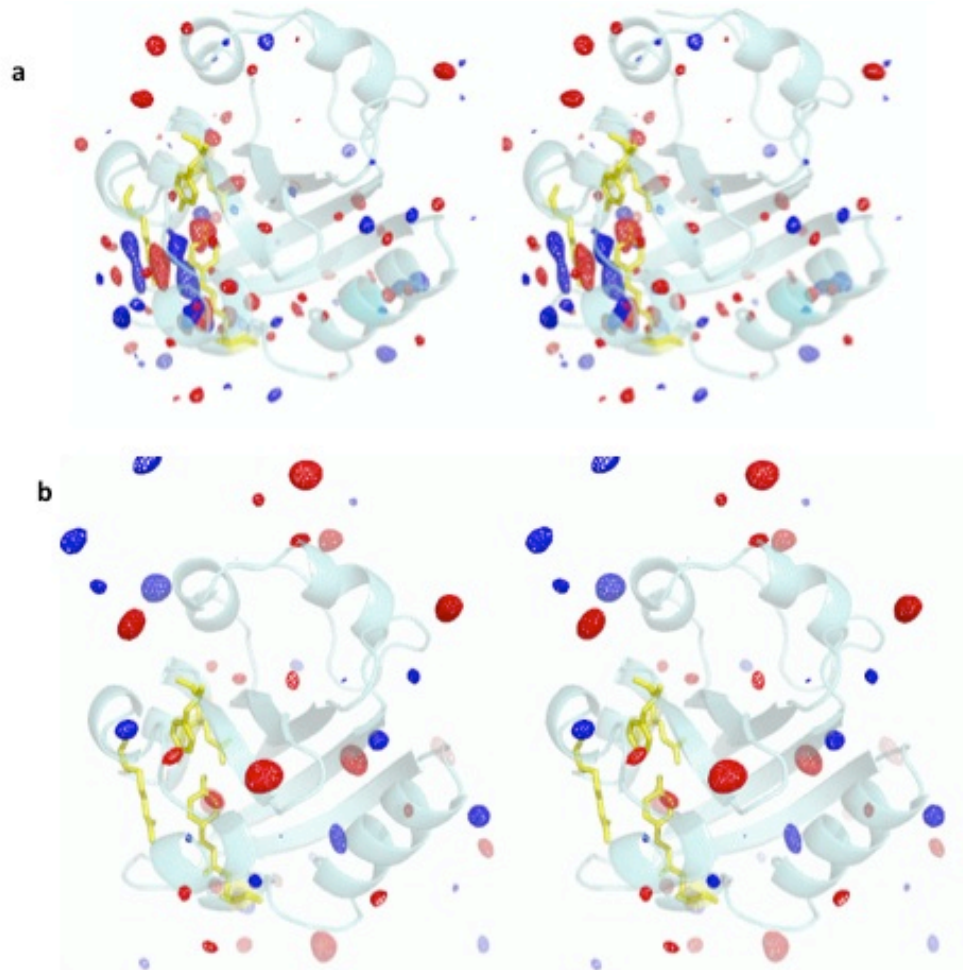
Figure 5.12: Superposition of the TR-SFX DED Map at 1-µs Time-Delay on the PYP Structure (cyan), at 3-Å and 5-Å Resolution. These images are stereo-representations. Contour levels: red/blue -3σ/+3σ. Some important chromophore pocket residues are marked in (a). (a) DED map was calculated with resolution extending to 3-Å. (b) DED map was calculated with resolution extending to 5-Å.

In another test, several DED maps were calculated based on random selection of a subset of diffraction patterns. The aim was to determine the minimum number of patterns needed to quantify the difference electron density features (See Table 5.2). Even at high R-split (defined by (White et al., 2012)) value, 4,000 diffraction patterns from the excited states were sufficient to notice the conformational change of the chromophore from *trans* to *cis*. Using only 2000 patterns from the excited state (at 1-µs time-point), movement of the sulfur atom could be seen (Fig. 5.13). But the *trans* to *cis* movement of the double bond could only be seen when 4,000 patterns were used in the data evaluation. It is noted that the minimum resolution and minimum number of diffraction patterns, required to see time-resolved changes depend on various factors – reaction progression, population in the excited states, symmetry or space group of the data set etc. PYP crystals, being of P6$_3$ space group (i.e., very high symmetry), require relatively fewer number of patterns to cover the reciprocal volume or to obtain a complete data set than what would be needed for a lower symmetry crystal lattice. This is caused by the higher order symmetry where most of the reflections are symmetry equivalents. The estimated minimum resolution range and minimum number of patterns required for time-resolved information would only be valid under these circumstances and caveats. Therefore, for other types of proteins (particularly for membrane proteins (e.g., PSII work in (Kupitz and Basu et al., 2014)), coupled with large structural changes), these numbers can vary greatly.

This time-resolved work proved the concept that high temporal resolution can be achieved by TR-SFX, establishing XFEL as a uniquely suited technique for time-resolved study of proteins. In future, it would be interesting to perform more time-resolved

experiments, pushing the limit of time-resolution higher to picosecond or femtosecond time-scale. That would open up a new avenue for studying ultrafast bio-molecular reactions.
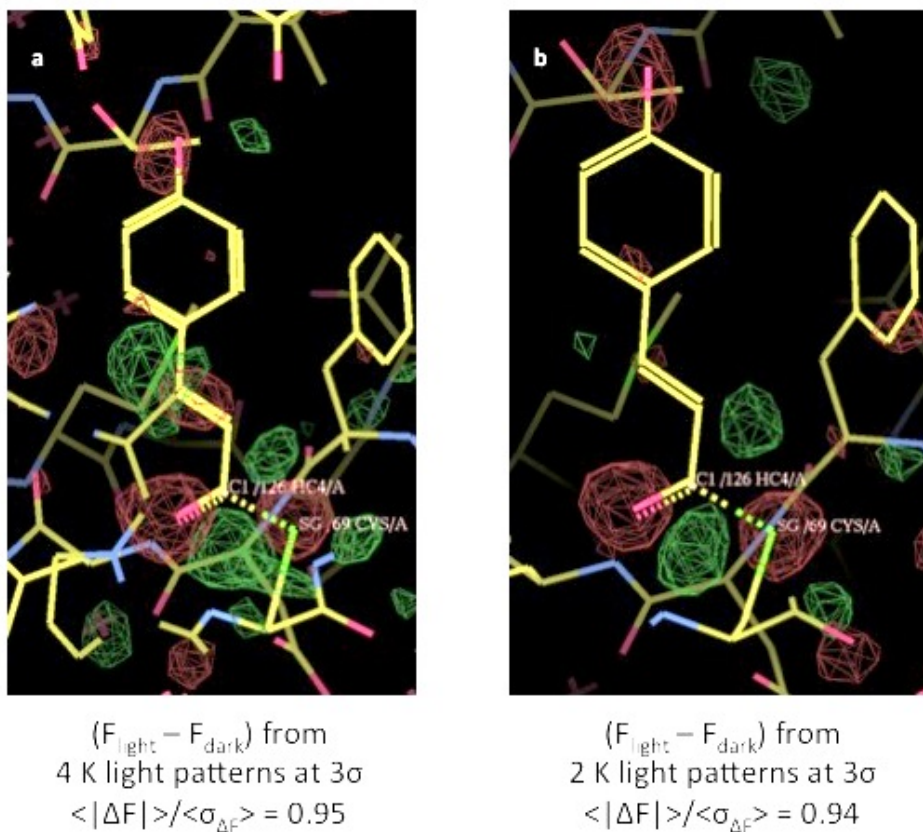
Figure 5.13: Difference Electron Density Maps with Randomly Selected 4,000 Diffraction Patterns (shown in (a)) and with 2,000 Diffraction Patterns (shown in (b)) from the 1-μs Excited State. Both maps are at 3.0-σ contour levels. It is shown that as minimum as 2,000 diffraction patterns are sufficient enough to see any conformational change.

Table 5.2: Data Statistics Based on Randomly Selected Different Fractions of Diffraction Patterns. These patterns are used to determine the minimum number of diffraction patterns required to reveal any difference features from the DED map. In addition last two columns also show the data statistics and difference structure factor amplitudes at artificially truncated resolutions.

Resolution	1.6 Å	1.6 Å	1.6 Å	1.6 Å	3.0 Å	5.0 Å
# Indexed Dark Light	32000 16000	16000 8000	8000 4000	2000 1000	64,496 31,255	64,496 31,255
R-split [%][a] Dark Light	9.8 (28.7) 13.4(50.8)	13.9 (42.9) 18.9 (73.2)	19.8 (61.2) 26.8 (95.8)	37.9 (120.6) 51.6 (209.9)	4.9 6.9	4.0 6.1
$<|\Delta F|>/<\sigma_{\Delta F}>$	7.8/7.2	9.8/9.8	22.0/18.2	23.7/24.6	14.8/13.0	13.5/14.5

6. SUMMARY OF ANCILLARY EXPERIMENTS

The development of SFX as a technique involves a large international collaboration, where our research group, led by Prof. Fromme, works with other groups across the world hand in hand on new experiments and discoveries. The expertise that I have gathered during my dissertation work on TR-SFX studies on Photosystem II and Photo-active yellow protein spans the wide range from X-ray diffraction data collection strategy, data reduction, data processing, determining structures from diffraction data, to crystallographic understanding within the SFX technique as well as conventional macromolecular crystallographic structure analysis. Since, the field of SFX is still blooming and under constant development, very few groups across the world have enough experiences and expertise in SFX. Therefore, we established collaborations with many groups on the application of SFX technique to address their important biological problems. I personally contributed to collect, process, analyze "new type" (See section 1.3) of data, collected in each of many SFX experiments performed in collaborations with many other groups. I have also contributed to a project based on electron diffraction data along with SFX data. I have collaborated during my dissertation with as many as nearly hundred scientists across the world, including working for them in beamline, elucidating optimum strategies for data collection, processing their data "on the fly", providing quick online feedback, and took the responsibilities of solving structures as quickly as possible. It is beyond the scope of this thesis to include full comprehensive descriptions of each of these projects. Thereby, in this chapter, abstracts from the selected publications along with highlighted results or major findings, have been summarized in separate sections.

6.1. Solving Protein Nanocrystals by Cryo-EM Diffraction: Multiple Scattering

Artifacts

Ganesh Subramanian, **Shibom Basu**, Haiguang Liu, Jian-Min Zuo, and John C. H. Spence **(2015)** *"Solving protein nanocrystals by cryo-EM diffraction: Multiple scattering artifacts," **Ultramicroscopy***, **148**, 87 – 93.

Abstract: The maximum thickness permissible within the single-scattering approximation for the determination of the structure of perfectly ordered protein microcrystals by transmission electron diffraction is estimated for tetragonal hen-egg lysozyme protein crystals using several approaches. Multislice simulations are performed for many diffraction conditions and beam energies to determine the validity domain of the required single-scattering approximation and hence the limit on crystal thickness. The effects of erroneous experimental structure factor amplitudes on the charge density map for lysozyme are noted and their threshold limits calculated. The maximum thickness of lysozyme permissible under the single-scattering approximation is also estimated using **R**-factor analysis. Successful reconstruction of density maps is found to result mainly from the use of the phase information provided by modeling based on the protein data base through molecular replacement (MR), which dominates the effect of poor quality electron diffraction data at thicknesses larger than about 200-Å. For perfectly ordered protein nanocrystals, a maximum thickness of about 1000-Å is predicted at 200-keV if MR can be used, using **R**-factor analysis performed over a subset of the simulated diffracted beams. The effects of crystal bending, mosaicity (which has recently been directly imaged by cryo-EM) and secondary scattering are discussed. Structure independent tests for single-scattering and new microfluidic methods for growing and sorting nanocrystals by size are reviewed (Subramanian et al., 2015).

This work addresses two important aspects in determining structures using cryo-EM diffraction methods. This work aimed to (1) determine the maximum thickness of the protein nanocrystals (based on model system lysozyme) can be permissible before multiple-scattering events become dominant and (2) to identify the significance of the tolerance to error in structure factor amplitudes (i.e., |F|) as compared to phase information.

Glaeser et al., (Glaeser et al., 1993) have evaluated the importance of multiple scattering

in electron diffraction of 2D crystals. They estimated that 100-Å thick 2D monolayer is the maximum limit at 200 keV electron beam energy. At larger thickness, the effects of multiple scattering events lead to Bragg intensities that have no relationship with single-scattering event, which is actually needed for the structure determination (Spence, 2013). The sensitivity of electron diffraction to the sample thickness has limited the use of cryo-EM diffraction in larger crystals. However, recently, Shi et al. (Shi et al., 2014), collected electron diffraction data on 0.5-μm thick lysozyme crystals and determined the structure using molecular replacement (MR) method at 3.0-Å resolution. Thereby it became important to understand whether Shi et al., were able to determine the structure either because of perfect phase information, obtained from a perfect lysozyme model or because of electron diffraction data can actually be collected at larger thickness.

The major finding of this work is the estimation of maximum permissible thickness of the perfectly ordered protein nanocrystals, which is 100-nm (or, 1000-Å) at 200-keV electron beam energy based on simulated electron diffraction data on lysozyme crystals. Here, R-factor is used as a metric to determine the permissible thickness of the sample. In Fig. 6.1.1, a plot of R-factor as a function of sample thickness at 200-keV electron beam energy shows that 100-nm would be thickness with reasonably correct R-factor (i.e., ≤ 0.3) and 230-nm would be absolute maximum limit at which R-factor becomes 0.59, resulting in arbitrary structure factor amplitudes. The lysozyme crystals have unit cell dimension of $a=b=79.1$-Å and $c = 37.8$-Å in $P4_32_12$ space group. Thereby, 100-nm thickness of lysozyme crystals will represent ~13 unit cells along a or b axes and ~25 unit cells along c direction.
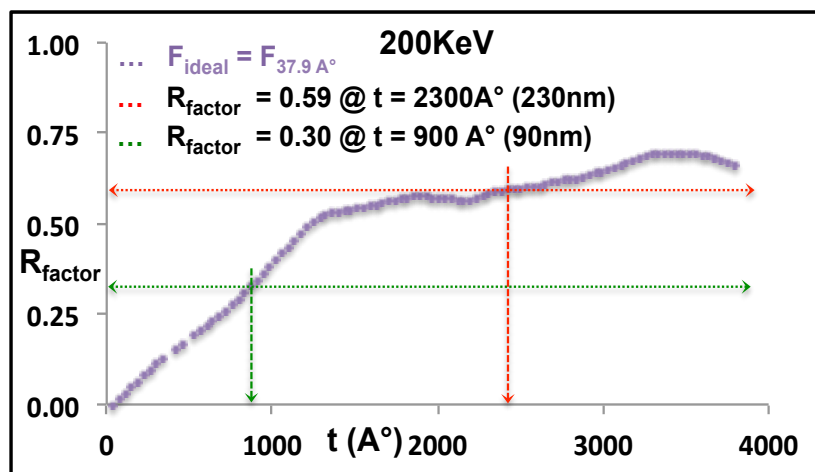
Figure 6.1.1: R-Factor Versus Thickness Plot for Lysozyme Crystals at 200-keV.

This observation questioned the electron diffraction data that has been collected by Shi et al., for 0.5-μm thick (i.e., much larger than our permissible thickness estimation) lysozyme crystals. Subsequently, it became important to understand the significance in tolerance to error in |F| when model building is performed by MR method.

Another major finding of this work is the estimation of tolerance to error in |F|, which is 34%, when phase information is provided by perfect model. We introduced random percentage of errors in structure factor amplitudes (i.e., |F|) extracted from lysozyme model 4ET8 (Boutet et al., 2012), keeping phase information unchanged. Then, we estimated the threshold error in |F|, beyond which similar amino acid sequences from hen and turkey lysozymes could not be distinguished. The amino acid sequence from hen-lysozyme contains His15 (i.e., Histidine residue at position 15 in the sequence), whereas turkey-lysozyme contains Leu15, everything else remains the same. For a model based on turkey-lysozyme and experiments based on hen-lysozyme, it would be expected that the overlap of the electron density on the model would not be perfect at the particular

181

position. In Fig. 6.1.2, $F_{obs} - F_{calc}$ difference electron density maps are shown at 3.0-σ

level, where the model is from turkey-lysozyme and |F| from hen-lysozyme. As expected,

a positive electron density (green mesh) falls outside the model at that particular position

for (a), (b), and (c) figs (corresponding to error in |F| up to 34%). In Fig. 6.1.2d,

(corresponding to error in |F| of 41%) it is noted that there is no positive electron density

falling outside the Leu15 residue in the model. This implies that at 41% error in |F|, hen-

lysozyme model cannot be distinguished from turkey-lysozyme model based on His15

residue against Leu15 residue. Thus, 41% error being maximum limit, this work suggests

34% could be permissible error limit in |F|, when building model by MR method.
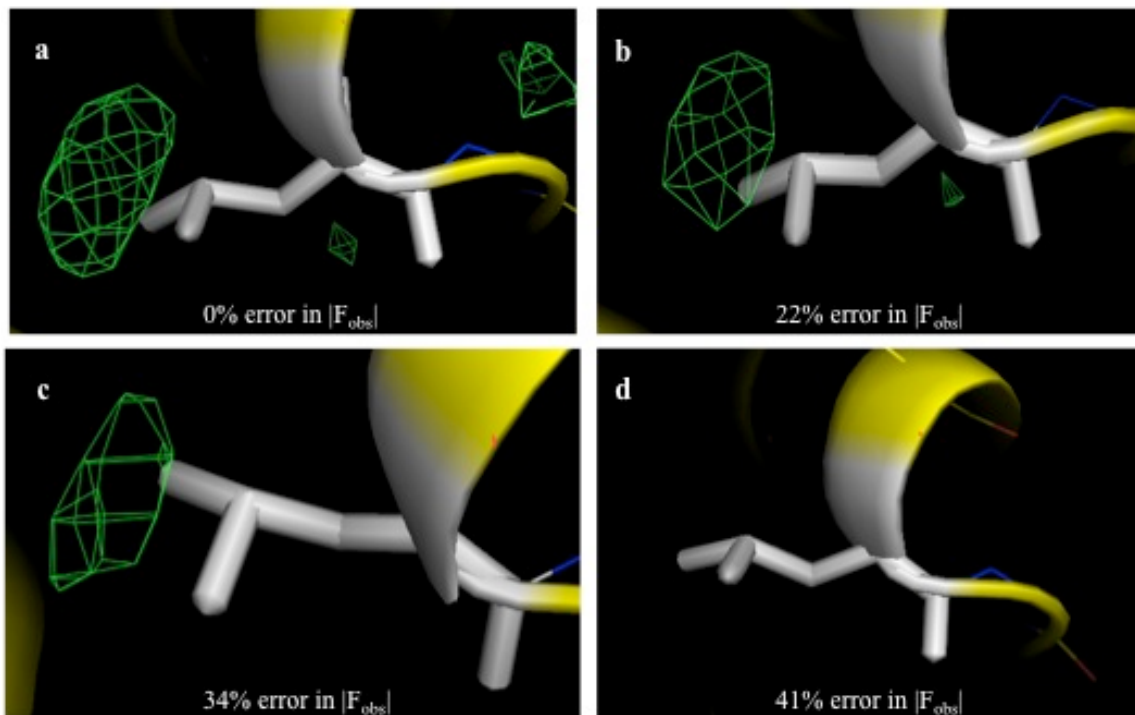
Figure 6.1.2: F_{obs}-F_{calc} Difference Maps for Turkey (Model)–Hen (Experiment) Lysozyme at 3.0-σ. Density outside the model (wire frame) indicates distinguishability, absent with 41% error in **|F|**.

However, this work established the importance of multiple scattering events in

electron diffraction method with increasing thickness of crystals, leading to erroneous structure factor amplitudes. The error in amplitudes can easily be overshadowed by the correct phase information during model building, producing erroneous structures with multiple scattering artifacts.

6.2. Serial Femtosecond Crystallography of G-protein Coupled Receptors

Wei Liu, Daniel Wacker, Cornelius Gati, Gye Won Han, Daniel James, Dingjie Wang, Garrett Nelson, Uwe Weierstall, Vsevolod Katritch, Anton Barty, Nadia A. Zatsepin, Dianfan Li, Marc Messerschmidt, Sébastien Boutet, Garth J. Williams, Jason E. Koglin, M. Marvin Seibert, Chong Wang, Syed T. A. Shah, **Shibom Basu**, Raimund Fromme, Christopher Kupitz, Kimberley N. Rendek, Ingo Grotjohann, Petra Fromme, Richard A. Kirian, Kenneth R. Beyerlein, Thomas A. White, Henry N. Chapman, Martin Caffrey, John C. H. Spence, Raymond C. Stevens, Vadim Cherezov (**2013**) *"Serial femtosecond crystallography of G protein-coupled receptors," **Science**, **342**: 1521 – 1524.*

Abstract: X-ray crystallography of G protein–coupled receptors and other membrane proteins is hampered by difficulties associated with growing sufficiently large crystals that withstand radiation damage and yield high-resolution data at synchrotron sources. We used an x-ray free-electron laser (XFEL) with individual 50-femtosecond-duration x-ray pulses to minimize radiation damage and obtained a high-resolution room-temperature structure of a human serotonin receptor using sub-10-micrometer microcrystals grown in a membrane mimetic matrix known as lipidic cubic phase. Compared with the structure solved by using traditional microcrystallography from cryo-cooled crystals of about two orders of magnitude larger volume, the room-temperature XFEL structure displays a distinct distribution of thermal motions and conformations of residues that likely more accurately represent the receptor structure and dynamics in a cellular environment (Liu et al., 2013)

This was the first work, where Lipidic Cubic Phase (LCP) was successfully used as a delivery medium for SFX data collection at the XFEL facility. Readers are referred to reviews on LCP for details (Caffrey and Cherezov 2009). There are plenty different and biologically challenging membrane proteins which are extremely difficult to crystallize in conventional methods with detergent micelles, but could be grown in LCP medium. Particularly, proteins from the GPCR family are challenging to grow as large crystals

even in the LCP medium. Thereby, it found its niche in SFX method, where micron or submicron sized crystals are sufficient enough to produce undamaged structure determination at the room temperature. This work's major finding includes the revealing of different dynamic conformations at the room temperature. A very experienced group, led by Vadim Cherezov, performed a SFX experiment (Liu et al., 2013) on micron-sized crystals (average size of 5-μm) of G protein-coupled receptors (GPCR), namely serotonin receptor at the CXI instrument of the LCLS on March, 2013.

This work is very important not only because it provides structural difference between room temperature and cryogenic conformations, but it also paves a new way to GPCR community to solve high-resolution structure using micro-crystals, grown in Lipid Cubic Phase (LCP) medium through SFX. The importance and the involvement of GPCRs in key physiological and sensory processes are well known to the scientific community. Amongst many other membrane proteins, GPCRs are extremely important drug targets but their structural studies are still highly challenging due to poor yield upon protein expression, poor stability of the receptor, and the conformational heterogeneity. Thereby, all these together make the entire crystallization process extremely difficult. So far, most of the GPCRs crystals are grown in a membrane-mimetic environment, lipidic cubic phase (LCP). This crystallization method can produce high-resolution and well diffracting crystals, but limited in size, which often hinders to get complete data set at high-resolution at synchrotron facility, even including micro-focus beamlines. In course of time, SFX came up as a blessing for the GPCRs community, because it is aimed at the use of nano-/micron-sized crystals. In this work, average size of 5-μm crystals was used.

184

For this work, a specific type of new injection system was developed based on LCP medium (Weierstall et al., 2014). This delivery system was designed to handle honey-like viscous medium, unlike previously used liquid jet injector. The development of this injector helped to reduce sample consumption from the order of mg to that of μg, which is a significant development (discussed in section 6.2). In this work, the protein consumption was 300-μg. As a result, this experiment produced 2.8-Å structure of $5HT_{2B}$, serotonin receptor at room temperature. This work also compares two structures, obtained from XFEL and Synchrotron – $5HT_{2B\text{-}XFEL}$ and $5HT_{2B\text{-}SYN}$. Both structures were solved in the same space-group $C222_1$. Room temperature XFEL structure had a bigger lattice parameter (slightly longer in *a* and *b* directions), leading to 2.1% increment in the unit cell size. Consequently, a rotation of 2.5° was noticed at the BRIL fusion domain at the receptor. The major findings include the displacement in the intracellular and extracellular loop regions (ICL and ECL). In $5HT_{2B\text{-}XFEL}$ structure, at the extracellular tip of helix II, Thr114 residue forms a stabilizing H-bond with the carbonyl oxygen of Ile110 residue, whereas in the $5HT_{2B\text{-}SYN}$ structure, a water-mediated kink was observed at this location (See Fig. 6.2.1). This deviation resulted in 2.0-Å movement of Cα at the tip of helix II and 3.4-Å movements in the ECL1 (Liu et al., 2013)(Ravelli and McSweeney, 2000). The room temperature XFEL structure has 21-$Å^2$ larger B-factor compared to cryogenic structure. This can be accounted as a consequence of thermal motion present in the XFEL structure due to room temperature. Interestingly, more mobile loop regions, i.e., ICL and ECL, had B-factor difference of close 100-$Å^2$. From the structural studies of cryogenic structures, it was well established that main structural difference between two

185

subtypes of serotonin receptors – $5HT_{1B}$ and $5HT_{2B}$ occurs at the extracellular tip helix V and ECL2, forming an additional helical turn, mediated by water molecule in $5HT_{2B}$ (Liu et al., 2013). But room temperature XFEL structure provides an extensive network of H-bond rather than water molecule stabilizes the conformation of ECL2. This observation concludes that perhaps water molecule does not play any structural role at this location. Thus, this work summarizes some major and important structural findings after solving $5HT_{2B}$ structure at 2.8-Å and also paves a path to GPCRs society, leading to determining high-resolution structure using submicron-sized crystals at XFEL facility.
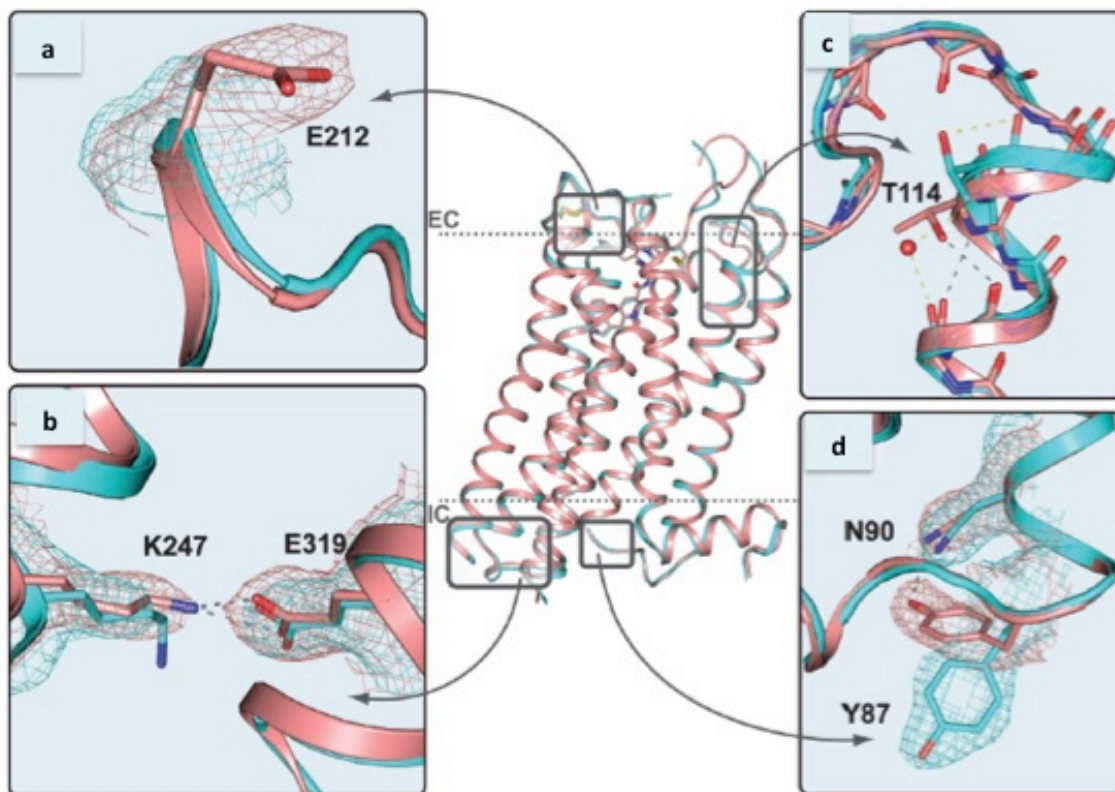
Figure 6.2.1: Comparison between Room-Temperature XFEL Structure (5HT$_{2B-}$ $_{XFEL}$) (light red) and Cryo-Cooled Synchrotron Structure (5HT$_{2B-SYN}$) (teal). Central image represents a backbone overlay of the two structures. Dashed lines correspond to membrane boundaries defined by the Orientation of Proteins in Membrane database (http://opm.phar.umich.edu). (a) Electron density for the Glu212 side chain is missing in 5-HT$_{2B-SYN}$ and fully resolved in 5-HT$_{2B-XFEL}$. (b) A salt bridge between Glu319 and Lys247 links intracellular parts of helices V and VI in the 5-HT$_{2B-XFEL}$ structure. In the 5-HT$_{2B-SYN}$ structure, Lys247 makes a hydrogen bond with Tyr1105 from the BRIL fusion protein. (c) Extracellular tip of helix II forms a regular helix in 5-HT$_{2B-XFEL}$ with Thr114, making a stabilizing hydrogen bond with the backbone carbonyl, whereas in 5-HT$_{2B-SYN}$, one water-stabilized kink is introduced at this position. (d) Tyr87 forms a hydrogen bond with Asn90 in 5-HT$_{2B-XFEL}$; this hydrogen bond is broken, and Tyr87 adopts a different rotamer conformation in the 5-HT$_{2B-SYN}$ structure. 2mF$_{obs}$-DF$_{calc}$ maps (contoured at 1σ level) are shown only around described residues (Source: Liu et al., 2013).

6.3. Lipidic Cubic Phase Injector Facilitates Membrane Protein – SFX

Uwe Weierstall, Daniel James, Chong Wang, Thomas A. White, Dingjie Wang, Wei Liu, John C.H. Spence, R. Bruce Doak, Garrett Nelson, Petra Fromme, Raimund Fromme, Ingo Grotjohann, Christopher Kupitz, Nadia A. Zatsepin, Haiguang Liu, **Shibom Basu**,

Daniel Wacker, Gye Won Han, Vsevolod Katritch, Se´bastien Boutet, Marc Messerschmidt, Garth J. Williams, Jason E. Koglin, M. Marvin Seibert, Markus Klinker, Cornelius Gati, Robert L. Shoeman, Anton Barty, Henry N. Chapman, Richard A. Kirian, Kenneth R. Beyerlein, Raymond C. Stevens, Dianfan Li1, Syed T.A. Shah, Nicole Howe, Martin Caffrey, Vadim Cherezov, **(2014)** *"Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography,"* **Nature Comm**. **5**: 3309.

Abstract: Lipidic cubic phase (LCP) crystallization has proven successful for high-resolution structure determination of challenging membrane proteins. Here we present a technique for extruding gel-like LCP with embedded membrane protein microcrystals, providing a continuously renewed source of material for serial femtosecond crystallography. Data collected from sub- 10-μm-sized crystals produced with less than 0.5-mg of purified protein yield structural insights regarding cyclopamine binding to the Smoothened receptor (Weierstall et al., 2014).

This work was the major driving force for the success of the research work described in the previous section 6.2. This work is based on designing a new type of sample injection system, developed using the principle of hydrodynamics to deliver crystals embedded in Lipidic cubic phase (LCP). LCP is a medium with very high viscosity (i.e., honey like consistency). There are many challenging membrane proteins (such as GPCR family), which cannot be crystallized in conventional detergent micelles method. LCP providing the necessary support for those membrane proteins growth is key to successfully crystallize them. In SFX, previously, liquid injector (Gas dynamic virtual nozzle, GDVN) was used. Liquid injector works with water medium with very high flow-rate, resulting in large sample consumption (~10-ml per experiment; See Table 6.3.1). This type of viscous jet could reduce the sample consumption from ml to μl of samples per structure. The design of the viscous jet is shown in Fig. 6.3.1. The paper describes not only the design of the viscous jet but also shows the structure determination of Smoothened Receptor in complex with the drug cyclopamine. The groundbreaking work has the most significant impact toward reduction in sample consumption, which is the

major disadvantage of the SFX technique. As long as the protein (soluble or membrane) is able to survive in the LCP medium, such viscous jet can easily be used for the SFX data collection.
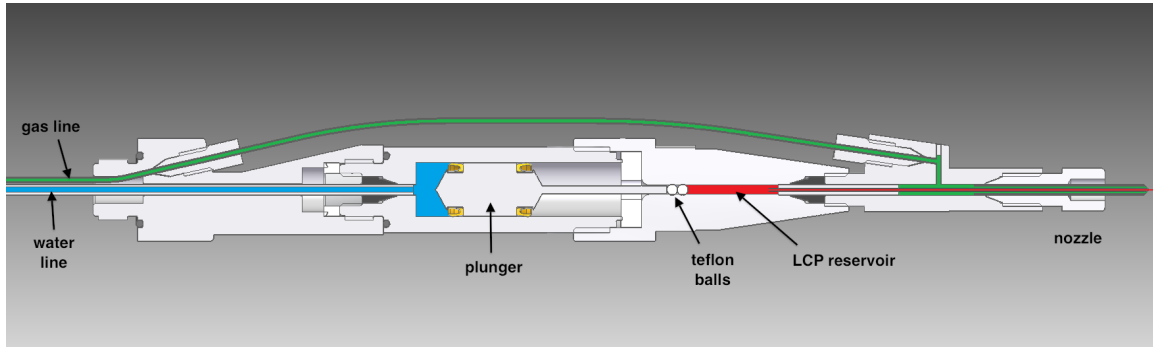
Figure 6.3.1: A Middle Section Through the LCP Injector. In operation, the device is attached via the leftmost threaded fitting to a nozzle rod (not shown) for insertion into the experimental chamber. Water (blue) and gas (green) lines are routed through the nozzle rod from the left, LCP (red) is extruded out from the nozzle on the right. Water at a pressure of up to 300-psi drives the hydraulic plunger, which amplifies the pressure 34 times to drive LCP through a capillary with an inner diameter of 10-50-µm. Two Teflon spherical beads are used to provide a tight seal against a pressure of up to 10,000-psi. The co-flowing gas is necessary for reliable extrusion and to maintain co-axial flow.

Table 6.3.1: Comparison of GDVN and LCP Injectors for Sample Consumptions and Hit Rates for Different SFX Experiments.

Protein	PS I	Lysozyme	5-HT$_{2B}$	DgkA
Crystal growth medium	Liquid	Liquid	LCP	LCP
Injector	GDVN	GDVN	LCP	LCP
Total frames recorded	1,850,000	1,471,615	4,217,508	1,987,632
Crystal hits[*]	112,725	66,442	152,651	263,435
Indexed hits	15,445	12,247	32,819	66,165
Hit rate, %	6	4.5	3.6	13
Indexing rate %	13.7	18.4	21.5	25
Sample consumption (liquid volume and protein mass)	10,000-µL 10-mg	10,800-µL 15-mg	100-µL 0.30-mg	46-µL 0.30-mg
Flow rate in µL/min	10	30	0.17	0.17
Consumed Protein per 10,000 indexed patterns	6.5-mg	12.2-mg	0.091-mg	0.045-mg

6.4. SFX Data Processing of Phycocyanin Crystals in Liquid Media and LCP

Raimund Fromme, Andrii Ishchenko, Markus Metz, Shatabdi Roy Chowdhury, **Shibom Basu**, Se´bastien Boutet, Petra Fromme, Thomas A. White, Anton Barty, John C. H. Spence, Uwe Weierstall, Wei Liu and Vadim Cherezov **(2015)** *"Serial Femtosecond Crystallography of Soluble Protein in Lipidic Cubic Phase," IUCrJ*, **2**, 545-551.

Abstract: Serial Femtosecond Crystallography (SFX) at X-ray Free-Electron Lasers (XFELs) enables high-resolution protein structure determination using micron and submicron-sized crystals at room temperature with minimal effects of radiation damage. SFX requires a steady supply of microcrystals intersecting the XFEL beam at random orientation. We have recently introduced an LCP-SFX method in which microcrystals of membrane proteins are grown and delivered for SFX data collection inside a gel-like membrane mimetic matrix, known as lipidic cubic phase (LCP) using a special LCP microextrusion injector. Here we demonstrate that LCP can also be used as a suitable carrier medium for microcrystals of soluble proteins enabling dramatic reduction in the amount of crystallized protein required for data collection as compared to crystals delivered by liquid injectors. High quality LCP-SFX data sets were collected for two

soluble proteins, lysozyme and phycocyanin using less than 0.1-mg of protein (Fromme et al., 2015).

This work shows the successful application of LCP injector to collect data on crystals, which were not grown in LCP medium. The micro-crystals (mean size of 10-μm) of Phycocyanin, a photosynthetic pigment protein from *Thermosynechococcus Elongatus*, were not grown in LCP medium but before injection, they were mixed into LCP medium and data were collected. This work shows results from two experiments – both were carried out in February 2014 at the CXI instrument in the LCLS. The first experiment was performed to collect data from Phycocyanin (PC) microcrystals, delivered in GDVN liquid injectors. Later another experiment was performed to collect data from PC microcrystals, delivered by LCP injectors. This is the first work, showing the stability and survival of a soluble protein (i.e., PC) in LCP medium along with successful data collection. The structure of PC from LCP medium was determined based on data collected in only 40-min, consuming only 3-μL of crystal suspension or 0.-mg protein. A comparison between PC structures from liquid jet and the LCP jet along with that from macro-crystallography was shown in Table 6.4.1.

Table 6.4.1: XFEL Data Collection and Refinement Statistics for Phycocyanin Crystals. This table compares data statistics from Macro-crystals, liquid jet, and LCP jet.

	Macro-crystal (3L0F)	GDVN injector	LCP injector
Data collection			
Average crystal size (μm)	200×200×300	10×10×5	10×10×5
Amount of protein used (mg)	N/A	30	0.1
Wavelength (Å)	1.00	1.45	1.56
Space group	H32	H32	H32
Resolution (Å)	19.0 – 1.35 (1.42-1.35)	36.4 – 1.95 (2.02-1.95)	31.6 – 1.75 (1.81-1.75)
Cell dimensions (Å)	187.1, 187.1, 59.8	186.4, 186.4, 60.3	187.1, 187.1, 60.5
No. hits / indexed images	N/A	36,118/16,689	18,794/6,629
No. total / unique reflections	600,252/86,960	7,520,260/32,291	6,171,418/44,284
I/σ(I)	19.15 (2.35)	2.92 (1.44)	2.94 (1.33)
Multiplicity	7.0(5.9)	358.63 (47.9)	139.36 (36.8)
Completeness, %	99.9 (99.7)	99.98 (99.8)	99.97 (100)
$CC_{1/2}$	0.999(0.799)	0.975 (0.34)	0.951 (0.32)
R_{split}, %	NA	31.52 (97.26)	39.05 (94.22)
R_{merge}, %	7.7 (84.9)	N/A	N/A
Refinement			
No. reflections	82,594 (6,097)	32,286 (3,188)	40,257 (4,000)
Test set	4,367 (296)	1,660 (176)	1,838 (183)
R_{work} / R_{free}, %	13.6(27.1)/17.5(33.1)	24.0(37.5)/28.7(41.9)	19.3 (50.4)/25.2(50.2)
No. atoms			
Protein	2,497	2,497	2,497
Water and other	573	258	300
B-factors ($Å^2$)			
Wilson B/ Overall B	16.3/16.5	15.3/24.1	33.8/37.5
Protein	20.7	23.9	37.0
Water and other	41.5	27.8	45.6
R.m.s bonds, Å / angles (°)	0.007/1.58	0.024/1.29	0.012/1.24
Ramachandran plot stats, %			
Favored	98.2	98.2	98.5
Allowed	1.8	1.8	1.5
Disallowed	0	0	0

6.5. A New Viscous Delivery Medium – Agarose for SFX Technique

Chelsie E. Conrad, **Shibom Basu**, Daniel James, Dingjie Wang, Alexander Schaffer, Shatabdi Roy-Chowdhury, Nadia A. Zatsepin, Andrew Aquila, Jesse Coe, Cornelius Gati, Mark S. Hunter, Jason E. Koglin, Christopher Kupitz, Garrett Nelson, Ganesh Subramanian, Thomas A. White, Yun Zhao, James Zook, Se´bastien Boutet, Vadim Cherezov, John C. H. Spence, Raimund Fromme, Uwe Weierstall and Petra Fromme, **(2015)** *"A novel inert crystal delivery medium for XFEL serial femtosecond crystallography," IUCrJ*, **2**, 421-430.

Abstract: Serial femtosecond crystallography (SFX) opened a new era in crystallography by permitting nearly damage-free, room temperature structure determination of challenging proteins, such as membrane proteins. In SFX, femtosecond X-ray free-electron laser pulses produce diffraction snapshots from nano and microcrystals delivered in a liquid stream, which leads to high protein consumption. We developed a slow-moving stream of agarose as a new crystal delivery medium for SFX. It has low background scattering, is compatible with both soluble and membrane proteins, and can deliver the protein crystals at a wide range of temperatures, down to 4 °C. Using this crystal-laden agarose stream, the structure of a multi-subunit complex, phycocyanin, was solved to 2.5-Å using 300-μg of microcrystals embedded into the agarose-medium post-crystallization. The agarose delivery method reduces protein consumption by at least 100 fold and has the potential to be used for a diverse population of proteins including membrane protein complexes (Conrad et al., 2015).

This work addressed the challenge owing to large sample consumption in SFX data collection. It showcased agarose as a new delivery medium, which is as viscous as LCP medium, but also inert, providing the stability to a wide range of various proteins, including membrane and soluble proteins. Agarose medium renders not only stability to various membrane proteins, including photosystem I and II, which were unstable in LCP medium, but also generate much less background scattering compared to other delivery media such as LCP. This work demonstrates the application of agarose in collecting a complete data set from phycocyanin (PC) micro-crystals in 72-min. The structure of PC

was determined at 2.5-Å in hexagonal P6$_3$ space group with different unit cell dimensions (a=b=153.4-Å and c=39.6-Å) compared to those from LCP or aqueous media.

In this work, the background scattering of agarose was compared with that of LCP medium quantitatively, resulting in 2.3 times less background noise from agarose medium than that from LCP in the diffused ring regions (See Fig. 6.5.1). The average scattered intensity from each medium delivered in a stream of the same width (50-μm) using detector readout events that contained no crystal diffraction was calculated. 13,902 frames from the agarose stream and 14,592 frames from the LCP stream data were analyzed. Frames that contained no scattering from the jets (owing to the jet temporarily fluctuating out of the path of the X-rays) were easily recognized on the basis of their very low photon counts (~10-20 detector units), and were excluded from the mean background calculation, leaving 9,147 and 8,326 frames with scattering from the LCP and agarose jets, respectively. Thereby, bias from large jet flow instabilities was avoided in the calculation of the mean radial intensities for each medium. To reduce the influence of shot-to-shot variations in the XFEL pulse intensities, each frame was scaled to the readings from the gas ionization detector upstream of the vacuum chamber at CXI, which is accurate to ~ 10%. Finally, the mean radial intensities from the LCP and agarose jets were scaled to be equal at a resolution of 2-Å, where neither medium should produce a background signal. As shown in Fig. 6.5.1, a broad peak corresponding to diffuse scattering from the lipid chains of LCP can be seen at 4.5-Å resolution. Diffuse scattering from agarose can be seen in the 3.3-Å region. The shaded regions represent the mean absolute deviation around the mean.

This paper extensively discusses the compatibility of the agarose medium with various crystallization conditions as well as temperatures. In addition to less sample consumption, it can also be used in vacuum and at ambient pressure. This injector is not as stable as LCP injector. However, the main attraction of the agarose injector is its low background scattering, particularly near 4.5- to 3.5-Å resolution range, where LCP and other delivery media render strong diffused ring. Putting all positive features together, this new delivery media, being inert to many proteins, may allow to investigate structures of rare proteins with low expression yields and hard to crystallize and extends the SFX methods.
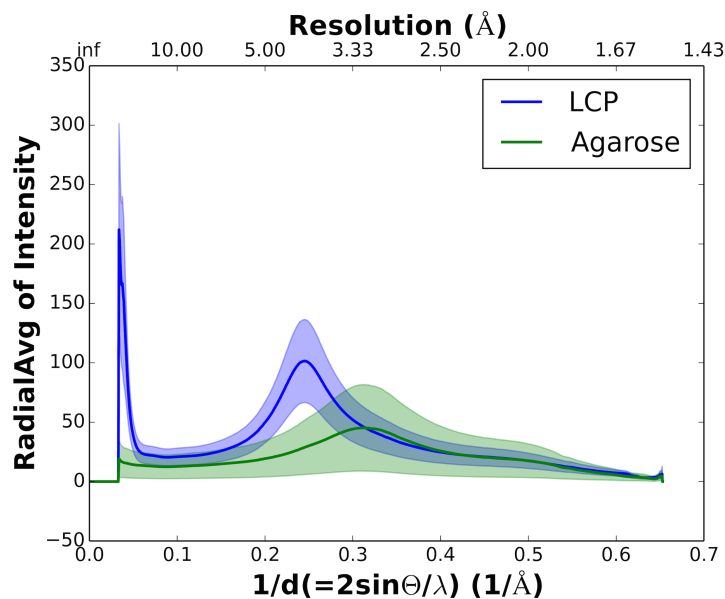
Figure 6.5.1: Diffuse Background Scattering Comparison between Agarose and Lipidic Cubic Phase. The figure shows $1/d$ in (Å^{-1}) on x-axis and mean radial intensity over total number of used frames from each medium on the y-axis. The blue line represents the mean radial intensity for LCP media as a function of $1/d$ (or, resolution in Å unit in second x-axis). The green line represents the mean radial intensity for agarose as a function of $1/d$ as well. The error or fluctuation in the radial intensity is quantified using mean absolute deviation for both media, shown in transparent region.

195

7. CONCLUSIONS AND FUTURE OUTLOOK

The time-resolved structures of Photosystem II presented here are the first ones showing interpretable conformational changes. The time-resolved work on Photo-active Yellow Protein (PYP) establishes XFEL as a technique for time-resolved studies at high spatial and temporal resolution. Both of these works have established significant advancements in science by indisputably demonstrating the practical utility of the TR-SFX method. These studies build a solid foundation for further advancement of time-resolved studies using SFX for basic science examining ultra-fast enzymatic processes.

SFX is still a new and emerging technique that is much in need of many advancements to support hardware and software. In future, this research would be directed to work on the development of novel methods for analysis of SFX data that is based on combining continuous diffused scattering and single-particle approach of model reconstruction. This new data analysis method will in the future enable SFX to increase attainable resolution and to yield independent phase information. In order to explore such novel data analysis approach, our team has been awarded beamtimes in the end of October 2015. Application of this new data evaluation strategy, in the future, promises to improve the information content extraction data for existing and future PSII data, thereby revealing heretofore unseen time-resolved structural changes at new levels of temporal and spatial resolution.

My research has also paved the way for new experiments, which endeavor to study ultra-fast enzymatic processes. Towards this end, I am a central member of a team that has been awarded beamtimes for such experiments in December 2015. If successful,

these experiments will constitute a breakthrough demonstrating TR-SFX as a technique enabling mankind to peer into processes never before seen (much as the first microscopes revealed an unseen world).

In summary, the development of SFX has advanced in recent years in dramatic fashion through landmark demonstrations. To fully realize the power of TR-SFX will require many more years of dedicated development by the scientific community. Nevertheless, the future of structural biology has forever been changed, and will continue to evolve with evolutionary advancements of TR-SFX. This is particularly true as new XFEL facilities become operational in next 4 years, including the groundbreaking attosecond free-electron laser facility at Arizona State University.

REFERENCES

Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (4), 352–367.

Ames, W.; Pantazis, D. A.; Krewald, V.; Cox, N.; Messinger, J.; Lubitz, W.; Neese, F. Theoretical Evaluation of Structural Models of the S2 State in the Oxygen Evolving Complex of Photosystem II: Protonation States and Magnetic Interactions. *J. Am. Chem. Soc.* **2011**, *133* (49), 19743–19757.

Aquila, A.; Hunter, M. S.; Doak, R. B.; Kirian, R. A.; Fromme, P.; White, T. A.; Andreasson, J.; Arnlund, D.; Bajt, S.; Barends, T. R. M.; et al. Time-Resolved Protein Nanocrystallography Using an X-Ray Free-Electron Laser. *Opt. EXPRESS* **2012**, *20* (3), 2706–2716.

Atkins, P. W.; De Paula, J. *Physical Chemistry*, 7th ed.; W.H. Freeman: New York, 2002.

Barty, A.; Caleman, C.; Aquila, A.; Timneanu, N.; Lomb, L.; White, T. A.; Andreasson, J.; Arnlund, D.; Bajt, S.; Barends, T. R. M.; et al. Self-Terminating Diffraction Gates Femtosecond X-Ray Nanocrystallography Measurements. *Nat. PHOTONICS* **2012**, *6* (1), 35–40.

Barty, A.; Kirian, R. A.; Maia, F. R. N. C.; Hantke, M.; Yoon, C. H.; White, T. A.; Chapman, H. Cheetah: Software for High-Throughput Reduction and Analysis of Serial Femtosecond X-Ray Diffraction Data. *J. Appl. Crystallogr.* **2014**, *47* (Pt 3), 1118–1131.

Bhat, T. N. Calculation of an OMIT Map. *J. Appl. Crystallogr.* **1988**, *21* (3), 279–281.

Bolduc, J. M.; Dyer, D. H.; Scott, W. G.; Singer, P.; Sweet, R. M.; Koshland, D. E.; Stoddard, B. L. Mutagenesis and Laue Structures of Enzyme Intermediates: Isocitrate Dehydrogenase. *Science* **1995**, *268* (5215), 1312–1318.

Borgstahl, G. E.; Williams, D. R.; Getzoff, E. D. 1.4 A Structure of Photoactive Yellow Protein, a Cytosolic Photoreceptor: Unusual Fold, Active Site, and Chromophore. *Biochemistry (Mosc.)* **1995**, *34* (19), 6278–6287.

Bourgeois, D.; Wagner, U.; Wulff, M. Towards Automated Laue Data Processing: Application to the Choice of Optimal X-Ray Spectrum. *Acta Crystallogr. D Biol. Crystallogr.* **2000**, *56* (8), 973–985.

Bourgeois, D.; Vallone, B.; Schotte, F.; Arcovito, A.; Miele, A. E.; Sciara, G.; Wulff, M.; Anfinrud, P.; Brunori, M. Complex Landscape of Protein Structural Dynamics Unveiled by Nanosecond Laue Crystallography. *Proc. Natl. Acad. Sci.* **2003**, *100* (15), 8704–8709.

Bourgeois, D.; Schotte, F.; Brunori, M.; Vallone, B. Time-Resolved Methods in Biophysics. 6. Time-Resolved Laue Crystallography as a Tool to Investigate Photo-Activated Protein Dynamics. *Photochem. Photobiol. Sci.* **2007**, *6* (10), 1047.

Boutet, S.; J Williams, G. The Coherent X-Ray Imaging (CXI) Instrument at the Linac Coherent Light Source (LCLS). *New J. Phys.* **2010**, *12* (3), 035024.

Boutet, S.; Lomb, L.; Williams, G. J.; Barends, T. R. M.; Aquila, A.; Doak, R. B.; Weierstall, U.; DePonte, D. P.; Steinbrener, J.; Shoeman, R. L.; et al. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* **2012**, *337* (6092), 362–364.

Brehm, W.; Diederichs, K. Breaking the Indexing Ambiguity in Serial Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70* (1), 101–109.

Brünger, A. T. Free R Value: A Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures. *Nature* **1992**, *355* (6359), 472–475.

Caffrey, M.; Cherezov, V. Crystallizing Membrane Proteins Using Lipidic Mesophases. *Nat. Protoc.* **2009**, *4* (5), 706–731.

Chapman, H. N.; Barty, A.; Bogan, M. J.; Boutet, S.; Frank, M.; Hau-Riege, S. P.; Marchesini, S.; Woods, B. W.; Bajt, S.; Benner, W. H.; et al. Femtosecond Diffractive Imaging with a Soft-X-Ray Free-Electron Laser. *Nat. Phys.* **2006**, *2* (12), 839–843.

Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U.; et al. Femtosecond X-Ray Protein Nanocrystallography. *NATURE* **2011**, *470* (7332), 73–77.

Chu, H. A.; Nguyen, A. P.; Debus, R. J. Amino Acid Residues That Influence the Binding of Manganese or Calcium to Photosystem II. 1. The Lumenal Interhelical Domains of the D1 Polypeptide. *Biochemistry (Mosc.)* **1995**, *34* (17), 5839–5858.

Conrad, C. E.; Basu, S.; James, D.; Wang, D.; Schaffer, A.; Roy-Chowdhury, S.; Zatsepin, N. A.; Aquila, A.; Coe, J.; Gati, C.; et al. A Novel Inert Crystal Delivery Medium for Serial Femtosecond Crystallography. *IUCrJ* **2015**, *2* (4), 421–430.

Cox, N.; Pantazis, D. A.; Neese, F.; Lubitz, W. Biological Water Oxidation. *Acc. Chem. Res.* **2013**, *46* (7), 1588–1596.

Dau, H.; Ztharieva, I.; Haumann, M. Recent Developments in Research on Water Oxidation by Photosystem II. *Curr. Opin. Chem. Biol.* **2012**, *16* (1-2), 3–10.

Davis, K. M.; Mattern, B. A.; Pacold, J. I.; Zakharova, T.; Brewe, D.; Kosheleva, I.; Henning, R. W.; Graber, T. J.; Heald, S. M.; Seidler, G. T.; et al. Fast Detection Allowing Analysis of Metalloprotein Electronic Structure by X-Ray Emission Spectroscopy at Room Temperature. *J. Phys. Chem. Lett.* **2012**, *3* (14), 1858–1864.

Debus, R. J.; Strickler, M. A.; Walker, L. M.; Hillier, W. No Evidence from FTIR Difference Spectroscopy That Aspartate-170 of the D1 Polypeptide Ligates a Manganese Ion That Undergoes Oxidation during the S0 to S1, S1 to S2, or S2 to S3 Transitions in Photosystem II. *Biochemistry (Mosc.)* **2005**, *44* (5), 1367–1374.

Debye, P. Interferenz von Röntgenstrahlen und Wärmebewegung. *Ann. Phys.* **1913**, *348* (1), 49–92.

Dekker, J. P.; Van Gorkom, H. J.; Brok, M.; Ouwehand, L. Optical Characterization of Photosystem II Electron Donors. *Biochim. Biophys. Acta BBA - Bioenerg.* **1984**, *764* (3), 301–309.

DePonte, D. P.; Weierstall, U.; Schmidt, K.; Warner, J.; Starodub, D.; Spence, J. C. H.; Doak, R. B. Gas Dynamic Virtual Nozzle for Generation of Microscopic Droplet Streams. *J. Phys. Appl. Phys.* **2008**, *41* (19), 195505.

Dilbeck, P. L.; Hwang, H. J.; Ztharieva, I.; Gerencser, L.; Dau, H.; Burnap, R. L. The D1-D61N Mutation in Synechocystis Sp. PCC 6803 Allows the Observation of pH-Sensitive Intermediates in the Formation and Release of O_2 from Photosystem II. *Biochemistry (Mosc.)* **2012**, *51* (6), 1079–1091.

Duisenberg, A. J. M. Indexing in Single-Crystal Diffractometry with an Obstinate List of Reflections. *J. Appl. Crystallogr.* **1992**, *25* (2), 92–96.

Emma, P.; Akre, R.; Arthur, J.; Bionta, R.; Bostedt, C.; Bozek, J.; Brachmann, A.; Bucksbaum, P.; Coffee, R.; Decker, F.-J.; et al. First Lasing and Operation of an Ångstrom-Wavelength Free-Electron Laser. *Nat. Photonics* **2010**, *4* (9), 641–647.

Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (4), 486–501.

Erk, B.; Boll, R.; Trippel, S.; Anielski, D.; Foucar, L.; Rudek, B.; Epp, S. W.; Coffee, R.; Carron, S.; Schorb, S.; et al. Imaging Charge Transfer in Iodomethane upon X-Ray Photoabsorption. *Science* **2014**, *345* (6194), 288–291.

Ewald, P. . Introduction to the Dynamical Theory of X-Ray Diffraction. *Acta Crystallogr. A.* **1969**, *A* (25), 103–108.

Fischer, N. The PsaC Subunit of Photosystem I Provides an Essential Lysine Residue for Fast Electron Transfer to Ferredoxin. *EMBO J.* **1998**, *17* (4), 849–858.

Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. *Proc. Natl. Acad. Sci.* **2011**, *108* (39), 16247–16252.

Fromme, P.; Witt, H. T. Improved Isolation and Crystallization of Photosystem I for Structural Analysis. *Biochim. Biophys. Acta BBA - Bioenerg.* **1998**, *1365* (1-2), 175–184.

Fromme, P.; Grotjohann, I. Overview of Photosynthesis. In *Photosynthetic Protein Complexes*; Fromme, P., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2008; pp 1–22.

Fromme, P.; Spence, J. C. Femtosecond Nanocrystallography Using X-Ray Lasers for Membrane Protein Structure Determination. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 509–516.

Fromme, P.; Melkozernov, A.; Jordan, P.; Krauss, N. Structure and Function of Photosystem I: Interaction with Its Soluble Electron Carriers and External Antenna Systems. *FEBS Lett.* **2003**, *555* (1), 40–44.

Fromme, R.; Ishchenko, A.; Metz, M.; Chowdhury, S. R.; Basu, S.; Boutet, S.; Fromme, P.; White, T. A.; Barty, A.; Spence, J. C. H.; et al. Serial Femtosecond Crystallography of Soluble Proteins in Lipidic Cubic Phase. *IUCrJ* **2015**, *2* (5), 545–551.

Galstyan, A.; Robertazzi, A.; Knapp, E. W. Oxygen-Evolving Mn Cluster in Photosystem II: The Protonation Pattern and Oxidation State in the High-Resolution Crystal Structure. *J. Am. Chem. Soc.* **2012**, *134* (17), 7442–7449.

Garman, E. F. Radiation Damage in Macromolecular Crystallography: What Is It and Why Should We Care? *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 4), 339–351.

Genick, U. K.; Borgstahl, G. E.; Ng, K.; Ren, Z.; Pradervand, C.; Burke, P. M.; Srajer, V.; Teng, T. Y.; Schildkamp, W.; McRee, D. E.; et al. Structure of a Protein Photocycle Intermediate by Millisecond Time-Resolved Crystallography. *Science* **1997**, *275* (5305), 1471–1475.

Genick, U. K.; Soltis, S. M.; Kuhn, P.; Canestrelli, I. L.; Getzoff, E. D. Structure at 0.85[thinsp]A Resolution of an Early Protein Photocycle Intermediate. *Nature* **1998**, *392* (6672), 206–209.

Ginn, H. M.; Messerschmidt, M.; Ji, X.; Zhang, H.; Axford, D.; Gildea, R. J.; Winter, G.; Brewster, A. S.; Hattne, J.; Wagner, A.; et al. Structure of CPV17 Polyhedrin Determined by the Improved Analysis of Serial Femtosecond Crystallographic Data. *Nat. Commun.* **2015**, *6*, 6435.

Glaeser, R. M.; Downing, K. H. High-Resolution Electron Crystallography of Protein Molecules. *Ultramicroscopy* **1993**, *52* (3-4), 478–486.

Grundmeier, A.; Dau, H. Structural Models of the Manganese Complex of Photosystem II and Mechanistic Implications. *Biochim. Biophys. Acta* **2012**, *1817* (1), 88–105.

Guinier, A. X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies; Dover Publications, New York, 1994.

Han, G.; Mamedov, F.; Styring, S. Misses during Water Oxidation in Photosystem II Are S State-Dependent. *J. Biol. Chem.* **2012**, *287* (16), 13422–13429.

Hart, P.; Boutet, S.; Carini, G.; Dragone, A.; Duda, B.; Freytag, D.; Haller, G.; Herbst, R.; Herrmann, S.; Kenney, C.; et al. The Cornell-SLAC Pixel Array Detector at LCLS. *SPIE* **2012**.

Henderson, R. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Q. Rev. Biophys.* **1995**, *28* (2), 171–193.

Henderson, R.; Moffat, J. K. The Difference Fourier Technique in Protein Crystallography: Errors and Their Treatment. *Acta Crystallogr. B* **1971**, *27* (7), 1414–1420.

Holton, J. M.; Frankel, K. A. The Minimum Crystal Size Needed for a Complete Diffraction Data Set. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (4), 393–408.

Holzwarth, A. R.; Muller, M. G.; Reus, M.; Nowaczyk, M.; Sander, J.; Rogner, M. Kinetics and Mechanism of Electron Transfer in Intact Photosystem II and in the Isolated Reaction Center: Pheophytin Is the Primary Electron Acceptor. *Proc. Natl. Acad. Sci.* **2006**, *103* (18), 6895–6900.

Hope, H. Cryocrystallography of Biological Macromolecules: A Generally Applicable Method. *Acta Crystallogr. B* **1988**, *44* (1), 22–26.

Hope, H.; Frolow, F.; von Böhlen, K.; Makowski, I.; Kratky, C.; Halfon, Y.; Danz, H.; Webster, P.; Bartels, K. S.; Wittmann, H. G.; et al. Cryocrystallography of Ribosomal Particles. *Acta Crystallogr. B* **1989**, *45* (2), 190–199.

Huang, Z.; Kim, K.-J. Review of X-Ray Free-Electron Laser Theory. *Phys. Rev. Spec. Top. - Accel. Beams* **2007**, *10* (3).

Hundelt, M.; Hays, A. M.; Debus, R. J.; Junge, W. Oxygenic Photosystem II: The Mutation D1-D61N in Synechocystis Sp. PCC 6803 Retards S-State Transitions without Affecting Electron Transfer from YZ to P680+. *Biochemistry (Mosc.)* **1998**, *37* (41), 14450–14456.

Ihee, H.; Rajagopal, S.; Srajer, V.; Pahl, R.; Anderson, S.; Schmidt, M.; Schotte, F.; Anfinrud, P. A.; Wulff, M.; Moffat, K. From The Cover: Visualizing Reaction Pathways in Photoactive Yellow Protein from Nanoseconds to Seconds. *Proc. Natl. Acad. Sci.* **2005**, *102* (20), 7145–7150.

Isobe, H.; Shoji, M.; Yamanaka, S.; Umena, Y.; Kawakami, K.; Kamiya, N.; Shen, J.-R.; Yamaguchi, K. Theoretical Illumination of Water-Inserted Structures of the CaMn4O5 Cluster in the S2 and S3 States of Oxygen-Evolving Complex of Photosystem II: Full Geometry Optimizations by B3LYP Hybrid Density Functional. *Dalton Trans.* **2012**, *41* (44), 13727.

Jackson, J. D. *Classical Electrodynamics*, 3rd ed.; Wiley: New York, 1999.

Jordan, P.; Fromme, P.; Witt, H.; Klukas, O.; Saenger, W.; Krauss, N. Three-Dimensional Structure of Cyanobacterial Photosystem I at 2.5 Angstrom Resolution. *Nature* **2001**, *411* (6840), 909–917.

Jung, Y. O.; Lee, J. H.; Kim, J.; Schmidt, M.; Moffat, K.; Šrajer, V.; Ihee, H. Volume-Conserving Trans–cis Isomerization Pathways in Photoactive Yellow Protein Visualized by Picosecond X-Ray Crystallography. *Nat. Chem.* **2013**, *5* (3), 212–220.

Kanady, J. S.; Tsui, E. Y.; Day, M. W.; Agapie, T. A Synthetic Model of the Mn_3Ca Subsite of the Oxygen-Evolving Complex in Photosystem II. *Science* **2011**, *333* (6043), 733–736.

Karplus, P. A.; Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **2012**, *336* (6084), 1030–1033.

Kern, J.; Alonso-Mori, R.; Tran, R.; Hattne, J.; Gildea, R. J.; Echols, N.; Glöckner, C.; Hellmich, J.; Laksmono, H.; Sierra, R. G.; et al. Simultaneous Femtosecond X-Ray Spectroscopy and Diffraction of Photosystem II at Room Temperature. *Science* **2013**.

Kirian, R. A.; Wang, X.; Weierstall, U.; Schmidt, K. E.; Spence, J. C. H.; Hunter, M.; Fromme, P.; White, T.; Chapman, H. N.; Holton, J. Femtosecond Protein Nanocrystallography—data Analysis Methods. *Opt. Express* **2010**, *18* (6), 5713.

Kirian, R. A.; White, T. A.; Holton, J. M.; Chapman, H. N.; Fromme, P.; Barty, A.; Lomb, L.; Aquila, A.; Maia, F. R. N. C.; et. al. Structure-Factor Analysis of Femtosecond Microdiffraction Patterns from Protein Nanocrystals. *Acta Crystallogr. A* **2011**, *67* (2), 131–140.

Kupitz C.; Basu, S.; Grotjohann, I.; Fromme, R.; Zatsepin, N. A.; Rendek, K. N.; Hunter, M. S.; Shoeman, R. L.; White, T. A.; Wang, D.; et al. Serial Time-Resolved Crystallography of Photosystem II Using a Femtosecond X-Ray Laser. *Nature* **2014**, *513* (7517), 261–265.

Kupitz, C.; Grotjohann, I.; Conrad, C. E.; Roy-Chowdhury, S.; Fromme, R.; Fromme, P. Microcrystallization Techniques for Serial Femtosecond Crystallography Using Photosystem II from Thermosynechococcus Elongatus as a Model System. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1647), 20130316–20130316.

Lincoln, C. N.; Fitzpatrick, A. E.; Thor, J. J. van. Photoisomerisation Quantum Yield and Non-Linear Cross-Sections with Femtosecond Excitation of the Photoactive Yellow Protein. *Phys. Chem. Chem. Phys.* **2012**, *14* (45), 15752.

Liu, W.; Wacker, D.; Gati, C.; Han, G. W.; James, D.; Wang, D.; Nelson, G.; Weierstall, U.; Katritch, V.; Barty, A.; et al. Serial Femtosecond Crystallography of G Protein-Coupled Receptors. *Science* **2013**, *342* (6165), 1521–1524.

Loll, B.; Kern, J.; Saenger, W.; Zouni, A.; Biesiadka, J. Towards Complete Cofactor Arrangement in the 3.0 Å Resolution Structure of Photosystem II. *Nature* **2005**, *438* (7070), 1040–1044.

Lomb, L.; Barends, T.; Kassemeyer, S.; Aquila, A.; Epp, S.; Erk, B.; Foucar, L.; Hartmann, R.; Rudek, B.; Rolles, D.; et al. Radiation Damage in Protein Serial Femtosecond Crystallography Using an X-Ray Free-Electron Laser. *Phys. Rev. B* **2011**, *84* (21).

Lomb, L.; Steinbrener, J.; Bari, S.; Beisel, D.; Berndt, D.; Kieser, C.; Lukat, M.; Neef, N.; Shoeman, R. L. An Anti-Settling Sample Delivery Instrument for Serial Femtosecond Crystallography. *J. Appl. Crystallogr.* **2012**, *45* (4), 674–678.

Luber, S.; Rivalta, I.; Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N.; Brudvig, G. W.; Batista, V. S. S1-State Model of the O2-Evolving Complex of Photosystem II. *Biochemistry (Mosc.)* **2011**, *50* (29), 6308–6311.

Margaritondo, G.; Rebernik Ribic, P. A Simplified Description of X-Ray Free-Electron Lasers. *J. Synchrotron Radiat.* **2011**, *18* (2), 101–108.

McCoy, A. J. Solving Structures of Protein Complexes by Molecular Replacement with *Phaser*. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63* (1), 32–41.

Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E. Orbital Steering in the Catalytic Power of Enzymes: Small Structural Changes with Large Catalytic Consequences. *Science* **1997**, *277* (5323), 202–206.

Moffat, K. Time-Resolved Crystallography and Protein Design: Signalling Photoreceptors and Optogenetics. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1647), 20130568–20130568.

Moffat, K.; Ren, Z. Synchrotron Radiation Applications to Macromolecular Crystallography. *Curr. Opin. Struct. Biol.* **1997**, *7* (5), 689–696.

Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53* (3), 240–255.

Navarro-García, F.; Sears, C.; Eslava, C.; Cravioto, A.; Nataro, J. P. Cytoskeletal Effects Induced by Pet, the Serine Protease Enterotoxin of Enteroaggregative Escherichia Coli. *Infect. Immun.* **1999**, *67* (5), 2184–2192.

Nave, C. Radiation Damage in Protein Crystallography. *Radiat. Phys. Chem.* **1995**, *45* (3), 483–490.

Neutze, R.; Moffat, K. Time-Resolved Structural Studies at Synchrotrons and X-Ray Free Electron Lasers: Opportunities and Challenges. *Curr. Opin. Struct. Biol.* **2012**, *22* (5), 651–659.

Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. Potential for Biomolecular Imaging with Femtosecond X-Ray Pulses. *Nature* **2000**, *406* (6797), 752–757.

Pal, R.; Negre, C. F. A.; Vogt, L.; Pokhrel, R.; Ertem, M. Z.; Brudvig, G. W.; Batista, V. S. S0-State Model of the Oxygen-Evolving Complex of Photosystem II. *Biochemistry (Mosc.)* **2013**, *52* (44), 7703–7706.

Perez Navarro, M.; Ames, W. M.; Nilsson, H.; Lohmiller, T.; Pantazis, D. A.; Rapatskiy, L.; Nowaczyk, M. M.; Neese, F.; Boussac, A.; Messinger, J.; et al. Ammonia Binding to the Oxygen-Evolving Complex of Photosystem II Identifies the Solvent-Exchangeable Oxygen Bridge (-Oxo) of the Manganese Tetramer. *Proc. Natl. Acad. Sci.* **2013**, *110* (39), 15561–15566.

Perman, B. Energy Transduction on the Nanosecond Time Scale: Early Structural Events in a Xanthopsin Photocycle. *Science* **1998**, *279* (5358), 1946–1950.

Powell, H. R. The Rossmann Fourier Autoindexing Algorithm in MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55* (Pt 10), 1690–1695.

Pushkar, Y.; Yano, J.; Sauer, K.; Boussac, A.; Yachandra, V. K. Structural Changes in the Mn4Ca Cluster and the Mechanism of Photosynthetic Water Splitting. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (6), 1879–1884.

Raszewski, G.; Renger, T. Light Harvesting in Photosystem II Core Complexes Is Limited by the Transfer to the Trap: Can the Core Complex Turn into a Photoprotective Mode? *J. Am. Chem. Soc.* **2008**, *130* (13), 4431–4446.

Ravelli, R. B.; McSweeney, S. M. The "Fingerprint" That X-Rays Can Leave on Structures. *Struct. Lond. Engl. 1993* **2000**, *8* (3), 315–328.

Read, R. J. New Ways of Looking at Experimental Phasing. *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59* (11), 1891–1902.

Redecke, L.; Nass, K.; DePonte, D. P.; White, T. A.; Rehders, D.; Barty, A.; Stellato, F.; Liang, M.; Barends, T. R. M.; Boutet, S.; et al. Natively Inhibited Trypanosoma Brucei Cathepsin B Structure Determined by Using an X-Ray Laser. *Science* **2013**, *339* (6116), 227–230.

Renger, G. Mechanism of Light Induced Water Splitting in Photosystem II of Oxygen Evolving Photosynthetic Organisms. *Biochim. Biophys. Acta* **2012**, *1817* (8), 1164–1176.

Renger, T.; Schlodder, E. Optical Properties, Excitation Energy and Primary Charge Transfer in Photosystem II: Theory Meets Experiment. *J. Photochem. Photobiol. B-Biol.* **2011**, *104* (1-2), 126–141.

Ren, Z.; Perman, B.; Šrajer, V.; Teng, T.-Y.; Pradervand, C.; Bourgeois, D.; Schotte, F.; Ursby, T.; Kort, R.; Wulff, M.; et al. A Molecular Movie at 1.8 Å Resolution Displays the Photocycle of Photoactive Yellow Protein, a Eubacterial Blue-Light Receptor, from Nanoseconds to Seconds [†]. *Biochemistry (Mosc.)* **2001**, *40* (46), 13788–13801.

Rhodes, G. *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*, 3rd ed.; Complementary science series; Elsevier/Academic Press: Amsterdam ; Boston, 2006.

Rivalta, I.; Brudvig, G. W.; Batista, V. S. Oxomanganese Complexes for Natural and Artificial Photosynthesis. *Curr. Opin. Chem. Biol.* **2012**, *16* (1-2), 11–18.

Robinson, I.; Gruebel, G.; Mochrie, S. Focus on X-Ray Beams with High Coherence. *New J. Phys.* **2010**, *12* (3), 035002.

Rossmann, M. G.; Blow, D. M. The Detection of Sub-Units within the Crystallographic Asymmetric Unit. *Acta Crystallogr.* **1962**, *15* (1), 24–31.

Rossmann, M. G.; van Beek, C. G. Data Processing. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55* (10), 1631–1640.

Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*; Garland Science: New York, 2010.

Sauter, N. K.; Hattne, J.; Grosse-Kunstleve, R. W.; Echols, N. New Python-Based Methods for Data Processing. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (7), 1274–1282.

Schlichting, I.; Berendzen, J.; Phillips, G. N.; Sweet, R. M. Crystal Structure of Photolysed Carbonmonoxy-Myoglobin. *Nature* **1994**, *371* (6500), 808–812.

Schmidt, M.; Rajagopal, S.; Ren, Z.; Moffat, K. Application of Singular Value Decomposition to the Analysis of Time-Resolved Macromolecular X-Ray Data. *Biophys. J.* **2003**, *84* (3), 2112–2129.

Schmidt, M.; Srajer, V.; Henning, R.; Ihee, H.; Purwar, N.; Tenboer, J.; Tripathi, S. Protein Energy Landscapes Determined by Five-Dimensional Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (12), 2534–2542.

Schotte, F. Watching a Protein as It Functions with 150-Ps Time-Resolved X-Ray Crystallography. *Science* **2003**, *300* (5627), 1944–1947.

Schotte, F.; Soman, J.; Olson, J. S.; Wulff, M.; Anfinrud, P. A. Picosecond Time-Resolved X-Ray Crystallography: Probing Protein Function in Real Time. *J. Struct. Biol.* **2004**, *147* (3), 235–246.

Schotte, F.; Cho, H. S.; Kaila, V. R. I.; Kamikubo, H.; Dashdorj, N.; Henry, E. R.; Graber, T. J.; Henning, R.; Wulff, M.; Hummer, G.; et al. Watching a Signaling Protein Function in Real Time via 100-Ps Time-Resolved Laue Crystallography. *Proc. Natl. Acad. Sci.* **2012**, *109* (47), 19256–19261.

Schröder, G. F.; Brunger, A. T.; Levitt, M. Combining Efficient Conformational Sampling with a Deformable Elastic Network Model Facilitates Structure Refinement at Low Resolution. *Structure* **2007**, *15* (12), 1630–1641.

Sétif, P. Q.; Bottin, H. Laser Flash Absorption Spectroscopy Study of Ferredoxin Reduction by Photosystem I: Spectral and Kinetic Evidence for the Existence of Several Photosystem I-Ferredoxin Complexes. *Biochemistry (Mosc.)* **1995**, *34* (28), 9059–9070.

Shi, D.; Nannenga, B. L.; Iadanza, M. G.; Gonen, T. Three-Dimensional Electron Crystallography of Protein Microcrystals. *eLife* **2013**, *2*.

Shinkarev, V. P.; Govindjee, null. Insight into the Relationship of Chlorophyll a Fluorescence Yield to the Concentration of Its Natural Quenchers in Oxygenic Photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (16), 7466–7469.

Siegbahn, P. E. M. Mechanism and Energy Diagram for O-O Bond Formation in the Oxygen-Evolving Complex in Photosystem II. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2008**, *363* (1494), 1221–1228; discussion 1228.

Siegbahn, P. E. M. Mechanisms for Proton Release during Water Oxidation in the S2 to S3 and S3 to S4 Transitions in Photosystem II. *Phys. Chem. Chem. Phys. PCCP* **2012**, *14* (14), 4849–4856.

Siegbahn, P. E. M. Substrate Water Exchange for the Oxygen Evolving Complex in PSII in the S1, S2, and S3 States. *J. Am. Chem. Soc.* **2013**, *135* (25), 9442–9449.

Sierra, R. G.; Laksmono, H.; Kern, J.; Tran, R.; Hattne, J.; Alonso-Mori, R.; Lassalle-Kaiser, B.; Glöckner, C.; Hellmich, J.; Schafer, D. W.; et al. Nanoflow Electrospinning Serial Femtosecond Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (11), 1584–1587.

Singh, A.; Singh, H. Time-Scale and Nature of Radiation-Biological Damage: Approaches to Radiation Protection and Post-Irradiation Therapy. *Prog. Biophys. Mol. Biol.* **1982**, *39* (2), 69–107.

Solem, J. C. Imaging Biological Specimens with High-Intensity Soft X Rays. *J. Opt. Soc. Am. B* **1986**, *3* (11), 1551–1565.

Spence, J. C. H. *High-Resolution Electron Microscopy*, Fourth edition.; Oxford University Press: Oxford, 2013.

Spence, J. C. H.; Chapman, H. N. The Birth of a New Field. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1647), 20130309–20130309.

Spence, J. C. H.; Weierstall, U.; Chapman, H. N. X-Ray Lasers for Structural and Dynamic Biology. *Rep. Prog. Phys.* **2012**, *75* (10), 102601.

Srajer, V.; Teng, T.; Ursby, T.; Pradervand, C.; Ren, Z.; Adachi, S.; Schildkamp, W.; Bourgeois, D.; Wulff, M.; Moffat, K. Photolysis of the Carbon Monoxide Complex of Myoglobin: Nanosecond Time-Resolved Crystallography. *Science* **1996**, *274* (5293), 1726–1729.

Steller, I.; Bolotovsky, R.; Rossmann, M. G. An Algorithm for Automatic Indexing of Oscillation Images Using Fourier Analysis. *J. Appl. Crystallogr.* **1997**, *30* (6), 1036–1040.

Subramanian, G.; Basu, S.; Liu, H.; Zuo, J.-M.; Spence, J. C. H. Solving Protein Nanocrystals by Cryo-EM Diffraction: Multiple Scattering Artifacts. *Ultramicroscopy* **2015**, *148*, 87–93.

Tenboer, J.; Basu, S.; Zatsepin, N.; Pande, K.; Milathianaki, D.; Frank, M.; Hunter, M.; Boutet, S.; Williams, G. J.; Koglin, J. E.; et al. Time-Resolved Serial Crystallography Captures High-Resolution Intermediates of Photoactive Yellow Protein. *Science* **2014**, *346* (6214), 1242–1246.

Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Adams, P. D.; Read, R. J.; Zwart, P. H.; Hung, L.-W. Iterative-Build OMIT Maps: Map Improvement by Iterative Model Building and Refinement without Model Bias. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64* (5), 515–524.

Uervirojnangkoorn, M.; Zeldin, O. B.; Lyubimov, A. Y.; Hattne, J.; Brewster, A. S.; Sauter, N. K.; Brunger, A. T.; Weis, W. I. Enabling X-Ray Free Electron Laser Crystallography for Challenging Biological Systems from a Limited Number of Crystals. *eLife* **2015**, *4*.

Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N. Crystal Structure of Oxygen-Evolving Photosystem II at a Resolution of 1.9 Å. *Nature* **2011**, *473* (7345), 55–60.

Vellieux, F. M. D.; Dijkstra, B. W. Computation of Bhat's OMIT Maps with Different Coefficients. *J. Appl. Crystallogr.* **1997**, *30* (3), 396–399.

Warren, B. E. *X-Ray Diffraction*, Dover ed.; Dover Publications: New York, 1990.

Weierstall, U.; Doak, R. B.; Spence, J. C. H.; Starodub, D.; Shapiro, D.; Kennedy, P.; Warner, J.; Hembree, G. G.; Fromme, P.; Chapman, H. N. Droplet Streams for Serial Crystallography of Proteins. *Exp. Fluids* **2008**, *44* (5), 675–689.

Weierstall, U.; Spence, J. C. H.; Doak, R. B. Injector for Scattering Measurements on Fully Solvated Biospecies. *Rev. Sci. Instrum.* **2012**, *83* (3), 035108.

Weierstall, U.; James, D.; Wang, C.; White, T. A.; Wang, D.; Liu, W.; Spence, J. C. H.; Bruce Doak, R.; Nelson, G.; Fromme, P.; et al. Lipidic Cubic Phase Injector Facilitates Membrane Protein Serial Femtosecond Crystallography. *Nat. Commun.* **2014**, *5*.

White, T. A. Post-Refinement Method for Snapshot Serial Crystallography. *Philos. Trans. R. Soc. B Biol. Sci.* **2014**, *369* (1647), 20130330–20130330.

White, T. A.; Kirian, R. A.; Martin, A. V.; Aquila, A.; Nass, K.; Barty, A.; Chapman, H. N. CrystFEL : A Software Suite for Snapshot Serial Crystallography. *J. Appl. Crystallogr.* **2012**, *45* (2), 335–341.

De Wijn, R.; van Gorkom, H. J. Kinetics of Electron Transfer from Q $_A$ to Q $_B$ in Photosystem II [†]. *Biochemistry (Mosc.)* **2001**, *40* (39), 11912–11922.

Wilson, A. J. C. Determination of Absolute from Relative X-Ray Intensity Data. *Nature* **1942**, *150* (3796), 152–152.

Witt, I.; Witt, H. T.; Di Fiore, D.; Rögner, M.; Hinrichs, W.; Saenger, W.; Granzin, J.; Betzel, C.; Dauter, Z. X-Ray Characterization of Single Crystals of the Reaction Center I of Water Splitting Photosynthesis. *Berichte Bunsenges. Für Phys. Chem.* **1988**, *92* (12), 1503–1506.

Yamaguchi, K.; Yamanaka, S.; Isobe, H.; Saito, T.; Kanda, K.; Umena, Y.; Kawakami, K.; Shen, J.-R.; Kamiya, N.; Okumura, M.; et al. The Nature of Chemical Bonds of the CaMn $_4$ O $_5$ Cluster in Oxygen Evolving Complex of Photosystem II: Jahn-Teller Distortion and Its Suppression by Ca Doping in Cubane Structures. *Int. J. Quantum Chem.* **2013**, *113* (4), 453–473.

Yang, X.; Ren, Z.; Moffat, K. Structure Refinement Against Synchrotron Laue Data: Strategies for Data Collection and Reduction. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54* (3), 367–377.

Yano, J.; Yachandra, V. K. Oxidation State Changes of the Mn4Ca Cluster in Photosystem II. *Photosynth. Res.* **2007**, *92* (3), 289–303.

Yano, J.; Yachandra, V. K. Where Water Is Oxidized to Dioxygen: Structure of the Photosynthetic Mn4Ca Cluster from X-Ray Spectroscopy. *Inorg. Chem.* **2008**, *47* (6), 1711–1726.

Yano, J.; Pushkar, Y.; Glatzel, P.; Lewis, A.; Sauer, K.; Messinger, J.; Bergmann, U.; Yachandra, V. High-Resolution Mn EXAFS of the Oxygen-Evolving Complex in Photosystem II: Structural Implications for the Mn $_4$ Ca Cluster. *J. Am. Chem. Soc.* **2005**, *127* (43), 14974–14975.

Yonath, A.; Leonard, K. R.; Weinstein, S.; Wittmann, H. G. Approaches to the Determination of the Three-Dimensional Architecture of Ribosomal Particles. *Cold Spring Harb. Symp. Quant. Biol.* **1987**, *52* (0), 729–741.

Zouni, A.; Witt, H.; Kern, J.; Fromme, P.; Krauss, N.; Saenger, W.; Orth, P. Crystal Structure of Photosystem II from Synechococcus Elongatus at 3.8 Angstrom Resolution. *NATURE* **2001**, *409* (6821), 739–743.

International Tables for Crystallography: Space-Group Symmetry, 1st ed.; Hahn, T., Ed.; Fuess, H., Hahn, T., Wondratschek, H., Müller, U., Shmueli, U., Prince, E., Authier, A., Kopský, V., Litvin, D. B., Rossmann, M. G., et al., Series Eds.; International Tables for Crystallography; International Union of Crystallography: Chester, England, 2006; Vol. A.

Photosynthetic Protein Complexes: A Structural Approach; Fromme, P., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, 2008.

BIOGRAPHICAL SKETCH

Shibom Basu was born at the suburban area of the greater Kolkata, in India on May 18, 1988. He completed his high school education from Bhatpara Amar Krishna Pathsala (H.S.) majoring in Science in 2006. He was admitted into renowned Presidency College, Kolkata to pursue his bachelor degree (i.e., B.Sc.) with Chemistry Honors, completing his degree program in 3 years (2006-2009). After finishing his B.Sc. degree in Chemistry, he competed in the all India entrance examination and was admitted into the Indian Institute of Technology (IIT) – Madras in 2009. There he pursued his master degree (M.Sc.) in Chemistry, completing his degree program in two years. While pursuing his M.Sc. degree in IIT, he carried out a summer project in 2010 for two months in the field of Computational Biophysics at Calcutta University. He worked on a project, titled "Homology Modeling and docking of ABF family transcription factors." During his one-year master thesis in IIT, he worked on "*ab initio* quantum dynamical calculation on proton and nitrogen diatomic molecular interaction". In this project, he solved time-dependent Schrodinger equation for multi-body system using Chebyshev Polynomial. After completion of his M.Sc. degree with standing 5th in the entire department, he came to the USA to pursue his PhD degree at Arizona State University in 2011. There he attended Department of Chemistry and Biochemistry. He performed his PhD research under the supervision of Prof. Petra Fromme on a very cutting-edge project, titled "Time-resolved X-ray crystallography using X-ray free-electron lasers". During his PhD, he collaborated with hundreds of scientists across the world, participated in various experiments, and published 12 highly cited research articles in high-ranking journals. His area of interests includes X-ray crystallography, X-ray optics, beamline instrumentation, serial crystallography and time-resolved studies of biological processes.