

PUMP-PROBE AND MIX-AND-INJECT EXPERIMENTS AT X-RAY FREE

ELECTRON LASERS

by

Suraj Pandey

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Physics

at

The University of Wisconsin-Milwaukee

December 2020

# ABSTRACT

PUMP-PROBE AND MIX-AND-INJECT EXPERIMENTS AT X-RAY FREE ELECTRON LASERS

by

Suraj Pandey

The University of Wisconsin-Milwaukee, 2020  
Under the Supervision of Professor Marius Schmidt

Time resolved serial femtosecond crystallography (TR-SFX) utilizes X-ray crystallography to visualize the reaction of molecules in real time at the atomic level. Crystals of biological macromolecules are exposed to powerful X-ray pulses. The X-ray radiation emitted by the crystal is then measured by an X-ray sensitive area detector that produces an image called a diffraction pattern. These patterns are analyzed to determine a three-dimensional atomic structure of the biological macromolecule.

The ultimate goal of TR-SFX is to make a “molecular movie” that shows the reaction dynamics of a biological process. For this, a reaction is started in a macromolecular crystal and its three-dimensional atomic structures at different time intervals are determined. When these structures are played in a timely order, a molecular movie is recorded. A perfect analogy of this in a real life would be taking pictures of someone dancing, and playing the succession of pictures to observe the dance.

TR-SFX is a method that requires an X-ray Free Electron Laser (XFEL). The XFEL produces X-ray pulses of tens of femtosecond duration with  $10^{12}$  photons per pulse. These pulses are so strong that the crystals are destroyed after being exposed to a single pulse. Since the X-ray pulses

are so short, diffraction patterns are collected by the detector just before the crystals are destroyed which is called “diffraction before destruction”. Most impressively, at XFELs, each diffraction pattern is obtained from a fresh crystal. As a result, both reversible and non-reversible reactions are placed in an equal footing and can be studied similarly.

Pump-probe TR-SFX and mix and inject crystallography (MISC) are two cornerstones of TR-SFX. In pump-probe TR-SFX, photoactive macromolecules within the crystals are activated using an optical laser called the pump, and the reaction is probed by XFEL pulses. Whereas in MISC, biomolecular crystals are mixed with a substrate, and the structural changes are probed by XFEL pulses in a time resolved fashion.

This dissertation presents the time-resolved studies of an enzyme called  $\beta$ -lactamase (BlaC) and two photoactive proteins - photoactive yellow protein (PYP) and phytochromes. MISC was used for the study of the enzymatic reaction of BlaC with an antibiotic called Ceftriaxone (CEF). This proof-of-principle experiment conducted at Linac Coherent Light Source (LCLS) operated by Stanford University in Menlo Park, California, showed how CEF binds at the active site of BlaC. Similarly, a pump-probe experiment on PYP was accomplished at the first megahertz XFEL, European XFEL (EuXFEL), in Hamburg, Germany. This experiment was performed to explore the picosecond regime of the photocycle of PYP and test the feasibility of TR-SFX at high repetition rates XFEL. Finally, another pump-probe experiment on phytochromes was conducted at Spring-8 Angstrom Compact free electron Laser (SACLA), in Harima, Japan. With this experiment, we have observed the Z-E isomerization of phytochromes for the first time and determined the previously unknown time-resolved structure at 1ps.

This dissertation also presents data acquisition techniques and processing methods used at XFELs. It explains methods to analyze millions of patterns obtained during an experiment, with the goal of solving the X-ray structures of different intermediates of the reaction. In summary, my dissertation explains everything from the beginning of the experiment to the production of a molecular movie.



*To my late father, Shambhu Prasad Pandey,  
and my loving mother, Rukmani Pandey.*

## TABLE OF CONTENTS

ABSTRACT

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

<b>1. Introduction .....</b>	<b>1</b>
1.1 X-rays and Crystallography .....	2
1.1.1 A Brief History of X-rays and Crystallography .....	2
1.1.2 Crystals .....	3
1.1.3 Protein Crystallization .....	4
1.2 Advanced X-ray sources .....	4
1.2.1 Third Generation Synchrotron Sources .....	6
1.2.2 Radiation Damage in MX .....	6
1.2.3 Diffraction before Destruction at XFELs .....	7
1.2.4 Comparison of various XFELs .....	9
1.3 Proteins investigated in this dissertation. ....	11
1.3.1 $\beta$ -Lactamase .....	11
1.3.2 Photoactive Yellow Protein .....	13
1.3.3 Phytochrome .....	15
<b>2. Materials and Methods.....</b>	<b>18</b>
2.1 Protein Purification and Crystallization .....	18

2.1.1 PYP .....	20
2.1.2 BlaC.....	21
2.1.3 Phytochrome .....	22
2.2 X-ray Crystallography .....	22
2.2.1 Crystal lattice and Reciprocal Lattice.....	22
2.2.2 Scattering of X-rays from an Atom.....	26
2.2.3 Scattering of X-rays from a Molecule .....	27
2.2.4 Scattering from a Crystal.....	28
2.2.5 Elastic Diffraction Condition and Ewald Sphere Construction .....	29
2.2.6 The Phase Problem and phasing algorithms .....	31
2.2.7 Mosaicity.....	33
2.2.8 Debye-Waller Factor and Wilson Plot .....	34
2.2.8.1. Debye Waller Factor.....	34
2.2.8.2 Absolute Scale .....	34
2.3 Serial Crystallography (SX).....	35
2.3.1 Serial Femtosecond Crystallography .....	36
2.3.2 Time-Resolved Serial Femtosecond Crystallography.....	38
2.3.3 Pump-probe TR-SFX.....	39
2.3.4 Mix and Inject Serial Crystallography (MISC) .....	40
2.4 Data Analysis at XFELs .....	45
2.4.1 SFX – From a data analysis point of view .....	45
2.4.2 Preprocessing of raw data using Cheetah - Hit Finding .....	46

2.4.3	Online monitoring during experiments .....	49
2.4.4	Indexing of Diffraction Patterns .....	50
2.4.5	Indexing ambiguity.....	51
2.4.6	Merging and Assessment of Data Quality .....	52
2.4.7	Analysis of Processed Data .....	53
2.4.7.1	Difference Electron Density (DED) Maps.....	53
2.4.7.2	Omit Difference Maps .....	55
2.4.7.3	Extrapolated Electron Density (EED) Maps .....	56
2.5	List of Attended Beamtimes at XFELs and my contributions.....	57
2.6	Mix-and-Inject Serial Crystallography (MISC) at the LCLS .....	58
2.7	Pump-probe TR-SFX at the EuXFEL.....	61
2.8	Pump-probe TR-SFX at the SACLA .....	64
<b>3.</b>	<b>Results.....</b>	<b>65</b>
3.1	MISC at the LCLS.....	65
3.2	Enzyme intermediates captured using MISC at the LCLS .....	68
3.3	Pump-probe TR-SFX at the EuXFEL.....	74
3.3.1	Control Experiment at 1.13 MHz .....	74
3.3.2	Picosecond time delay data collection at 564 kHz .....	78
3.4	Primary structural photo-response of Phytochrome captured using pump-probe TR-SFX at SACLA.....	84
<b>4.</b>	<b>Discussions .....</b>	<b>87</b>

4.1 MISC at XFELs .....	87
4.2 Structure dynamics of pCA chromophore after photoexcitation.....	89
4.3 Primary structural photoresponse of phytochrome .....	90
4.4 TR-SFX at the high repetition XFEL, EuXFEL.....	93
4.5 Pump-probe and MISC experiments at XFELs .....	94
4.5.1 From oscillations on ultrafast timescales to transient intermediates .....	95
4.5.2 Rapid ligand screening to support structure-based drug design .....	98
4.5.3 From TR-SFX to time-resolved single particles imaging .....	99
<b>5. References.....</b>	<b>100</b>
<b>Curriculum Vitae .....</b>	<b>118</b>

## LIST OF FIGURES

Figure 1.1 Unit cell with lattice parameters	3
Figure 1.2 Comparison of peak brilliance of storage rings and Free electron Lasers	5
Figure 1.3 Radiation power and micro-bunching of the electrons	9
Figure 1.4 Core structure of penicillin and cephalosporin	11
Figure 1.5 $\beta$ -lactamase enzyme with a catalytic cleft	11
Figure 1.6 Reaction of $\beta$ -lactamase with Ceftriaxone	12
Figure 1.7 Photoactive Yellow Protein and its chromophore	14
Figure 1.8 Phytochromes and Biliverdin	16
Figure 2.1 Schematic crystallization phase diagram	19
Figure 2.2 PYP microcrystals	21
Figure 2.3 Scattering of X-rays by an atom	26
Figure 2.4 Two-dimensional crystal lattice	28
Figure 2.5 Ewald sphere construction	30
Figure 2.6 Experimental setup for SFX	36
Figure 2.7 Experimental setup for Pump probe TR-SFX	39
Figure 2.8 Experimental setup for MISC	41
Figure 2.9 A rectangular crystal of dimension $2a$ , $2b$ and $2c$	42
Figure 2.10 Dark calibration image produce by a detector	46
Figure 2.11 An example of a diffraction pattern	47
Figure 2.12 A GUI of Cheetah	48
Figure 2.13 A layouts of online monitoring programs	49

Figure 2.14 Found spots and Predicted spots in a diffraction pattern	50
Figure 2.15 Experimental setup for data collection using T-junction	58
Figure 2.16 Crystal packing for BlaC	59
Figure 2.17 Injector used in MISC	60
Figure 2.18 Pulse patterns produced by EuXFEL	61
Figure 2.19 Experimental setup used at EuXFEL for pump probe TR-SFX	63
Figure 3.1 Electron density map of BlaC	67
Figure 3.2 Electron density map of second experiment on BlaC	73
Figure 3.3 Pump-probe experimental scheme at 1.13 MHz	74
Figure 3.4 DED in PYP chromophore	75
Figure 3.5 Hitrate and indexing rate	76
Figure 3.6 Pump-probe experimental scheme at 564 kHz	78
Figure 3.7 DED in PYP chromophore	79
Figure 3.8 DED in PYP chromophore pocket of a time series	82
Figure 3.9 DED in biliverdin of phytochrome	85
Figure 4.1 Changes in difference signal as a function of laser fluences	92
Figure 4.2 Intermediates in photocycle of PYP	96
Figure 4.3 Schematic experimental setup for multi-compound screening	98

## LIST OF TABLES

Table 1.1 Comparison of various XFELs	10
Table 2.1 Seven crystallographic points	24
Table 2.2 Diffusion times for various crystal sizes	44
Table 2.3 List of beamtimes that I attended	57
Table 2.4 Time-points collected in this dissertation	64
Table 3.1 Data collection and refinement statistics from 1 <sup>st</sup> MISC experiment	66
Table 3.2 Data collection and refinement statistics for shard crystal form	69
Table 3.3 Data collection and refinement statistics for needle crystal form	71
Table 3.4 Data collection statistics for experiment at EuXFEL at 1.13 MHz	77
Table 3.5 Data collection statistics for experiment at EuXFEL at 564 KHz	80
Table 3.6 Data collection statistics for phytochrome at SACLA	84
Table 4.1 Geometry of pCA chromophore	89



## ACKNOWLEDGEMENTS

This accomplishment would not have been possible if a lot of eminent people had not provided enough support and advice. Firstly, I would like to convey my sincere respect and gratitude to my advisor, Professor Marius Schmidt. I am very grateful for his mentorship, scientific supports, constant encouragement, and motivations as well as the freedom he has provided throughout the work period. His enthusiasm, passion, and dedication for scientific work has motivated me to continually improve and produce the best of my efforts. His methodological way of advising and out-of-box thinking has made entire graduate studies exciting and enjoyable. My choice to join his group for graduate studies couldn't have been better.

I express my gratitude to Prof. Abbas Ourmazd, Prof. Peter Schwander, Prof. Ionel Popa, and Prof. Andy Pacheco for their willingness to be members of my Ph.D. committee. I have a sincere appreciation for providing valuable comments, suggestions, and not to mention, their precious time.

Next, I would like to extend my acknowledgments to former and present members of the Schmidt Lab. I want to thank Jason Tenboer, Chris Kupitz, Tyler Norwood, and Glenn Niles for their continuous support and help when I was a novice to the field. I also want to thank my colleague and a good friend Ishwor Poudyal for all those wonderful suggestions and an important discussion we had. Also, to the newest member of the lab, Tek Narsingh Malla. I am very optimistic about your future and wish you all the best for your future endeavors. I also like to thank Ms. Kate Valerius (Physics graduate program assistant) for her continuous help throughout the graduate studies period.

I couldn't remain without acknowledging my friends Dhruva Adhikari, Narayan Dahal, Manish Shrestha, and once again, Ishwor Poudyal. We came to UW-Milwaukee together from Nepal and started this wonderful journey of graduate studies. I am very grateful to you guys for your friendship and wonderful support. I would also like to include Netra Sharma, Resham Rana, and Dammar Badu in the same list. My stay in Milwaukee and studies at UW-Milwaukee would not have been this fun and amazing without all of you guys. I also want to express my deepest gratitude to my sibling brothers, whom I am greatly indebted for providing me continuous support and inspiration.

Finally, I want to dedicate this degree to my loving wife Dikshya. Your continuous support, love, and understanding have always instilled positivity and energy in me and have helped to tackle every obstacle in my life.

# 1. Introduction

Proteins are complex biomolecules that play many significant roles in an organism such as promoting motility, harnessing solar energy, stopping infections, etc. (Alper, 1994; Brand, Ligtenberg and Veerman, 2014). They are composed of one or more polypeptide chains. Each chain consists of amino acids that are covalently bound together. The sequence of amino acids determines the unique three-dimensional structure and the specific function of a protein.

There are different types of proteins: enzymes, for example catalyze biological reactions; photoactive proteins perform several types of reactions when photoactivated (Björn, 2015), while other proteins transmit signals within an organism. Different types of proteins have different and unique functions. To perform their specific functions, proteins are required to change their structures. In order to understand how proteins function, we must determine their three-dimensional molecular structures and then observe how their structures change to accomplish their functions.

There are three main techniques to determine the three-dimensional structure of biological macromolecules. They are Nuclear Magnetic Resonance (NMR) spectroscopy, cryo-Electron Microscopy (EM), and Macromolecular X-ray Crystallography (MX). As of 07/28/2019, there are 154,243 structures (Source: <https://www.rcsb.org/stats/growth/overall>) deposited in the protein data bank (Berman et al., 2002). Out of these structures, 3,538 and 12,701 are determined by NMR and EM respectively, while 137,586 (~89% of total structures deposited) are determined by MX. This shows that the MX is so far the most prolific technique used to determine 3-dimensional molecular structures.

The advent of X-ray Free Electron Lasers (XFELs) has initiated a new era in MX. XFELs produce extremely brilliant X-ray pulses with a duration of tens of femtoseconds. They help to outrun radiation damage and allow the collection of high-resolution data from microcrystals. The femtosecond X-ray pulses can explore ultra-short time scales. This allows the investigation of structural displacements on the femtosecond time scale. In this dissertation, we show how to investigate protein structural changes by employing a method called time-resolved serial femtosecond crystallography (TR-SFX).

There are several XFELs in the world – each with a unique data analysis pipeline. To successfully process the data obtained at these XFELs, various techniques and methods are needed to be applied in sequential order. These techniques are completely different for different XFELs. This study also focuses on the data acquisition techniques applied at several XFELs with the emphasis on the data analysis methods used to study protein dynamics.

## **1.1 X-rays and Crystallography**

### **1.1.1 A Brief History of X-rays and Crystallography**

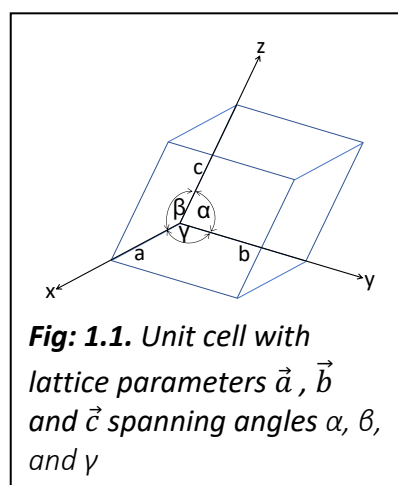
X-rays were discovered by Wilhelm Conrad Roentgen in 1895 which won him the newly created Nobel prize in 1901 (Röntgen, 1898). Subsequently, in 1912, Max Von Laue discovered that crystals diffract X-rays (Friedrich, Knipping and Laue, 1913). This work made it possible to derive interatomic distances within the crystals. W.L Bragg and his father W.H Bragg showed that determination of atomic structure is possible by solving the crystal structures of NaCl, KCl, KBr, and KI (Bragg, 1912, 1913, 1914; Bragg, W.H, Bragg, 1949). Nearly 45 years later, the first three-dimensional protein structure on sperm whale myoglobin was solved by John Kendrew and

coworkers (Kendrew, 1958; Kendrew et al., 1958). A few years later, the structure of hemoglobin was solved by Max Perutz and colleagues (Perutz et al., 1960). These discoveries, along with several other, became a milestone of X-ray crystallography (Blake et al., 1965; Brodersen et al., 2000).

X-rays are forms of electromagnetic radiation. They have wavelengths much shorter than UV radiation. They are highly energetic and penetrating waves that cannot be focused by lenses. They have photon energies between 100eV to 100keV. X-rays scatter weakly when they interact with molecules. But when the molecules are arranged in a 3-dimensional lattice, diffraction is amplified in certain directions. Accordingly, large and well-ordered crystals are required for MX. When these crystals are exposed to X-ray radiation, diffraction patterns are recorded by an X-ray sensitive detector. The diffraction patterns are processed and analyzed to determine the structures of macromolecules and their complexes at near-atomic resolution.

### 1.1.2 Crystals

Crystals consist of periodic arrays of atoms or molecules in all three dimensions. They are repetitive translations of the unit cell content along principal axes. The unit cell is the smallest repeating unit of a crystal which completely reflects the symmetry of the crystal. The lengths of the unit cell vector and the angles between them are called cell parameters or



lattice constants. The lengths are usually denoted as a, b, and c spanning the angles  $\alpha$ ,  $\beta$ , and  $\gamma$  (Fig. 1.1). For further, in-depth discussion of crystal properties, see Material and Methods (Section 2.2.1).

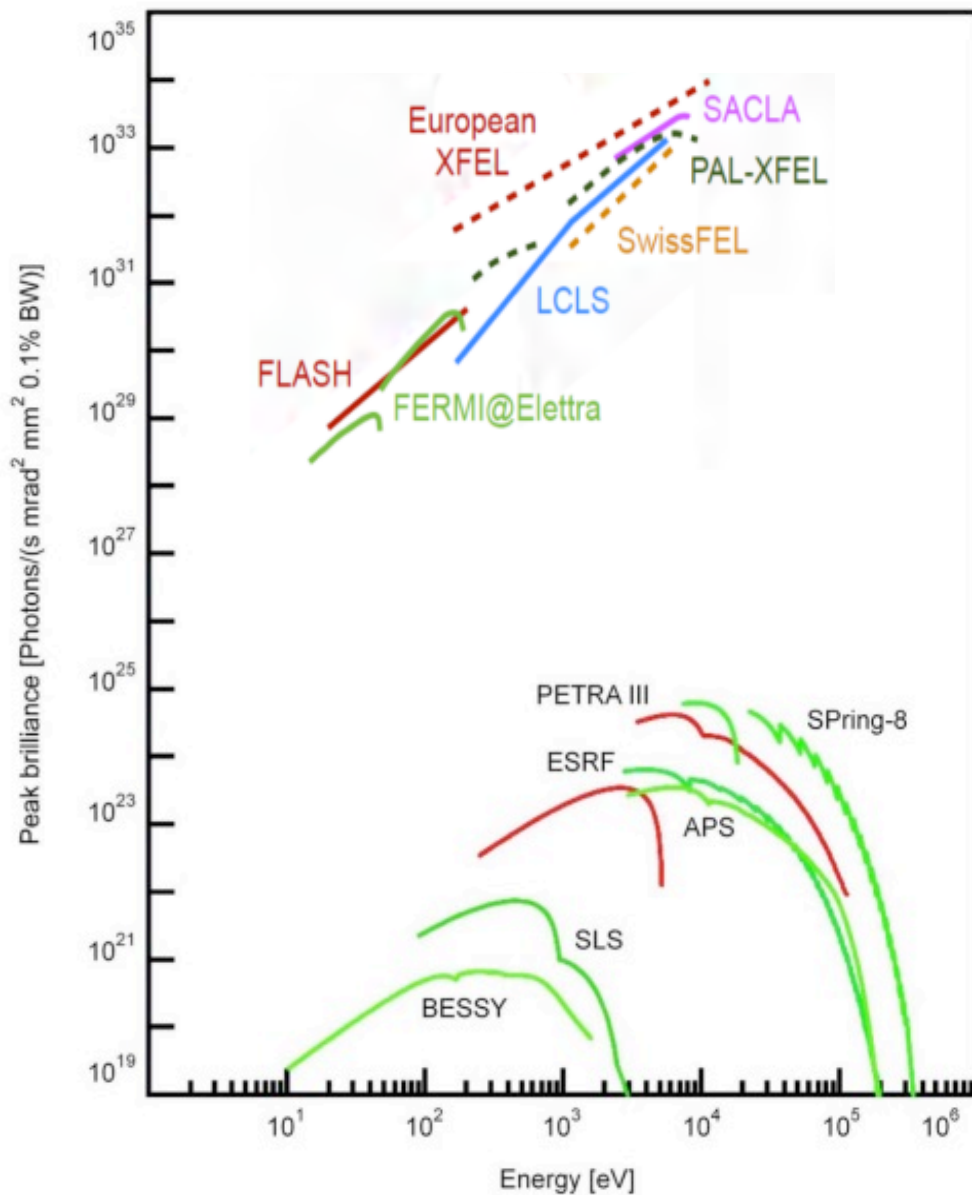
### 1.1.3 Protein Crystallization

Unlike crystals of small molecules, protein crystals are soft and sometimes gel-like. Protein crystallization is typically carried out in vitro by using purified proteins. Crystals might also occur naturally. For example, sperm whale myoglobin crystals were discovered on ship decks and used by Kendrew to determine the first three-dimensional protein structure (Kendrew, 1958; Kendrew et al., 1958). Recently, protein crystals were even grown within living cells and their structures were solved (Redecke et al., 2013). However, up to date, only a few naturally occurring protein crystals are reported (Doye and Poon, 2006; Koopmann et al., 2012; Duszenko et al., 2015).

### 1.2 Advanced X-ray sources

With relatively large crystals, of around a few hundreds of microns in diameter, it is possible to collect high resolution data by using laboratory X-ray sources. For small crystals, it is essential to use sources with more powerful X-ray intensities.

In the last few decades, the brilliance of X-ray sources has increased dramatically. The current strongest synchrotrons, the “Positron Elektron Tandem Ring Anlage” (PETRA) III at the Deutsches Elektronen Synchrotron (DESY) and the “Super Photon ring-8 GeV” (Spring-8) produce around  $10^{10}$  photons per second on the average. They have peak brilliance of around  $10^{25}$  photons/s/mrad<sup>2</sup>/mm<sup>2</sup>/0.1%BW. The current strongest XFEL – the European XFEL (EuXFEL) in Hamburg produces an average flux of  $10^{12}$  photons per pulse. This amounts to around  $10^{16}$  photons per second. The EuXFEL produces femtosecond pulses with a peak brilliance of  $10^{33}$  photons/s/mrad<sup>2</sup>/mm<sup>2</sup>/0.1%BW (Tschechschner et al., 2017).



**Fig. 1.2.** Comparison of peak brilliance of storage rings and Free electron Lasers as a function of Photon Energy (eV). The dashed lines represent recently operational XFELs. (Source: modified from Robinson et al., 2010)

(FLASH: Free Electron Laser at Hamburg, Germany, FERMI: Free Electron laser Radiation for Multidisciplinary Investigations, Germany, BESSY: Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung, Germany, SLS: Swiss Light Source, Switzerland, ESRF: European Synchrotron Radiation Facility, France)

### 1.2.1 Third Generation Synchrotron Sources

Synchrotrons are brilliant X-ray sources where electrons travel around a closed path. Third-generation synchrotron sources use special arrays of magnets called insertion devices, also called wigglers or undulators (Winick et al., 1981). Insertion devices cause electrons to travel on an undulating trajectory. This results in the emission of highly brilliant and forward-directed synchrotron radiation. This radiation spans over a wider bandwidth of wavelengths and produces a so-called “pink” polychromatic beam. The beamline “14 ID” at Advanced Photon Source (APS), Chicago generates X-rays tunable to 7-20 keV with a band-pass of  $\Delta E/E = 5 \times 10^{-2}$ . The beamline delivers 100ps short X-ray pulses with  $3 \times 10^{10}$  photons per pulse within a beam size of  $0.020 \times 0.015 \text{ mm}^2$  (FWHM) (Source: <https://biocars.uchicago.edu/facilities/14-id-beamline/>). With these X-ray pulse characteristics, (i) diffraction patterns from single pulses are difficult to obtain, (ii) processes faster than about 100 ps cannot be observed, and (iii) radiation damage cannot be outrun (see further down).

### 1.2.2 Radiation Damage in MX

Currently, synchrotrons are the most widely used X-ray sources for protein crystallography experiments. Since X-rays are ionizing, they cause significant radiation damage to protein crystals. The mechanism of X-ray radiation damage is not fully understood yet and is subject to intense research activity (Owen, Rudiño-Piñera and Garman, 2006; Ravelli and Garman, 2006; Garman, 2010).

In 1962, Blake and Phillips showed that the radiation damage to the proteins is linearly proportional to the dose received by crystals (Blake and Phillips, 1962). They showed that each 8



keV photon is capable of disrupting 70 molecules and disordering about 90 protein molecules more. They also suggested that the protein molecules suffer significant structural damage from the dose. This was later confirmed by J.R Helliwell in 1988 using data collected from des-pentapeptide insulin crystals (Helliwell, 1988). General effects of radiation damage are an increase in temperature factor (a decrease of the Debye Waller Factor, see below), the change in unit cell parameter, the decrease in diffraction power, and the loss of high-resolution data.

There have been extensive studies in this field that suggest that the X-ray radiation protein crystal damage is stochastic in nature (Nave, 1995). Several other studies show that the effect of the radiation damage can be reduced by cryo-cooling the crystals (Blake and Phillips, 1962; Owen, Rudiño-Piñera and Garman, 2006; Roedig et al., 2016). But cryo-cooling the sample might lead to freezing artifacts. This might interfere with the determination of the native protein structure. Most importantly, protein dynamics are effectively frozen out, and the protein is not functional. This motivated scientists to use femtosecond pulses from X-ray Free Electron Laser (XFEL) that outrun radiation damage due to the famous “Diffraction before destruction” principle (see below) and produce essentially radiation damage-free diffraction patterns at ambient temperatures.

### **1.2.3 Diffraction before Destruction at XFELs**

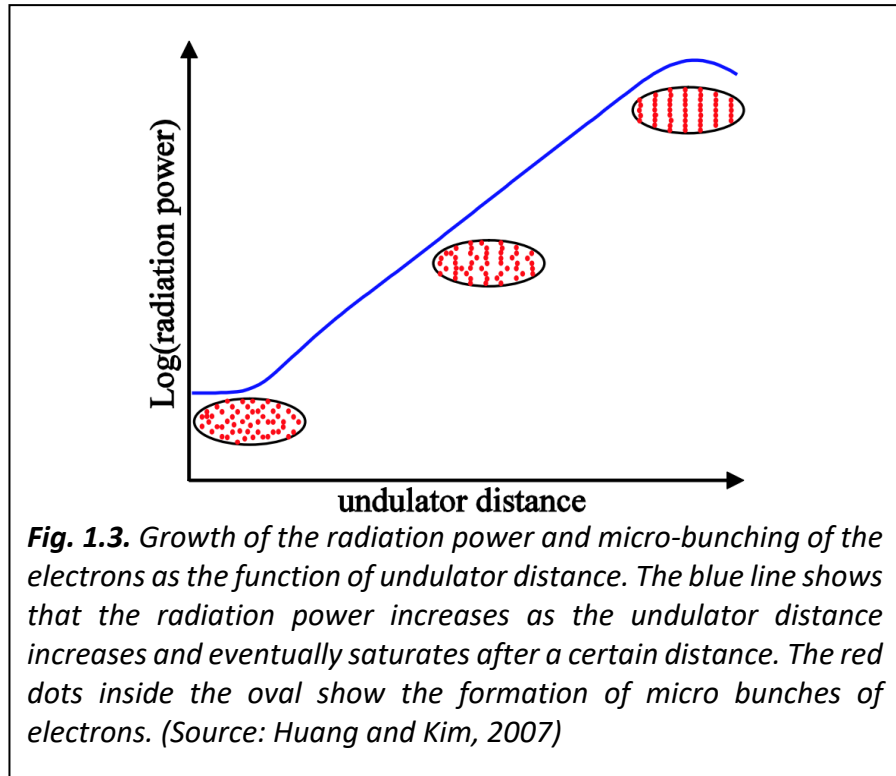
The idea of “Diffraction before destruction” to outrun radiation damage was first mentioned by J.C Solem in 1986 (Solem, 1986). He suggested that the radiation damage barrier may be extended at a very high dose rate and a very short exposure time. If the X-ray pulses are very intense and extremely short, the sample crystals are destroyed, yet diffraction patterns are collected right before their destruction. When the pulses are short enough, diffraction patterns

are essentially free of radiation damage. This concept was predicted with computer simulations in 2000 (Neutze et al., 2000) and experimentally shown in 2006 using a microscopic cowboy figure etched in silicon nitride (Chapman et al., 2006). After the commissioning of the first X-ray Free Electron Laser (XFEL), the Linac Coherent Light Source (LCLS) in 2009, the “diffraction before destruction” principle was also experimentally verified for protein microcrystals (Chapman et al., 2011).

As synchrotrons produce X-ray pulses of a few hundred picoseconds, important processes that occur on significantly shorter time scales are difficult if not impossible to explore. Ultra-short pulses produced by the XFELs can be used to explore the reaction processes that occur even at the femtosecond time scale. In addition to this, even though the pulses are ultra-short, high resolution diffraction patterns can still be collected at the XFELs (Lomb et al., 2011; Boutet et al., 2012)

Although the XFEL has the word “Laser” in its name, the mechanism of generating XFEL radiation is conceptually different from that of a conventional optical laser. An XFEL consists of a linear accelerator and a long undulator. Bunches of electrons are accelerated to relativistic velocities using a linear accelerator. During the passage through the accelerator, each electron bunch is passed through several bunch compressors, where electrons of different energies even out. The electrons then enter a long undulator where they emit electromagnetic radiation. The passage through the undulator causes electrons to interact with their own, previously emitted, electromagnetic field. This causes the electrons to “micro-bunch”. The micro bunches are separated by the X-ray wavelength  $\lambda$ . They coherently emit X-rays. The radiation power then does not scale with  $N$  ( $N$  is the number of electrons) but with  $N^2$ . With a sufficiently long

undulator and a bright electron beam, the radiation power is amplified exponentially until saturation (Fig. 1.3). This process is called self-amplified spontaneous emission (SASE) (Huang and Kim, 2007).



#### 1.2.4 Comparison of various XFELs

The first hard XFEL source of the world, the LCLS is located at Stanford Linear Accelerator Center (SLAC), Menlo Park, California. It became operational in 2009. So far, six more XFELs became operational since then. The LCLS-II at SLAC and the Shanghai high repetition rate XFEL and Extreme Light Facility (SHINE) in China are still under construction. Table 1.1 compares various XFELs in the world. European XFEL (EuXFEL) at Hamburg, Germany, LCLS at California, and Spring-8 Angstrom Compact free electron Laser (SACLA) in Japan are the XFEL facilities where we performed our experiments. This dissertation will primarily focus on these XFELs.

**Table 1.1.** Comparison of various, upcoming and existing XFEL sources in the world in the chronological order. The XFELs in the bold box represent the XFELs that we visited for experimental beamtimes. (Source: [https://www.xfel.eu/facility/comparison/index\\_eng.html](https://www.xfel.eu/facility/comparison/index_eng.html))

	LCLS	SACLA	SwissFEL	PAL-XFEL	EuXFEL	LCLS II CuRF	LCLS II SCRF	SHINE
Location	USA	Japan	Switzerland	South Korea	Germany	USA	USA	China
Start of Commissioning	2009	2011	2016	2016	2017	2019	2020	2025
Accelerator Technology	N.C	N.C	N.C	N.C	S.C	N.C	S.C	S.C
Length of Facility (km)	3	0.75	0.74	1.1	3.4	3	3	3.1
Max E. beam energy (GeV)	14.3	8.5	5.8	10	17.5	15	5	8
Minimum Wavelength (Å)	1.5	0.8	1.0	0.6	0.5	0.5	2.5	0.5
Frequency	120	60	100	60	27k	120	1M	1M
Pulse length (fs)(FWHM)	20-200	30-200	20-100	50-200	2-100	10-50	10-50	-
Detectors	CS-PAD	MPCCD	JUNGFRAU	PAL	AGIPD	ePix	-	-
Detector pixel Size (µm × µm)	110×110	50×50	75×75	40×40	200 ×200	100 ×100	-	-
No. of pixels/unit	184×195	512×1024	512×1024	512×128	64×64	384×352	-	-
Max. Num of Photons	3500	2700	10000	15000	10000	11000	-	-
Framerate (Hz)	120	30	2000	60	3250	120	-	-

Note: **CuRF** – Copper Radio Frequency, **SCRF** - Super Conducting Radio Frequency, **SwissFEL** – Switzerland’s X-ray free-electron Laser, **N.C** – Normal Conducting, **S.C**-Super Conducting, **AGIPD** – Adaptive Gain Integrating Pixel Detector, **CSPAD** – Cornell-SLAC Pixel Array Detector, **MPCCD** – Multiport charged coupled detector, **JUNGFRAU** – adjusting gain detector for the Aramis user station at SwissFEL, **PAL** - Pohang Accelerator Laboratory

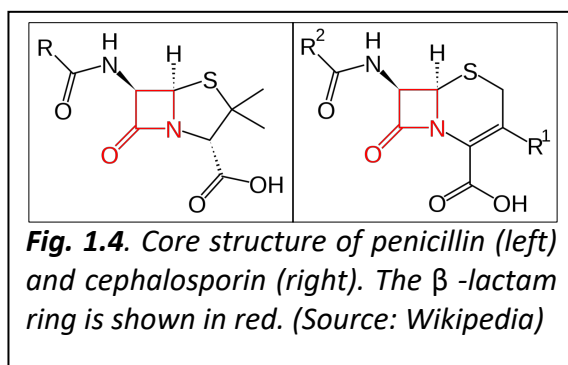
### 1.3 Proteins investigated in this dissertation.

In this work, several proteins were investigated at various XFELs. They are either model systems to assess the feasibility of TR-SFX experiments at new XFELs, or they are used to answer specific biological questions.

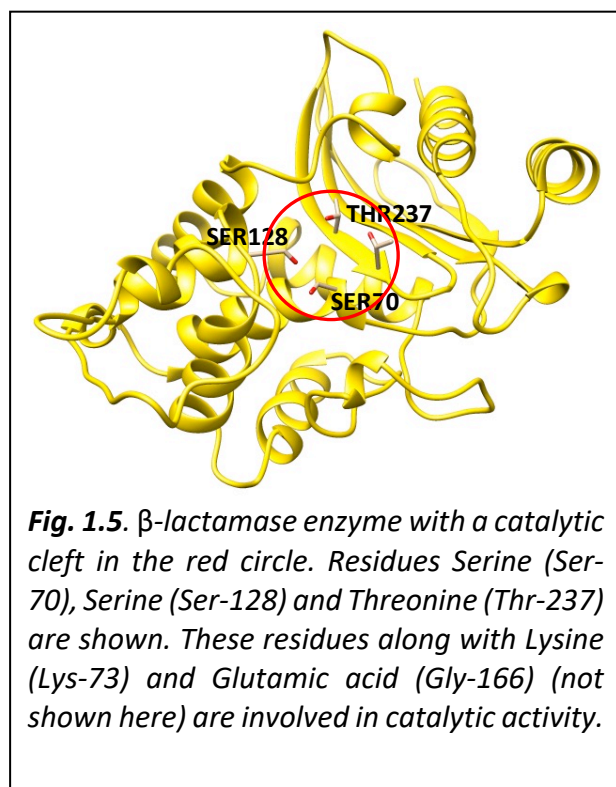
#### 1.3.1 $\beta$ -Lactamase

$\beta$ -lactams are classes of antibiotics that contain the  $\beta$ -lactam ring in their chemical structure (Fig. 1.4). They are frequently used for the prevention and treatment of bacterial infections (Lewis, 2013). They include a broad range of molecules such as penicillin derivatives (penams), cephalosporins (cephams) (Page, 2004). The  $\beta$ -lactam ring is also found in other classes of antibiotics such as penems and carbapenems (Perez-Llarena and Bou, 2009; Palzkill, 2013).

Ceftriaxone (CEF) is a 3<sup>rd</sup> generation cephalosporin antibiotic. Unlike, penicillin which features a 5-membered thiazolidine ring, CEF features a 6-membered dihydrothiazine ring attached to the  $\beta$ -lactam ring (Fig. 1.4). It is a broad-spectrum antibiotic used in the treatment of various bacterial infections. However, bacterial resistance

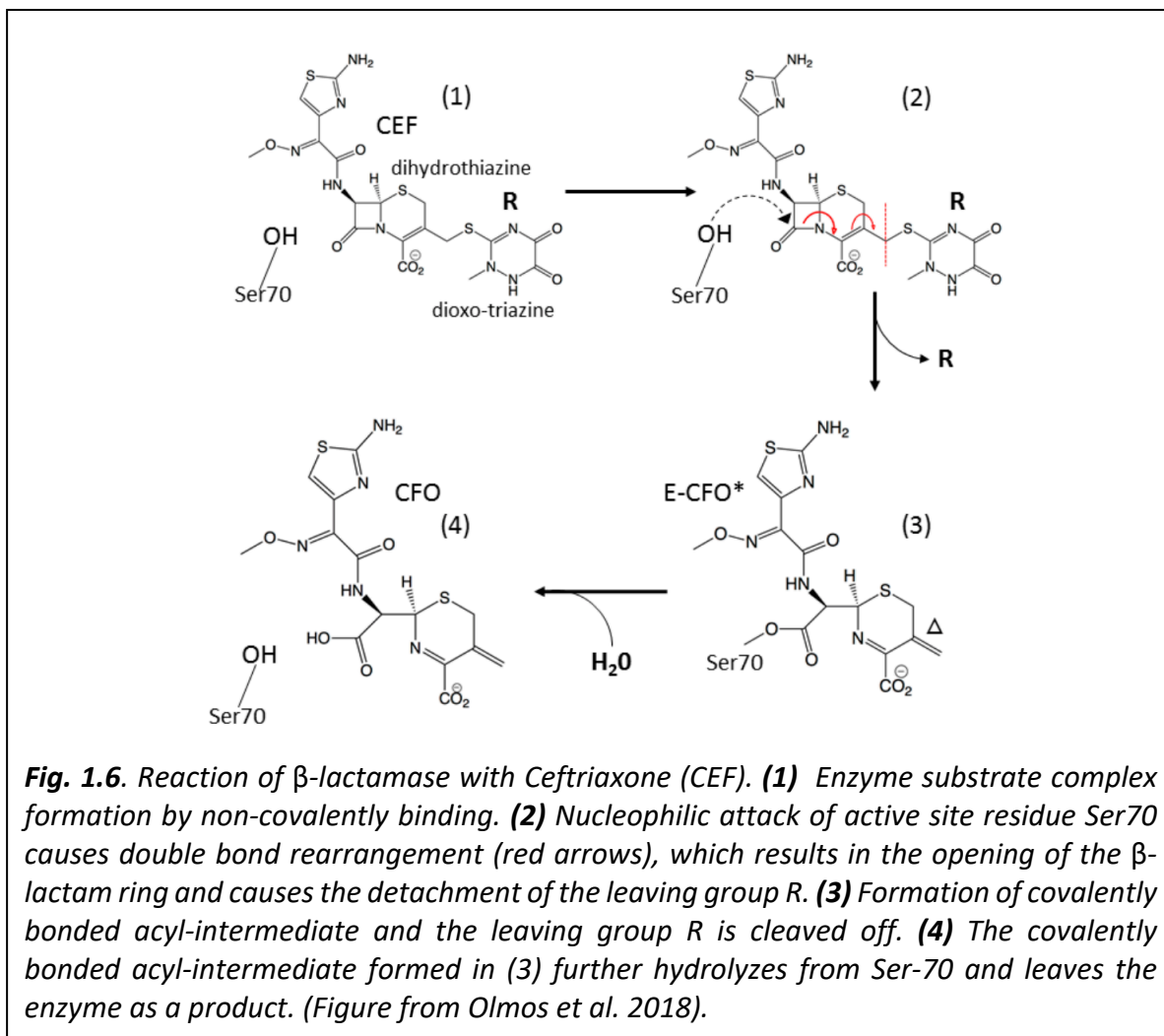


**Fig. 1.4.** Core structure of penicillin (left) and cephalosporin (right). The  $\beta$ -lactam ring is shown in red. (Source: Wikipedia)



**Fig. 1.5.**  $\beta$ -lactamase enzyme with a catalytic cleft in the red circle. Residues Serine (Ser-70), Serine (Ser-128) and Threonine (Thr-237) are shown. These residues along with Lysine (Lys-73) and Glutamic acid (Gly-166) (not shown here) are involved in catalytic activity.

against CEF was observed shortly after its widespread use (Kong, Schneper and Mathee, 2010).  $\beta$ -lactams are among the most widely used antibiotics due to their broad activity against both gram-negative and gram-positive pathogens(Zeng and Lin, 2013).



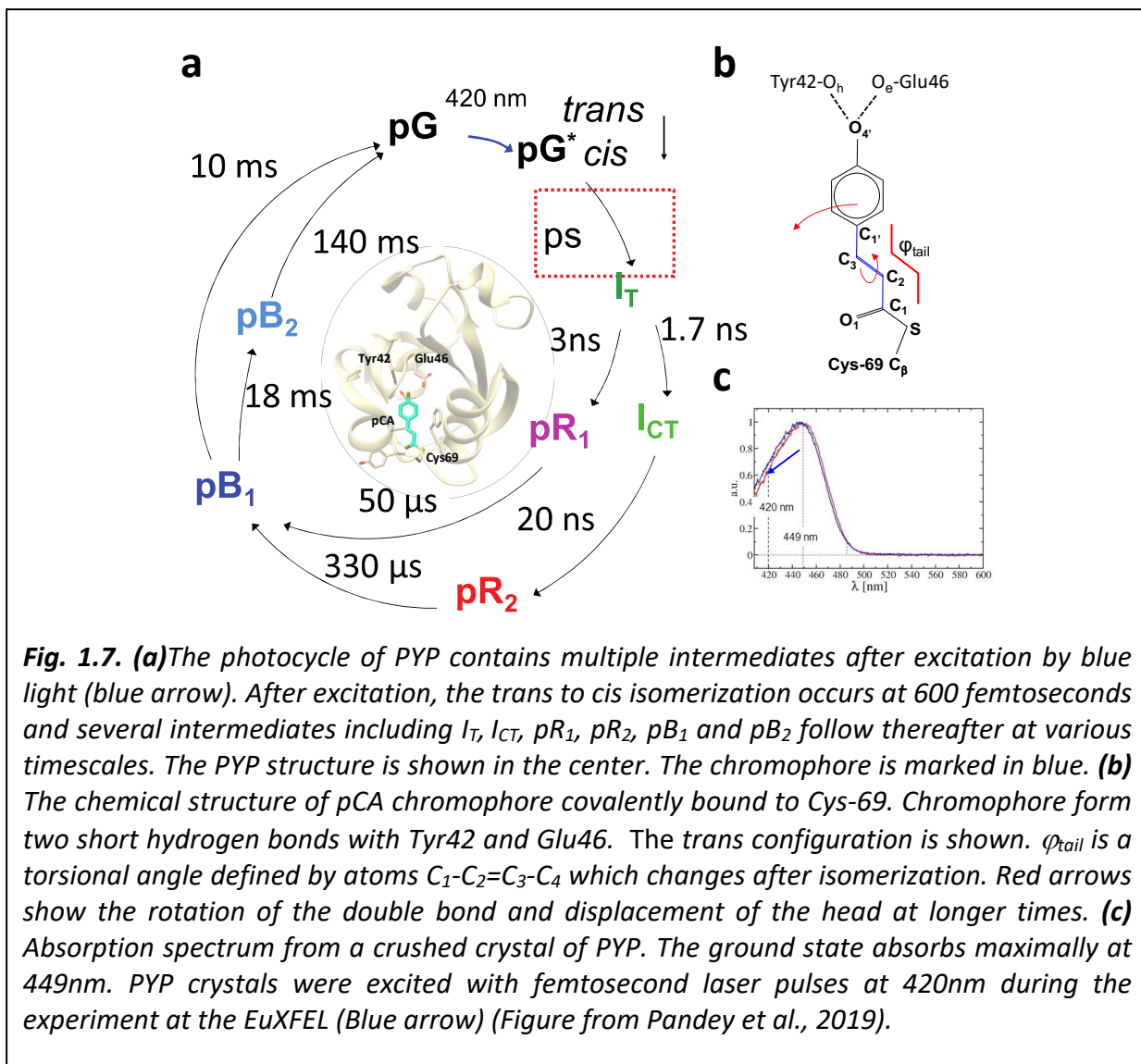
$\beta$ -lactamases are enzymes produced by bacteria. They provide multi-resistance to  $\beta$ -lactam antibiotics. The  $\beta$ -lactamase (BlaC), we used for our investigation, is an enzyme produced by *Mycobacterium tuberculosis* (Mtb). Mtb is a causative agent of tuberculosis which is the reason for nearly around a million deaths per year. BlaC has a molecular weight of about 30 KDa and has around 280 amino acids. Fig. 1.5 shows the  $\beta$ -lactamase enzyme with the catalytic cleft.

The evolution of BlaC has rendered the various classes  $\beta$ -lactam antibiotics ineffective, thereby evading the treatment of tuberculosis (Tremblay, Fan and Blanchard, 2010).

BlaC catalyzes the opening of the  $\beta$ -lactam ring of CEF by a nucleophilic attack of the active site via Ser-70 residue (Fig. 1.6(1)). As a result, a reaction intermediate forms where the antibiotics species is covalently bound to the enzyme. This species is called the acyl-intermediate (Fig. 1.6(2)) (Tremblay, Fan and Blanchard, 2010). The  $\beta$ -lactam ring-opening inactivates the antibiotic. The leaving group R (Fig. 1.6(2-3)) is cleaved off (Boyd and Lunn, 1979; Dubée et al., 2012). After about 1 s, the inactive antibiotic hydrolyzes and leaves the enzyme (Fig. 1.6 (d)). In this thesis, we outline TR-SFX experiments on the interaction of CEF with the BlaC on different time scales. With these time-resolved experiments, we aim to determine the previously unknown structures of the enzyme-substrate complex and the acyl-intermediate with CEF. This time-resolved experiment was performed at the LCLS, USA.

### **1.3.2 Photoactive Yellow Protein**

Photoactive yellow protein (PYP) is a water-soluble blue-light photoreceptor. It was discovered in 1985 by Meyer in a purple sulfur bacterium called *Halorhodospira halophila* (Meyer, 1985). This small molecule of 14KDa is known for its yellow color and its reactivity with light. PYP displays a typical PAS (Per-ARNT-Sim) domain fold. The PAS domain is important in sensing and signal transduction (Pellequer et al., 1998; Taylor and Zhulin, 1999). Due to the potential biomedical significance of the PAS domain's structural change, its mass producibility, and interesting physicochemical properties (Kort et al., 1996), PYP belongs to one of the most widely studied proteins in MX.



PYP contains a chromophore called para-hydroxycinnamic acid or para-coumaric acid (pCA). The chromophore is covalently bound to Cys-69 and forms 2 short hydrogen bonds with Tyr-42 and Glu-46 (Fig. 1.7 b) (Kort et al., 1996; Anderson, Crosson, and Moffat, 2004). Upon absorption of blue light, PYP enters a reversible photocycle. As shown in Fig. 1.7a, the photocycle starts with the *trans* to *cis* isomerization of the  $C_2=C_3$  double bond (Fig. 1.7b) within a few hundred femtoseconds after photoactivation. Thereafter, several intermediates are populated

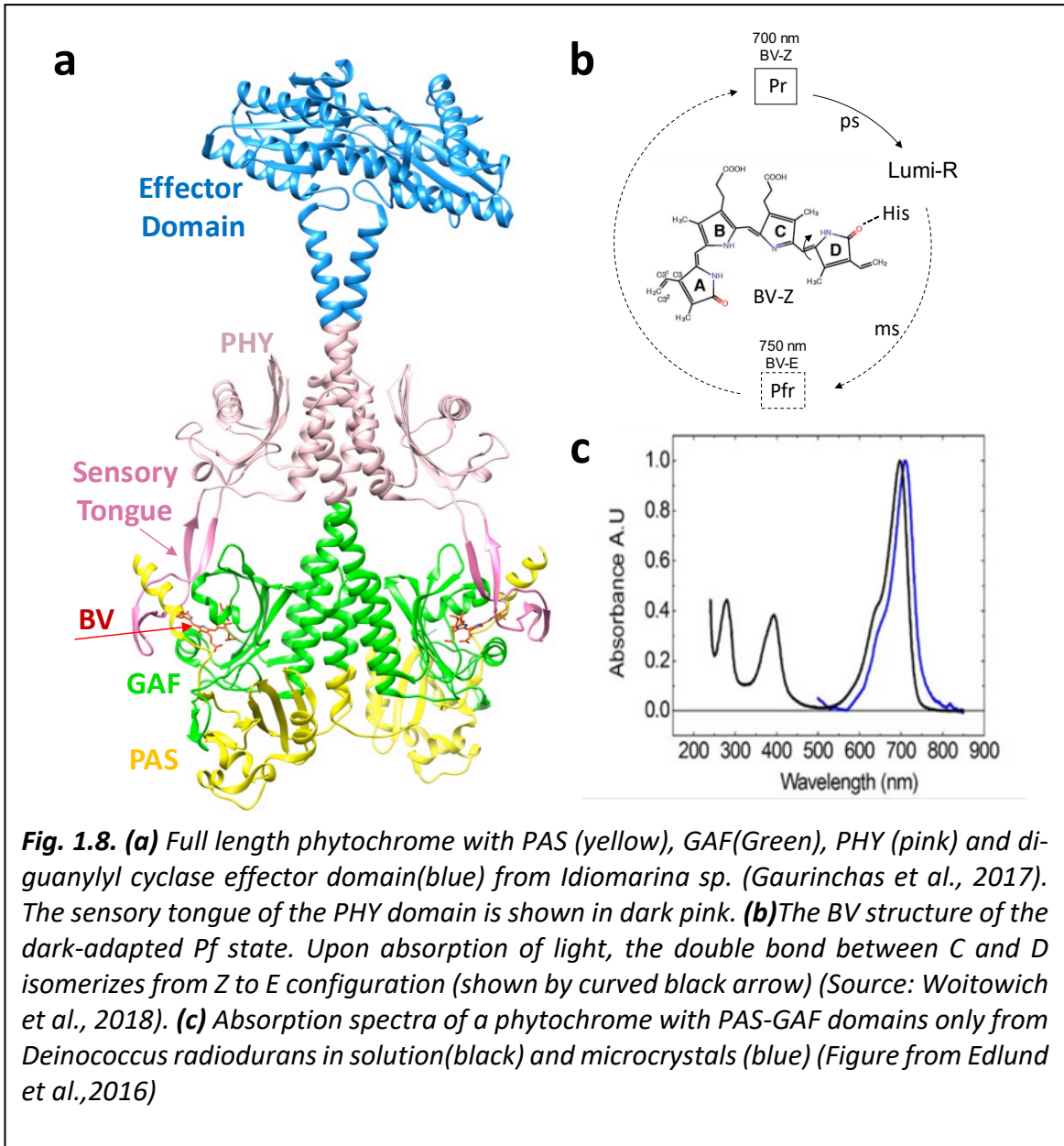


on different timescales (Tenboer et al. 2014; Pande et al. 2016; Tripathi et al. 2012; Schmidt et al. 2013). Here, we show the results of a TR-SFX experiment on PYP on the structurally poorly characterized picosecond regime. This experiment was a pioneering TR-SFX experiment performed at the recently operational EuXFEL, Germany (Wiedorn et al., 2018; Mancuso et al., 2019).

### 1.3.3 Phytochrome

Phytochromes belong to the family of red-light photoreceptors that control light responses in various organisms. They were first characterized in plants with homologs in both photosynthetic and non-photosynthetic bacteria known as bacteriophytochromes (Davis, Vener and Vierstra, 1999). Phytochromes detect light and trigger intracellular signaling cascades, which regulate many light-dependent phenomena such as shade avoidance and seed germination in plants (Bae and Choi, 2008; Hughes, 2013; Pandey, Poudyal and Malla, 2020). They also control the expression of photosynthetic proteins in plants and cyanobacteria (Gan et al., 2014).

The photosensory core module (PCM) of phytochromes consists of PAS (PER, ARNT, SIM), GAF (cGMP phosphodiesterase, adenylate cyclase, FhIA) and PHY (Photochrome-specific GAF related) domains (Takala *et al.*, 2016). An effector domain with enzymatic activity is covalently attached to the PHY domain (Auldridge and Forest, 2011). The PAS and GAF domains typically harbor a covalently bound bilin chromophore called biliverdin (BV) (Woitowich *et al.*, 2018). BV is made of heme-derived open-chain tetrapyrrole (pyrrole rings A-D) (Fig. 1.8). Similarly, phycocyanobilin (PCB) is found in cyanobacterial phytochromes and phytochromobilin (PΦB) in plant phytochromes. BV, PCB, and PΦB are chemically very similar and differ mainly in the number and the arrangement of their conjugated double bonds (Schmidt et al., 2007).



**Fig. 1.8. (a)** Full length phytochrome with PAS (yellow), GAF(Green), PHY (pink) and di-guanylyl cyclase effector domain(blue) from *Idiomarina* sp. (Gaurinchas et al., 2017). The sensory tongue of the PHY domain is shown in dark pink. **(b)**The BV structure of the dark-adapted Pf state. Upon absorption of light, the double bond between C and D isomerizes from Z to E configuration (shown by curved black arrow) (Source: Woitowich et al., 2018). **(c)** Absorption spectra of a phytochrome with PAS-GAF domains only from *Deinococcus radiodurans* in solution(black) and microcrystals (blue) (Figure from Edlund et al.,2016)

Classical phytochromes change their structures upon absorption of red light. The absorption of light results in a so-called Z to E isomerization of its central chromophore. Z stands for the German word “zusammen” and means close-by or together, just another expression for a cis configuration. E denotes the German word “entgegen” which translates to opposite, or trans. The chromophore isomerizes about the C15=C16 double bond between bilin rings C and D (Fig.

1.8b). The signal is then transmitted to the effector domain (Burgie et al., 2014) (Fig. 1.8b). Phytochromes make the structural transition from the Pr state (red absorbing) to the Pfr state (far-red absorbing) after they absorb the light of suitable wavelength (Rockwell et al., 2009; Takala et al., 2014).

In this dissertation, we show results from a time-resolved experiment on the chromophore binding domain (CBD), consisting only of the PAS and GAF domains, of the *Deinococcus radiodurans* bacterial phytochrome (DrBphP). Results of this time-resolved crystallographic experiment were the first published on any phytochromes. Experiments were performed at SACLA, Japan.

## 2. Materials and Methods

### 2.1 Protein Purification and Crystallization

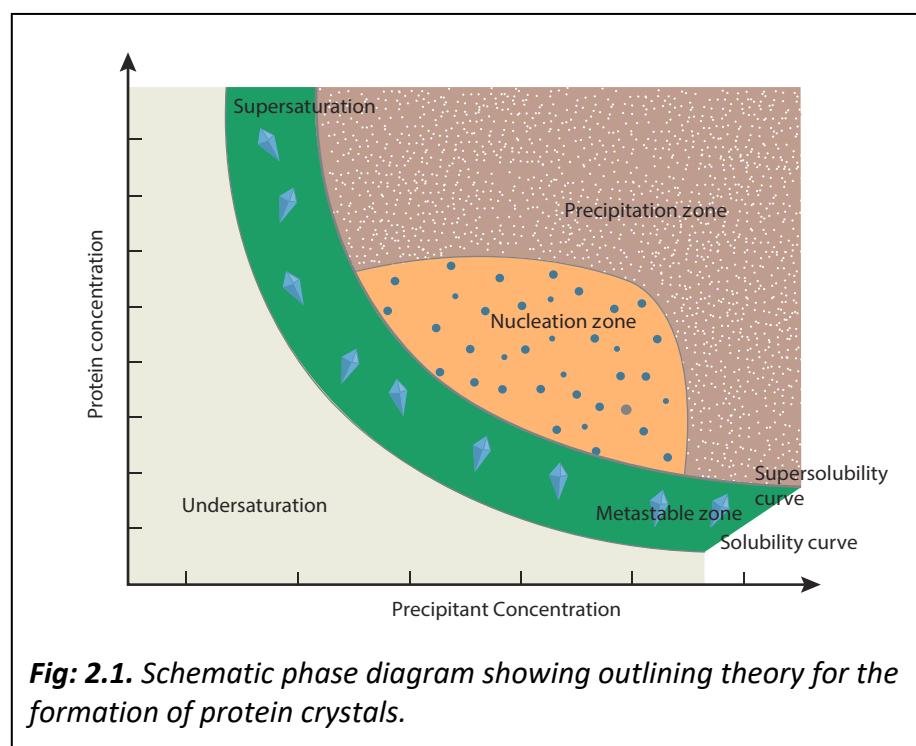
Protein purification is a vital component of any X-ray crystallographic experiments. Only highly purified proteins can crystallize properly. The methods for purification are different for different proteins and depend on the nature of proteins. They will be discussed separately for each protein that we studied in the upcoming chapters.

Although protein crystallization is a difficult process, the principle idea is very simple. Crystallization is carried out by adding a precipitating agent to the purified protein solution. This gradually removes the hydration shell of the protein thereby allowing interactions between protein molecules. There may be three outcomes of a crystallization method (McPherson and Gavira, 2014). They are:

- i) The addition of the precipitating agent may show no reaction. This may happen when the concentration of protein and/or precipitant is too low, or the agent does not precipitate the protein at all.
- ii) The addition of the precipitating agent may result in amorphous precipitation. The amorphous precipitation could be due to protein denaturation caused by unspecific hydrophobic interactions.
- iii) The protein crystallizes.

The formation of crystals depends on various factors such as temperature, pH value, and concentration. Conformational heterogeneity may also be one of the biggest limiting factors for successful crystallization. Stable, and homogenous protein solutions are required.

Protein crystallization can be understood using a phase diagram shown in Fig. 2.1. The phase diagram is divided into undersaturation, supersaturation, metastable, nucleation, and precipitation zones. The protein is stable in the undersaturation zone. Crystallization is initiated in the nucleation zone. Crystal growth occurs in the metastable zone. In the precipitation zone, the protein forms an amorphous precipitate.



Nowadays, high-throughput robotic screening has made the mapping of phase diagrams much easier. However, it is still impossible to accurately predict the outcome. High throughput trial and error methods are still abundantly used (<https://hwi.buffalo.edu/high-throughput-crystallization-center/>).

The following sections provide brief information about the purification and crystallization methods for the various proteins that we used for our investigations.

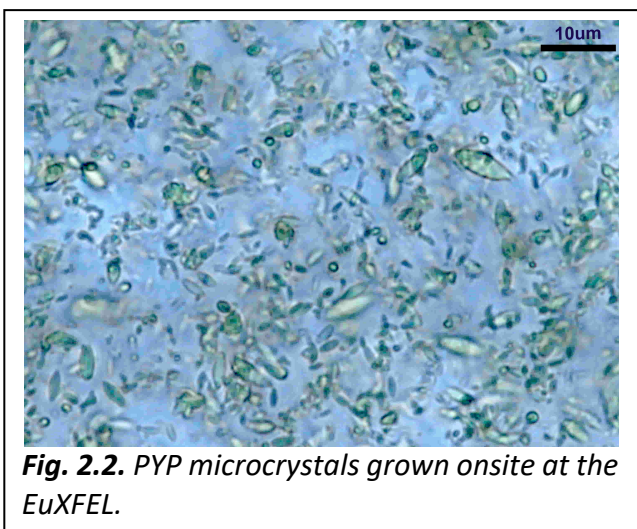
### 2.1.1 PYP

Overexpression and purification were carried out at the UW-Milwaukee. DNA from the original bacterium *H. halophila* is expressed in the host bacterium *E. coli* (Meyer, 1985; Meyer et al., 1987; Kort et al., 1996). Expression is achieved at 16° C for ~22 hours in 7.5L Luria-Bertani (LB) culture medium. Bacterial cells were harvested by centrifugation. Cell walls were ruptured by sonication. Activated p-coumaric anhydride (pCA) was added. pCA binds autocatalytically to the apo-PYP. PYP was purified using His60 Ni Superflow resin. After cleaving off the N-terminal His-tag, pure PYP was obtained by anion chromatography on Q-sepharose. The purified PYP was stored at -80°C at concentrations varying from 30-50 mg/ml. The PYP was shipped to Hamburg for onsite crystallization. The crystallization was carried out shortly before the beamtime at the EuXFEL.

For crystallization, the purified PYP is concentrated to 100mg/ml. 2-3mls of the concentrated PYP are then transferred into a glass vial. Under vigorous stirring, 4 mol/L and pH 7.0 sodium malonate is added into the protein solution to a final concentration of 3.3 mol/L. The required volume of Malonate can be calculated as:

$$volume(malonate) = volume(PYP) \times \frac{conc\_mal(desired)}{conc\_mal(stock) - conc\_mal(desired)},$$

where conc\_mal is the concentration of sodium malonate. In our protocol the stock concentration of malonate is 4 mol/L and the final (desired) concentration is 3.3 mol/L.



**Fig. 2.2.** PYP microcrystals grown onsite at the EuXFEL.

The protein becomes cloudy right after addition of malonate. The suspension is stirred for around 8 hours in a closed vial and then is set aside to rest for another 24 hrs. After that, the solution is centrifuged in a 15ml conical tube for a few minutes at 8000 x g. The centrifugation causes the microcrystals to swim towards the top of the

solution, while clear and seemingly transparent solution will remain at the bottom. The clear solution is carefully removed with a syringe and exchanged with 2.8 mol/L sodium malonate, pH 7 to reduce the viscosity of the crystal suspension to facilitate liquid injection into the X-ray interaction region. The crystals are then filtered twice with 10 $\mu$ m stainless steel filter. The filtered suspension is used for the TR-SFX experiment.

### 2.1.2 BlaC

BlaC was overexpressed, purified, and crystallized by Christopher Kupitz, a postdoc at the time with Prof. Schmidt. Expression was achieved at 25 $^{\circ}$  C for ~20 hrs in 5L of terrific broth (TB). Bacterial cells were harvested using centrifugation. Harvested cells were lysed using a sonicator. Cellular debris were removed via centrifugation. The BlaC was purified using His60 Ni Superflow resin and size exclusion chromatography with Sephadex 200. The His-tag was not cleaved off as it did not interfere with crystallization. The final purified BlaC was then stored at -80 $^{\circ}$ C and shipped to the LCLS for onsite crystallization.

The crystallization was carried out right before the beamtime at the LCLS. Two different types of crystal forms were produced for the experiment. They are needle crystal form and shard crystal form. Needle crystal forms were grown with 35% PEG 1000 and sodium acetate (pH 5.0) using free interface method (Salemme, 1972). Shard crystal forms were grown with ammonium phosphate (pH 4.1) as a precipitant using a stir method. Crystallization processes for both crystal forms are described in Olmos et al., 2018.

### **2.1.3 Phytochrome**

The experiment on the DrBphP-CBD was performed in collaboration with the Westenhoff group from the University of Gothenburg. The Westenhoff group produced and crystallized the protein. I was involved in data analysis and data interpretation and did not participate in purification and crystallization.

## **2.2 X-ray Crystallography**

### **2.2.1 Crystal lattice and Reciprocal Lattice**

The unit cell parameters (Section 1.1.2) are combined in several ways to 14 Bravais lattices. These lattices are grouped into 7 fundamental lattice types or crystallographic systems namely triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal, and cubic. 14 Bravais lattices can be also categorized as primitive(P), body-centered(I), face-centered (F), base centered (A, B or C), and rhombohedral (R).

The symmetry operators of a crystalline lattice are rotations, mirrors, inversion, glide planes, and screw axes. The sets of symmetry operations that leave the central point unmoved while mapping a point lattice into itself are called crystallographic point groups. There are 32



available point groups. Table 2.1 shows the 7 crystal systems and their point and Laue groups. There are 230 space groups for 3-dimensional crystals. However, only 65 space groups are possible for protein crystals. It is because amino acids have a chiral center around C $\alpha$  atoms. Due to this chirality, glide planes, mirrors, or inversion centers cannot be observed.

A crystallographic experiment determines interatomic distance within a crystal. As diffraction always add center of symmetry in the data, instead of 32 point groups, only 11 Laue groups represent the symmetry of diffraction patterns. The others remain hidden. There have been extensive studies on crystals, their symmetry, and space groups. For detailed information on these topics, refer to the “International Tables of Crystallography” Vol A(Hahn, 2006).

**Table 2.1:** Seven Crystallographic Lattice points and their corresponding Bravais lattices, point and Laue groups.

Crystal System	Parameters	Bravais Lattice	Point Groups	Laue Groups
<b>Cubic</b>	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	Primitive(P) Body Centered(I) Face Centered(F)	$23, m\bar{3}$	$m\bar{3}$
			$432, \bar{4}3m, m\bar{3}m$	$m\bar{3}m$
<b>Tetragonal</b>	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	Primitive (P) Body Centered(I)	$4, \bar{4}, 4/m$	$4/m$
			$42m, 4/mmm$	$4/mmm$
<b>Orthorhombic</b>	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	Primitive (P) Body Centered (I) Base Centered (C) Face Centered(F)	$222, mm2, mmm$	$mmm$
			$3, \bar{3}$	$\bar{3}$
			$32, 3m, \bar{3}m$	$\bar{3}m$
			$3, \bar{3}, 321, 312, 31m, 3m1, \bar{3}1m, 6, \bar{6}, 6/m$	$6/m$
<b>Hexagonal</b>	$a = b \neq c$ $\alpha = \beta = 90^\circ,$ $\gamma = 120^\circ$	Primitive (P)	$622, 6mm, 6m2, 6/mmm$	$6/mmm$
			$2, m, 2/m$	$2/m$
<b>Monoclinic</b>	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$	Primitive (P) Base Centered (C)	$2, m, 2/m$	$2/m$
<b>Triclinic</b>	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$	Primitive	$1, \bar{1}$	$\bar{1}$

The periodic crystal lattice is represented as a reciprocal lattice in angular space also known as the reciprocal space. The reciprocal lattice plays an important role in the studies of periodic structures. If  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$  are denoted as lattice (or basis) vectors for the directions of translation inside the 3-dimensional crystal in real space, then their corresponding translations

in reciprocal space are represented as reciprocal lattice vectors  $\vec{a}^*$ ,  $\vec{b}^*$  and  $\vec{c}^*$  respectively. The reciprocal lattice vectors can be calculated as:

$$\vec{a}^* = \frac{\vec{b} \times \vec{c}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \quad (2.1a)$$

$$\vec{b}^* = \frac{\vec{c} \times \vec{a}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \quad (2.1b)$$

$$\vec{c}^* = \frac{\vec{a} \times \vec{b}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \quad (2.1c)$$

From these above equations we obtain the following conditions:

$$\vec{a} \cdot \vec{b}^* = \vec{a} \cdot \vec{c}^* = \vec{c} \cdot \vec{b}^* = \vec{b} \cdot \vec{c}^* = \vec{b} \cdot \vec{a}^* = \vec{c} \cdot \vec{a}^* = 0 \quad (2.2a)$$

$$\vec{a} \cdot \vec{a}^* = \vec{b} \cdot \vec{b}^* = \vec{c} \cdot \vec{c}^* = 1 \quad (2.2b)$$

Equations 2.2a suggest that  $\vec{a}^*$  is perpendicular to the plane formed by  $\vec{b}$  and  $\vec{c}$ ,  $\vec{b}^*$  is perpendicular to the plane formed by  $\vec{c}$  and  $\vec{a}$ , and  $\vec{c}^*$  is perpendicular to the plane formed by  $\vec{a}$  and  $\vec{b}$ . Additionally, the equations 2.2b are used to solve the magnitude of reciprocal lattice vectors. These relations also suggest that the longer the translation in real space, the smaller will be the distance between lattice points in reciprocal space.

### 2.2.2 Scattering of X-rays from an Atom

When electromagnetic radiation interacts with atoms, electrons oscillate in the presence of the electric field and emit dipole radiation. The scattering depends on the number of electrons and the position of the electrons in an atom.

The probability of finding an electron at any point  $\vec{r}$  is given by the electron density  $\rho(\vec{r})$ .

The electron density  $\rho(\vec{r})$  of the atom scatters the X-ray within an infinitesimal volume element  $dV$ . Scattering and then interference results from each partial wave.

Consider only two volume elements, as shown in figure 2.3. The total path difference ( $\Delta$ ) between waves scattered by

these two elements is given by

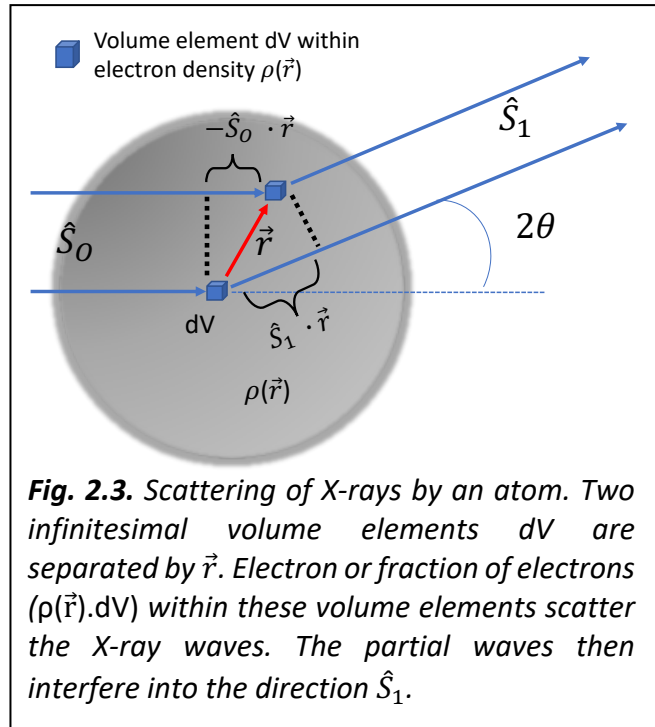
$$\Delta = (\hat{S}_1 \cdot \vec{r} - \hat{S}_0 \cdot \vec{r}) = (\hat{S}_1 - \hat{S}_0) \cdot \vec{r} = \vec{S} \cdot \vec{r}$$

And the phase difference  $\phi$  is given by

$$\phi = \frac{2\pi}{\lambda} \Delta = 2\pi(\vec{H} \cdot \vec{r}), \text{ where } \vec{H} \text{ is } \frac{(\hat{S}_1 - \hat{S}_0)}{\lambda}.$$

Finally, the atomic form factor,  $f$ , can be obtained by integrating over the electron density of the entire atom,

$$f = \sqrt{\sigma_e} \int_{V(\text{atom})} \rho(\vec{r}) e^{2\pi i \vec{H} \cdot \vec{r}} d\vec{r}, \quad 2.3$$



**Fig. 2.3.** Scattering of X-rays by an atom. Two infinitesimal volume elements  $dV$  are separated by  $\vec{r}$ . Electron or fraction of electrons ( $\rho(\vec{r}) \cdot dV$ ) within these volume elements scatter the X-ray waves. The partial waves then interfere into the direction  $\hat{S}_1$ .

where  $\sqrt{\sigma_e}$  is Thompson scattering length.

Eq. 2.3 shows that the atomic form factor is the Fourier transform of the electron density of an atom. It is used for describing the scattering efficiency of a given atom in a given direction. It decreases as the scattering angle increases. The angular dependence is tabulated for each atom species in the International Table of Crystallography.

### 2.2.3 Scattering of X-rays from a Molecule

Since a molecule is an assembly of atoms, the structure factor of a molecule with N atoms is the complex addition of the scattering contribution from all atoms present in a molecule. The structure factor of a molecule is given by,

$$\mathbf{F}^{mol}(\vec{H}) = \sum_{j=1}^N f_j e^{2\pi i \vec{H} \cdot \vec{r}_j}, \quad 2.4$$

where  $f_j$  is the atomic scattering factor, and  $r_j$  the position vector of the  $j^{\text{th}}$  atom of the molecule

Similarly, the structure factor of molecules can also be written as the integral over the electron density.

$$\mathbf{F}^{mol}(\vec{H}) = \int_V \rho(\vec{r}) e^{2\pi i \vec{H} \cdot \vec{r}} d\tau, \quad 2.5$$

where  $\rho(\vec{r})$  is an electron density at position  $r$

## 2.2.4 Scattering from a Crystal

A macroscopically large protein crystal contains on the order of  $10^{12}$  to  $10^{14}$  protein molecules, each with a large number of atoms. The structure factor of a crystal consisting of  $M$  atoms can be written as,

$$\mathbf{F}^{Crys}(\vec{H}) = \sum_{m=1}^M f_m e^{2\pi i \vec{H} \cdot \vec{R}_m}, \quad 2.6$$

where  $\vec{R}_m$  is the position of each atom with respect to the origin of the coordinate system as shown in figure 2.4.  $\vec{R}_m$  can be separated into the position vector  $\vec{r}_p$  of an atom in a unit cell and the lattice vector  $\vec{R}_{lv}$  so that (Fig. 2.4),

$$\vec{R}_m = \vec{r}_p + \vec{R}_{lv}.$$

$\vec{R}_{lv}$  can be written as the linear combination of the lattice parameters  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$ .

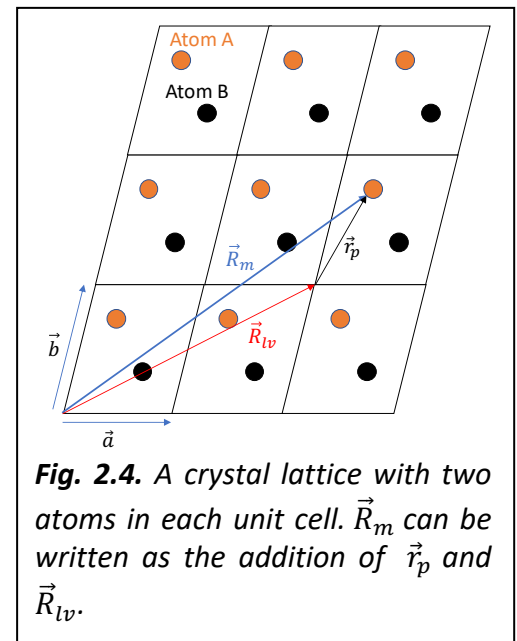
$$\vec{R}_{lv} = u\vec{a} + v\vec{b} + w\vec{c},$$

where  $u$ ,  $v$ , and  $w$  are integers.

Then Eqn. 2.4 becomes

$$\mathbf{F}^{Crys}(\vec{H}) = \sum_{n=1}^N e^{2\pi i \vec{R}_{lv} \cdot \vec{H}} \sum_{j=1}^{atoms} f_j e^{2\pi i \vec{r}_p \cdot \vec{H}}. \quad 2.7$$

The first sum in the Eqn. 2.7 is called the lattice factor. The second term is the structure factor of the content of the unit cell. The index  $j$  sums over all atoms within the unit cell which consists of protein molecules and their symmetric counterparts.



Finally, the lattice factor can be written as:

$$\sum_{n=1}^N e^{2\pi i \vec{R}_n \cdot \vec{H}} = \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} \sum_{w=0}^{W-1} e^{2\pi i \vec{H} \cdot (u\vec{a} + v\vec{b} + w\vec{c})} = \sum_{u=0}^{U-1} e^{2\pi i u \vec{H} \cdot \vec{a}} \sum_{v=0}^{V-1} e^{2\pi i v \vec{H} \cdot \vec{b}} \sum_{w=0}^{W-1} e^{2\pi i w \vec{H} \cdot \vec{c}},$$

where U, V, and W are the number of unit cells along the  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$  directions respectively such that  $U \cdot V \cdot W = N$ , which is the number of unit cells in the crystal.

Since, U, V, and W are usually large, each of the sums in the above equation is generally zero. However, if  $\vec{H} \cdot \vec{a} = h$ ,  $\vec{H} \cdot \vec{b} = k$ ,  $\vec{H} \cdot \vec{c} = l$  with h, k, and l integers, the sum will be different from zero. These are known as Laue conditions. If all three conditions are simultaneously fulfilled, the structure factor will be amplified by N.

### 2.2.5 Elastic Diffraction Condition and Ewald Sphere Construction

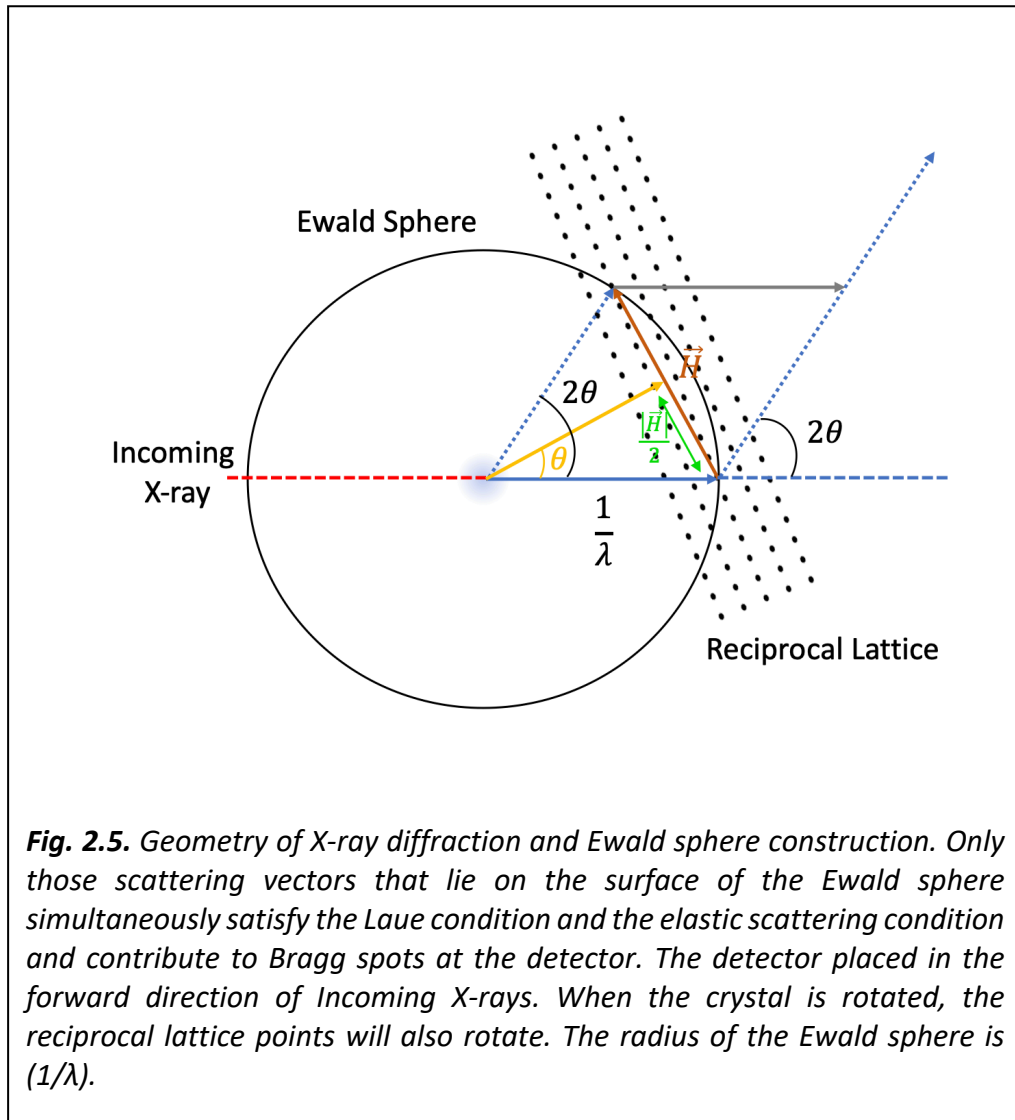
Each set of reflecting lattice planes in real space can be represented as a discrete reciprocal lattice point in reciprocal space. If  $\vec{a}$ ,  $\vec{b}$  and  $\vec{c}$  are denoted as lattice vectors for the directions of translation inside the 3-dimensional crystal in real space, then their corresponding translation reciprocal space is represented as reciprocal lattice vectors  $\vec{a}^*$ ,  $\vec{b}^*$  and  $\vec{c}^*$  respectively (Section 2.2.1). The scattering vectors can be written in terms of reciprocal lattice vectors and the Laue integers as

$$\vec{H}_{hkl} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*.$$

With this, it can be shown that

$$|\vec{H}_{hkl}| = \frac{1}{d}, \quad 2.8$$

where  $d$  is the interplanar distance between two reflecting lattice planes.



X-ray scattering is considered an elastic process. The wavelength doesn't change during the scattering process. During the scattering process, only those scattering vectors that lie on the surface of the Ewald sphere satisfy the Laue and the elastic scattering conditions and give rise to Bragg reflection. Ewald sphere has a radius of  $1/\lambda$ . In addition to this, only those lattice points that have the scattering vector smaller than  $2/\lambda$  can potentially give rise to Bragg reflections.

In general, the reciprocal lattice points do not lie on the Ewald sphere. To collect diffraction data, it is necessary to rotate the crystal. As for serial crystallography (discussed later



in this chapter), due to the randomness of microcrystals during sample flow, reciprocal lattice points lie randomly on the Ewald sphere. Therefore, a dataset can be completed just by collecting a sufficient number of diffraction patterns.

The Ewald sphere construction can be used to derive the Bragg's equation. Fig. 2.5, shows that:

$$\sin \theta = \frac{|\vec{H}_{hkl}|}{\frac{2}{\lambda}} .$$

Using  $|\vec{H}_{hkl}| = \frac{1}{d}$  from Eqn. 2.6, we can write,

$$2d \sin \theta = \lambda .$$

This equation is called Bragg's Law, where  $\lambda$  is the wavelength of the X-ray radiation, and  $\theta$  is the angle between the scattered wave and the reflecting lattice plane. If the Bragg condition is fulfilled, waves that are elastically scattered from a specific set of lattice planes (with coprime Miller indices hkl or an integer multiple) in the crystal are in phase and constructively interfere.

### 2.2.6 The Phase Problem and phasing algorithms

The main aim of a diffraction experiment is to calculate the electron density of a molecule. From Eqn. 2.3, we can see that the structure factor is the Fourier transform of the electron density of a molecule. Hence, the electron density of a molecule is the inverse Fourier transform of its structure factor.

Therefore, the electron density of the molecules can be written as

$$\rho(XYZ) = \frac{1}{V} \sum_{hkl} \vec{F}_{hkl} e^{-2\pi i(hX + kY + lZ)} , \quad 2.9$$

where  $V$  is the volume of the unit cell. The integration in the Eqn. 2.5 is replaced by summation because, ultimately due to the periodicity of the crystal, diffraction occurs in discrete directions rather than continuously.

Since  $\vec{F}_{hkl}$  is a complex vector quantity, we can write  $\vec{F}_{hkl} = |\vec{F}_{hkl}|e^{i\alpha_{hkl}}$ . And the electron density becomes

$$\rho(XYZ) = \frac{1}{V} \sum_{hkl} |\vec{F}_{hkl}| e^{-2\pi i(hX + kY + lZ)} e^{i\alpha_{hkl}}, \quad 2.10$$

where  $\alpha$  is the phase angle for each hkl plane in reciprocal space.

Since X-ray crystallography is a lens-less imaging technique, only structure factor amplitudes can be obtained from the diffraction pattern. The information about the phase is lost. This problem is known as the “phase problem” in X-ray crystallography. Eqn. 2.8 shows that the electron density of the molecule can only be calculated when both the amplitudes and the phases of the structure factors are known. Different approaches are used to solve the phase of scattered waves. Since these methods are out of the scope of this dissertation, only a brief overview will be provided.

a) Direct methods. These methods are based on the positivity and atomicity of electron density.

It is applied for a small molecule with 1000 or few atoms per unit cell. Similarly, for very small molecules with only a few atoms, the Patterson map can be used to determine the structure by “educated guessing”. While an electron density map is obtained from the Fourier transform of the structure factor  $\vec{F}_{hkl}$ , a Patterson map is obtained from the Fourier transform of Intensities  $|\vec{F}_{hkl}|^2$  which are directly available from the diffraction experiment.

- b) Isomorphous replacement is used for deriving the phases from heavy-atom derivatives. The heavy-atom derivatives are created by soaking the crystals in a solution of heavy atom compounds.
- c) Anomalous diffraction at various wavelengths is exploited for determining the phase.
- d) Another approach called “Molecular Replacement” (MR) is extensively used for phase retrieval. MR can be used if a protein structure can be found in the protein data bank that is very similar to the one under investigation. The homologous, known structure is used as a search model to find the unknown orientations and positions of the protein molecules in the crystal. Only measured intensities or amplitudes are necessary for this search. Phases are calculated from the oriented model and combined with the measured amplitudes to calculate new electron density maps to solve the unknown structure of the protein of interest. For more information on all of these methods, refer to “The Phase problem” by G. Taylor (Taylor, 2003).

### **2.2.7 Mosaicity**

Protein crystals are not perfectly ordered. They consist of many small arrays called mosaic blocks which are only approximately aligned with each other. Due to mosaicity, reciprocal lattice points are not points, but resemble spherical caps, spanning a small angular range. When such a spherical cap intersects with the Ewald sphere, the intensity of only a part of the Bragg reflection is collected. This becomes a problem since only the intensity integrated over the entire reflection range is proportional to the structure factor amplitude squared. In traditional crystallography, this is solved by rotating the crystal and integrating over the intensity distributions measured for each Bragg reflection. In serial crystallography, still exposures are collected, and rotation is not

possible. This is solved by averaging a large number of observations for the same Bragg reflection (see chapter 2.4.1).

## 2.2.8 Debye-Waller Factor and Wilson Plot

### 2.2.8.1. Debye Waller Factor.

Atoms always undergo thermal vibration about their mean position. The vibration is seen even at absolute zero temperature and its amplitude increase as the temperature increases. Due to this, the position of atoms can be slightly disordered in the random direction from their original position. Proteins are flexible molecules. The flexibility of molecules leads to structural heterogeneity i.e. the position of atoms is slightly displaced from their equilibrium position. These displacements contribute to the Debye-Waller Factor (T).

$$T = \exp \left[ -B \left( \frac{\sin \theta}{\lambda} \right)^2 \right], \quad 2.11$$

where B is the temperature factor (B-factor). If  $\langle x^2 \rangle$  is the mean square displacement of the atom, then B is given by

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

This isotropic B-factor can be factored out from the atomic form factor by using

$$f^B = f^0 \exp \left[ -B \left( \frac{\sin \theta}{\lambda} \right)^2 \right], \quad 2.12$$

where  $f^0$  is an atomic form factor without any displacement.

### 2.2.8.2 Absolute Scale

While processing diffraction data, it is important to scale the experimental data with the theoretical model or the absolute data. This scaling requires an overall scale factor and a

correction term to factor the overall B-factor attenuation. This isotropic B-factor scaling is called Wilson scaling (Wilson, 1942).

The measured, average intensity  $\langle I_{obs} \rangle$  in each resolution shell, is proportional to the absolute intensity  $\langle I_{abs} \rangle$ . i.e.  $\langle I_{obs} \rangle \propto \langle I_{abs} \rangle$

Squaring Eqn. 2.12 and taking the sum over all of the atoms in a crystal lattice, we can write

$$\langle I_{obs} \rangle = \langle I_{abs} \rangle \cdot k \cdot \exp \left[ -B \left( \frac{\sin \theta}{\lambda} \right)^2 \right]$$

where  $\langle I_{obs} \rangle = \sum_i^{atoms} f_i^B$ ,  $\langle I_{abs} \rangle = \sum_i^{atoms} f_i^0$  and k is proportionality constant.

Finally, the equation becomes,

$$\ln \frac{\langle I_{obs} \rangle}{\langle I_{abs} \rangle} = \ln k - 2B \left( \frac{\sin \theta}{\lambda} \right)^2 \quad 2.13$$

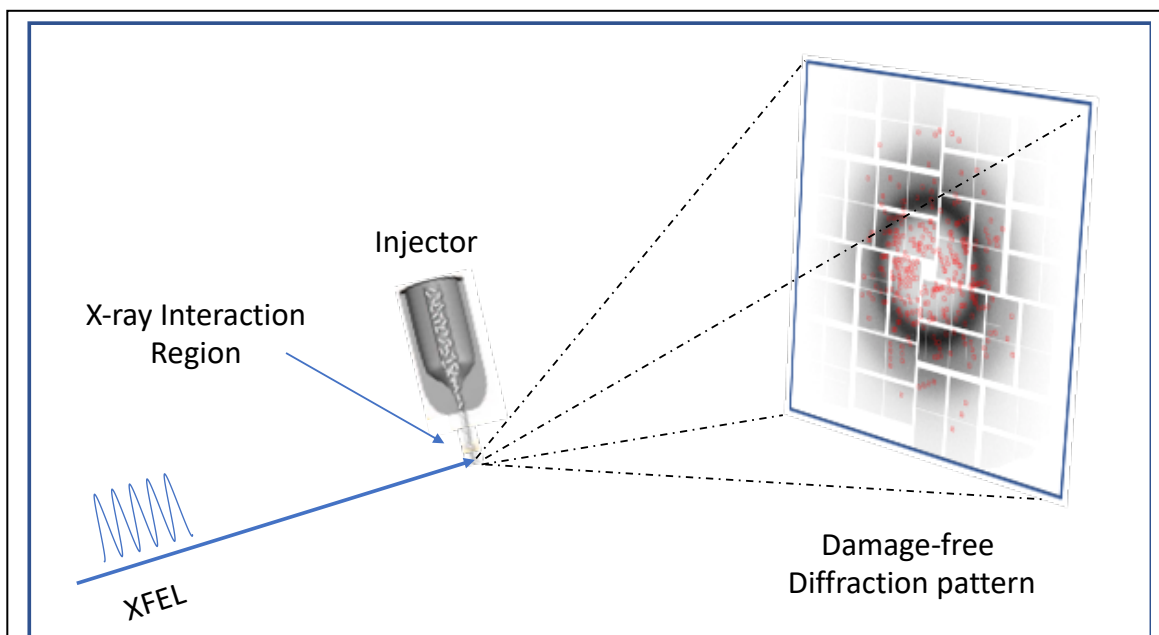
The plot of  $\frac{\langle I_{obs} \rangle}{\langle I_{abs} \rangle}$  versus  $\left( \frac{\sin \theta}{\lambda} \right)^2$  is referred to as Wilson plot. The plot should represent a straight line of form  $y(x) = \ln(k) - 2B \cdot x$ . But the Wilson plot obtained from all X-ray data generally doesn't follow a straight line because atoms are not randomly distributed within the crystal. Therefore, the slope of the fitted line provides the estimate of overall B-factor (B) for a particular set of diffraction data. Similarly, the intercept of the line provides the estimate of scale factor k.

### 2.3 Serial Crystallography (SX)

In MX, a complete dataset is collected with only a single or a few large individual crystals. If MX is performed at room temperature significant radiation damage might occur. In addition to this, growing a large single crystal may be challenging. With serial crystallography (SX), large crystals are not required, and the radiation damage is less of an issue. In this method, many small crystals are serially injected into the X-ray interaction region at room temperature (Pothineni et

al., 2014). The smaller crystals are relatively easier to produce and allow homogeneous activation by light or effective diffusion of chemical compounds.

### 2.3.1 Serial Femtosecond Crystallography



**Fig. 2.6.** The experimental setup for SFX experiment at LCLS. Randomly oriented single microcrystals in their mother liquor flowing through the gas-focused liquid jet are exposed to X-ray pulse and the diffraction patterns are collected in CSPAD detector.

Serial Femtosecond Crystallography (SFX) is a crystallographic technique developed specifically for data collection at XFEL facilities. The femtosecond X-ray pulses enable the collection of radiation damage-free data due to the “Diffraction before destruction” (Section 1.2.3) principle (Neutze et al., 2000; Chapman et al., 2011). Several other studies (Lomb et al., 2011; Barty et al., 2012) also showed that even though the X-ray dose is beyond the tolerable limit, damage is not perceived due to very short exposure time. As the crystals are destroyed already after a single exposure in SFX, a new crystal is required for every new X-ray pulse. This necessitates the delivery of crystals serially into the X-ray interaction region.

For serial femtosecond crystallography, new sample injection methods were required. The sample delivery systems such as the gas dynamic virtual nozzle (GDVN) are specifically designed for SFX (DePonte et al., 2008; Weierstall, Spence and Doak, 2012). The GDVN produces a liquid jet from microcrystals suspended in their mother liquor. The jet, a few microns in diameter, delivers the microcrystals in random orientation into the X-ray interaction region where they are probed by X-ray pulses. Hence, randomly oriented diffraction patterns will be recorded by the detector. Fig. 2.6 shows the setup for SFX experiment conducted at LCLS.

The X-ray beam produced by the XFEL can be focused to a very small focal spot of a few microns. Crystal sizes on the micrometer length scale are acceptable, which completely alleviates the requirement of large single crystals (Boutet et al., 2012; Kupitz, Basu, et al., 2014; Kupitz, Grotjohann, et al., 2014).

Serial crystallography (SX) was developed to work at the synchrotron as well. As this thesis focuses on XFEL based techniques, SX has been mentioned only briefly. The problem of radiation damage is still prevalent at synchrotrons as the pulse duration of around 100ps is not fast enough such that the diffraction-before-destruction principle holds. Two methods are used to decrease the radiation damage: (i) by using significantly larger crystals, and (ii) by using a large number of larger microcrystals. The aim of both methods is to reduce the amount of absorbed dose. In the first method, the volume is increased so that the crystal can be exposed to more X-ray radiation without surpassing the dose limit. The second method distributes the dose over a large number of small crystals. With SX, the experiments can be conducted at room temperature, which is a prerequisite for time-resolved crystallography. Recently, several SX experiments were

successfully accomplished by using monochromatic and pink-beam serial crystallography (Meents et al., 2017; Martin-Garcia et al., 2019).

### **2.3.2 Time-Resolved Serial Femtosecond Crystallography**

Time-Resolved Crystallography (TRX) utilizes X-ray crystallography to visualize reactions of proteins in four dimensions (x, y, z, and time). This technique allows the determination of the high-resolution structures of intermediates populated along the reaction pathway. With TRX, it is also possible to extract the chemical kinetics from the same set of crystallographic data (Schmidt et al., 2003, 2004, 2013; Schmidt, Ihee, et al., 2005; Schmidt, Nienhaus, et al., 2005; Schmidt, 2013, 2019; Pandey, Poudyal and Malla, 2020).

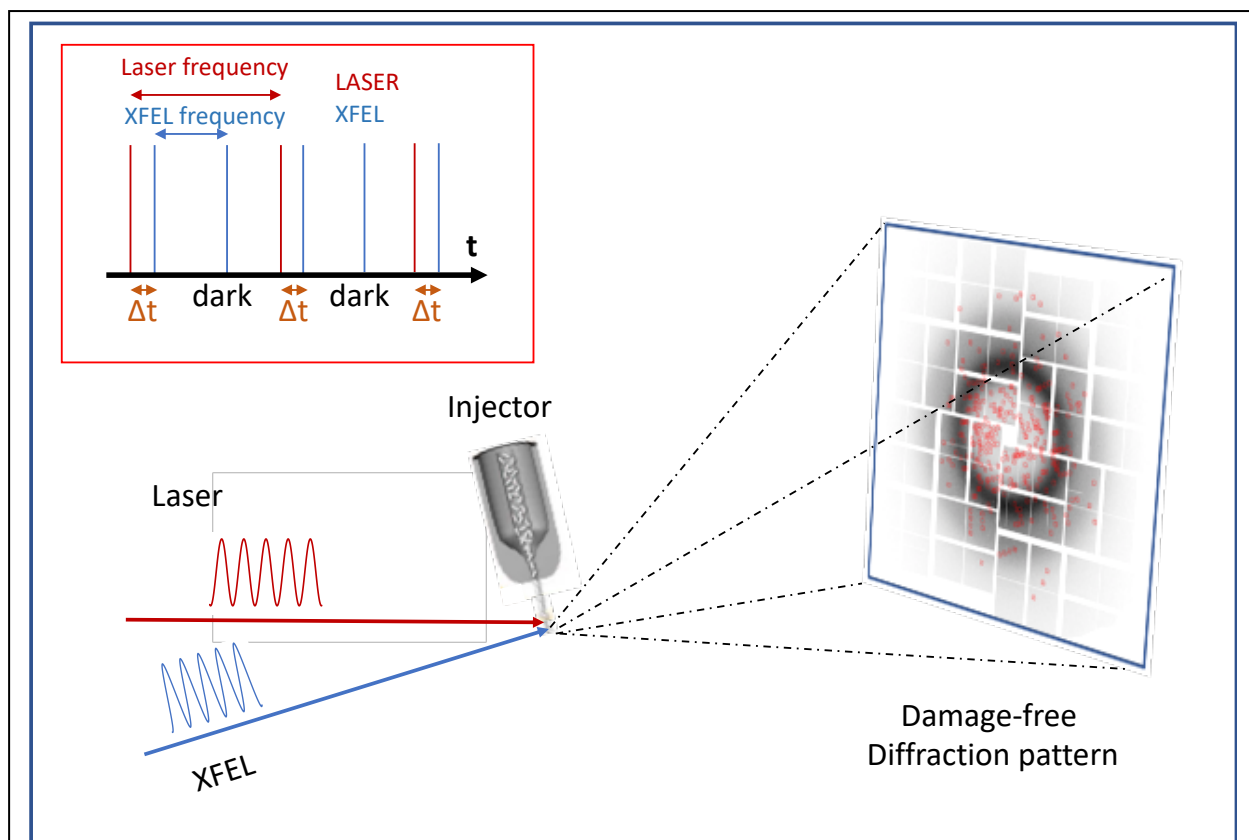
The monochromatic beam of synchrotrons requires an exposure time of hundreds of microseconds to seconds to produce a diffraction pattern (Helliwell, 1985). However, the interesting reaction intermediates are populated on much faster time scales. Time-resolved Laue Crystallography (TR-LX) was developed to probe reactions on much faster time scales (Moffat, Szebenyi and Bilderback, 1984; Schlichting et al., 1990; Ren and Moffat, 1994; Ren et al., 1999; Helliwell et al., 2002). TR-LX uses a polychromatic beam of X-rays (Moffat, 1989, 1998). With TR-LX, a reaction with a time scale of 100ps and longer can be explored.

The advent of XFELs has changed the way experiments are conducted. With the ability to probe the reaction on a femtosecond scale, it complements synchrotron sources. With XFELs, the previously inaccessible time scales faster than 100 ps including the femtosecond time scale can be probed (Tenboer et al., 2014; Barends et al., 2015; Pande et al., 2016). The technique is called time-resolved serial femtosecond crystallography (TR-SFX). TR-SFX probes reactions in



crystals which are either activated by the light source or triggered diffusion of the substrate into the crystals. The different types of TR-SFX methods are explained below.

### 2.3.3 Pump-probe TR-SFX



**Fig. 2.7.** A schematic diagram for a typical pump-probe experiment at XFELs. The macromolecular microcrystals are injected into the X-ray interaction region. The sample crystals are activated by laser light (red arrow) of appropriate energy and wavelength and then probed by X-ray pulse (blue arrow) from XFELs. The red box (top left) shows the pump-probe experimental scheme as a function of time. In the given scheme, the laser light (red bar) starts the reaction and the X-ray pulse (blue bar) probe the reaction after the time delay  $\Delta t$  (orange arrow). In the scheme, X-ray data collection as a dark is interleaved after each light activation pulse.

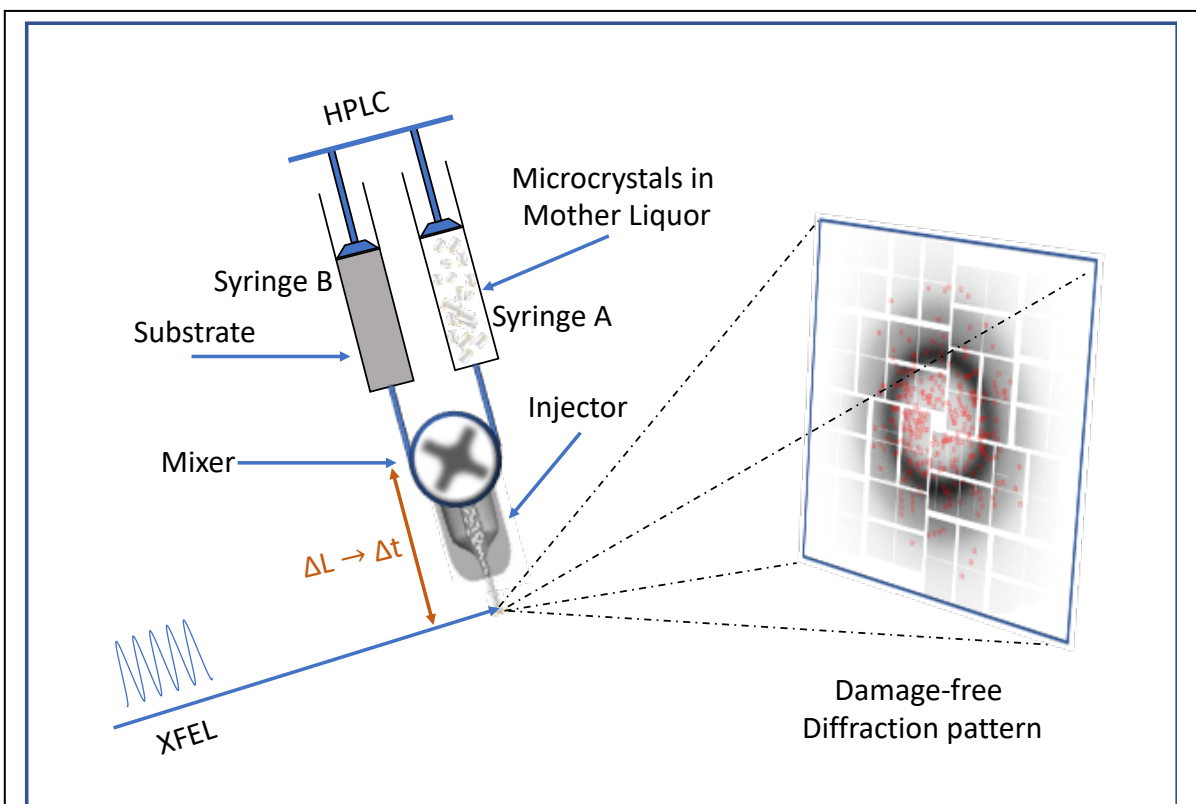
If the reaction is activated (pumped) by a light source prior to probing with X-ray pulses, this method is called pump-probe TR-SFX. An obvious requirement is that the protein has to be light sensitive. Apart from that, the reaction triggered through photoactivation should be

functionally significant. Fig. 2.7 shows a schematic diagram for a typical TR-SFX experiment at the XFELs. There have been several pioneering studies, such as on PYP (Tenboer et al., 2014; Pande et al., 2016; Pandey et al., 2019), myoglobin (Barends et al., 2015), bacteriorhodopsin (Nango et al., 2016), photosystem II (Kern et al., 2013; Kupitz, Basu, et al., 2014; Young et al., 2016; Suga et al., 2017), etc. which show that it is possible to follow and describe cyclic and non-cyclic reactions by photoactivation.

#### **2.3.4 Mix and Inject Serial Crystallography (MISC)**

Enzymes are proteins that can catalyze reactions. Ever since the first structure of the enzyme was solved (Blake et al., 1965; Johnson and Phillips, 1965), understanding the catalytic reaction of a biomolecule at the molecular level has been a dream for structural biologists. Even though several pioneering works have been done on reactions triggered by photoactivation, structural studies on irreversible enzymatic reactions are sparse.

Standard crystallography uses trapping methods to trap particular intermediates within the crystals before X-ray data are collected. These trapping methods include changing the pH value (Yamashita et al., 2003), using inhibitors (Miller et al., 2002) or flash cooling to slow down the turnover rate (Fiedler et al., 2002; Nienhaus et al., 2005; Bourgeois and Weik, 2009; Tremblay, Fan and Blanchard, 2010). Fast intermediates cannot be trapped using these methods because the time resolution is not sufficient. There had been several experiments where the enzymatic reaction was triggered by soaking an inactive, so-called caged substrate into the crystals (Schlichting et al., 1990; Goeldner and Givens, 2005). The caged compound is activated by a laser flash. The progress is then probed by an X-ray pulse. While this method has great potential, only



**Fig. 2.8.** A schematic diagram for a typical Mix and Inject experiments at XFELs. The macromolecular microcrystals in their mother liquor in syringe A and substrate in syringe B are mixed in a mixer. Both syringes are pushed using High-Performance Liquid Chromatography (HPLC) pumps. The mixture is then injected in the X-ray interaction region. During mixing, the substrate diffuses into the crystals thereby reacting with the macromolecules inside the crystals. The time delay  $\Delta t$  is determined by the length traveled by mixture ( $\Delta L$ ) and the velocity of mixture before probed by an X-ray pulse.

a few time-resolved experiments have so far been reported. These difficulties demand that a technique is developed to probe the enzymatic reaction at room temperature.

Mix and Inject Serial Crystallography (MISC) is a crystallographic technique where biomolecular crystals are mixed with a substrate and the following structural changes are probed by X-rays. Fig. 2.8 shows a schematic diagram for a typical MISC experimental setup at XFELs. As shown in the figure, crystals and substrates are mixed “on the fly”. The mixture is then injected into an X-ray interaction region and probed by an X-ray pulse.

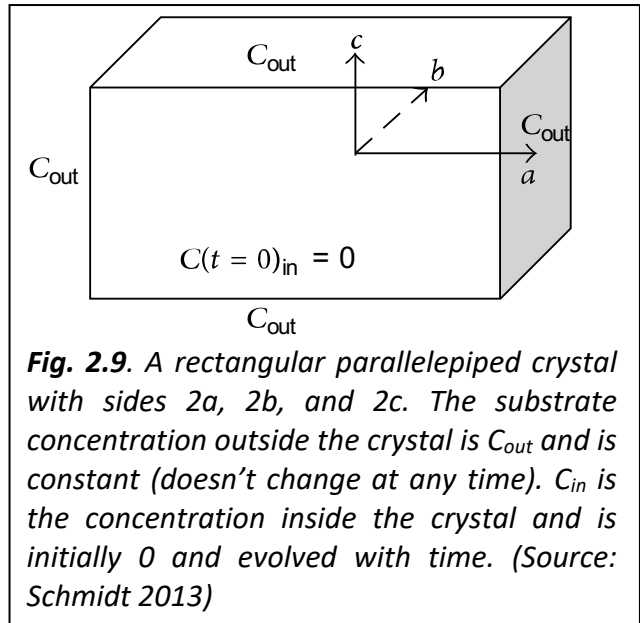
A MISC experiment is fundamentally a diffusion experiment and is based on Fick's second law. Fick's second law describes the dependence of diffusion of a substrate in solution as a function of time and is given by,

$$D\nabla^2 C = \frac{\partial C}{\partial t} , \quad 2.14$$

where  $D$  is a diffusion coefficient,  $\nabla^2$  is Laplace operator and  $C$  is the time-dependent concentration of the diffusing substrate.

Eqn. 2.14 is a second-order partial differential equation which can be solved by separation of variables with appropriate boundary conditions. The solution can further be understood by using a simple case of substrate diffusing in a rectangular parallelepiped crystal of length  $2a$ ,  $2b$ , and  $2c$  as shown in Fig. 2.9.

Further assumptions are made to idealize the situation. They are (i) Substrate concentration outside the crystal is constant all the times, (ii) substrate diffuse freely into the crystal, (iii) substrate mixes with crystal instantaneously. Using, these boundary conditions, we can get the solution to the Eqn. 2.14 which is given by.



**Fig. 2.9.** A rectangular parallelepiped crystal with sides  $2a$ ,  $2b$ , and  $2c$ . The substrate concentration outside the crystal is  $C_{out}$  and is constant (doesn't change at any time).  $C_{in}$  is the concentration inside the crystal and is initially 0 and evolved with time. (Source: Schmidt 2013)

$$C(x, y, z, t) = C_{out} \left[ 1 - \frac{64}{\pi^3} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(-1)^{l+m+n}}{(2l+1)(2m+1)(2n+1)} \times \cos \frac{(2l+1)\pi x}{2a} \cos \frac{(2m+1)\pi y}{2b} \cos \frac{(2n+1)\pi z}{2c} \times e^{-\alpha t} \right], \quad 2.15$$

where

$$\alpha = \frac{D\pi^2}{4} \left[ \frac{(2l+1)^2}{a^2} + \frac{(2m+1)^2}{b^2} + \frac{(2n+1)^2}{c^2} \right]. \quad 2.16$$

The concentration of substrate inside the crystal is dependent on the concentration outside the crystal and the time-dependent term that contains the characteristics time delay  $\tau_D$  in the exponential term in Eqn. 2.15. The characteristics time delay is given by:

$$\tau_D = \frac{1}{\alpha}$$

Such that,

$$\tau_D = \frac{4}{D\pi^2 \left[ \frac{(2l+1)^2}{a^2} + \frac{(2m+1)^2}{b^2} + \frac{(2n+1)^2}{c^2} \right]}. \quad 2.1$$

This characteristics time describes the time required for the concentration of the substrate inside the crystal to reach  $1 - 1/e$  of  $C_{out}$ , which is 69% of the initial concentration

Table 2.2 shows diffusion times for various crystal sizes.

**Table 2.2.** Diffusion times for various crystal sizes

<b>Crystal Size (<math>\mu\text{m}^3</math>)</b>	<b><math>\tau_D</math></b>
400 × 400 × 1600	16 sec – 60sec
300 × 400 × 500	9.5sec
10 × 20 × 30	15ms
3 × 4 × 5	1ms
1 × 2 × 3	150 $\mu\text{s}$
0.5 × 0.5 × 0.5	17 $\mu\text{s}$
0.1 × 0.2 × 0.3	1.5 $\mu\text{s}$

## 2.4 Data Analysis at XFELs

### 2.4.1 SFX – From a data analysis point of view

In SFX, each diffraction pattern is produced from a single X-ray pulse and a single microcrystal. When the microcrystals in their mother liquor are injected into the X-ray interaction region, diffraction patterns are obtained from randomly oriented crystals. No effort is made to target a single crystal.

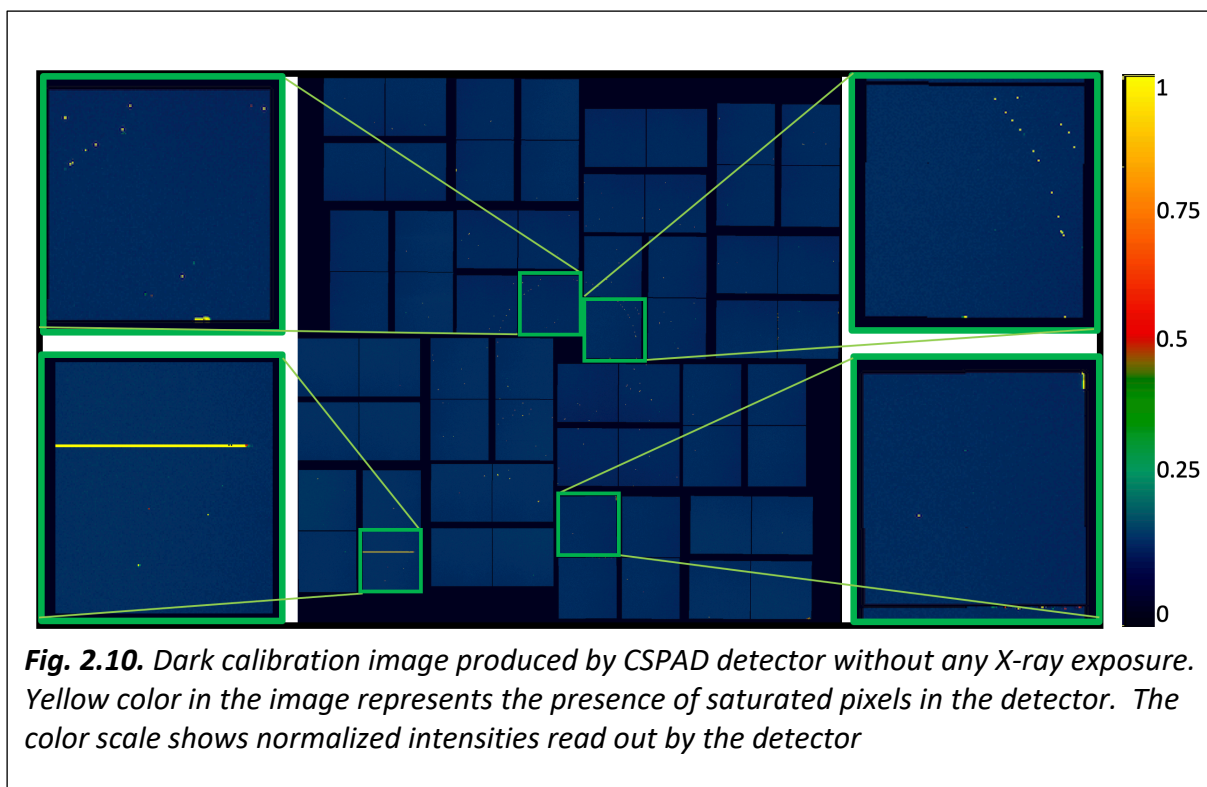
The “hit rate” is the number of crystal hits per second. The hit rate depends on the concentration of crystals in the jet, and the crystal quality. In some cases, crystals do not diffract at all, and the hit rate close to 0%. It is also necessary to avoid very high hit rate. The high hit rate may produce multi-crystal hits and may complicate the data analysis (Hunter et al., 2014). Typically, hit rate less than 50% is best for the SFX.

In S(F)X each reflection has to be measured multiple times to reconstruct the integrated reflection intensity (Kirian et al., 2011). A complete SFX dataset typically has tens of terabytes of data. Each dataset can have several hundred thousand to millions of diffraction patterns, some with and some without Bragg reflections. New approaches are used, and new algorithms had to be developed to specially process and analyze the data collected at the XFELs.

Data analysis at XFELs means extracting meaningful information from all diffraction patterns. It is necessary to follow a particular set of steps and use several suites of programs. Data analysis steps mainly have two parts: Hit finding and Indexing. We use Cheetah (Barty et al., 2014) for hit-finding and CrystFEL (White et al., 2012) for indexing. They will be described in detail in the following chapters.

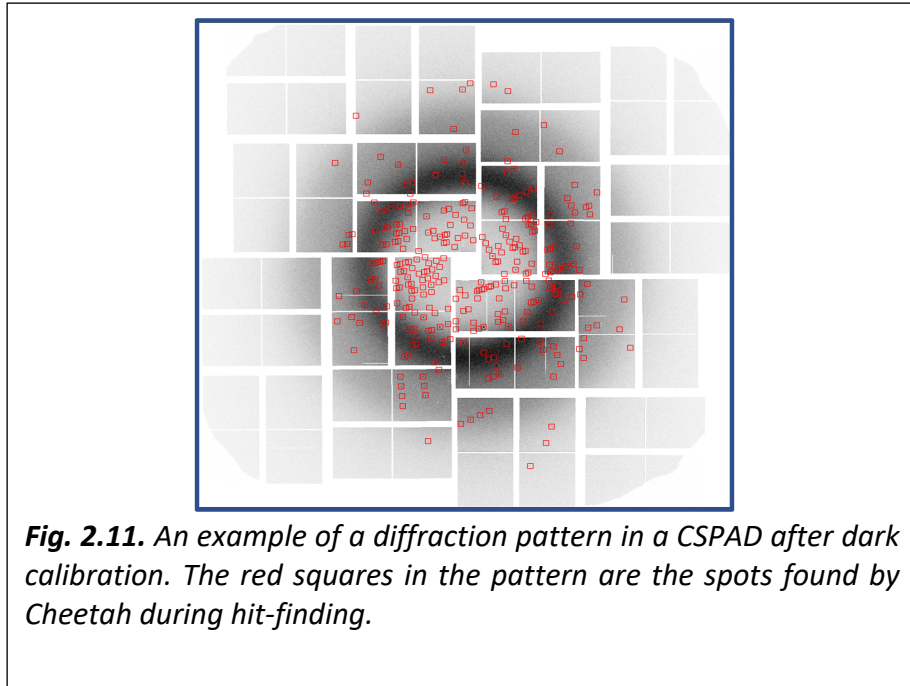
## 2.4.2 Preprocessing of raw data using Cheetah - Hit Finding

Diffraction patterns that contain Bragg reflections are called hits. Cheetah separates hits from the pool of all diffraction patterns. For this, Cheetah first determines the static detector offset. It is estimated from frames collected in the absence of an X-ray pulse (dark). Pixels that constantly give either too high or too low pixel-values due to physical damage of the pixel are flagged and masked out. Figure 2.10 shows electronic noise and bad, hot, and saturated pixels produced by the CSPAD. This electronic noise, bad and saturated pixels are subtracted from each diffraction pattern.



The local background subtraction is then performed to account for the intensity in the diffraction pattern generated due to the liquid surrounding the crystal. To determine the local background value, a pixel box of a certain size is considered around the pixel that contains the





Bragg reflection. The median of the background intensity inside the box of the pixel is estimated and subtracted. The background-corrected data is used for final Bragg peak detection. Based on various parameters such as threshold intensity, signal to noise ratio (SNR) and a certain number of spots in a pattern, etc., hits are separated (Fig. 2.11). Information about hits along with other required experimental information from the original raw file produced by the detector is stored in files following the hierarchical data format, version 5 (HDF5, <https://www.hdfgroup.org/solutions/hdf5/>).

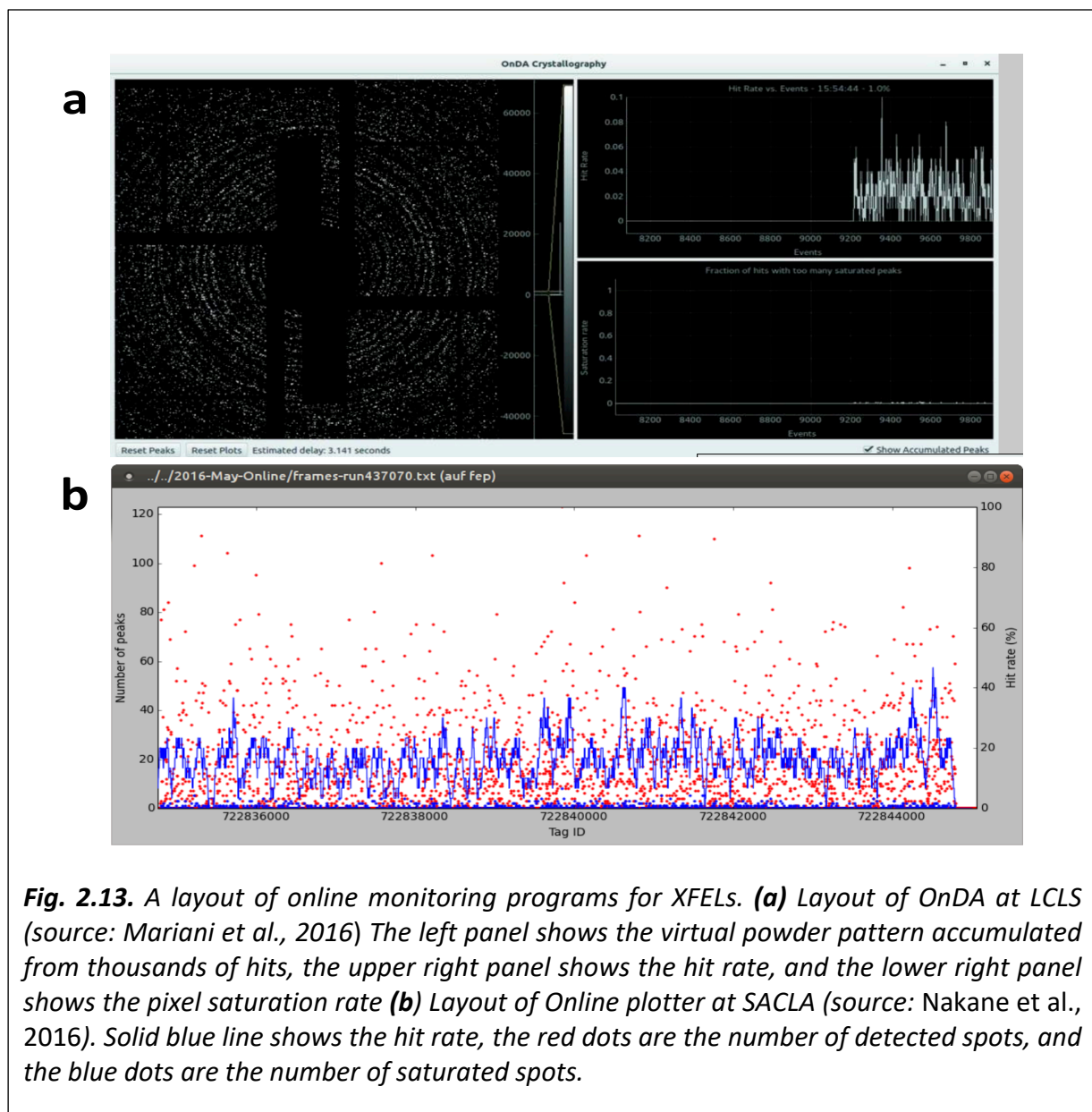
Run	Dataset	XTC	Cheetah	CrystFEL	H5Directory	Nprocessed	Nhits	lindex	Hitrate%
181	darkcal	Ready	Finished	---	r0041-darkcal	2412	0	---	0.00
180	ricemixed_30...	Ready	Finished	---	r0042-asum...	71362	3368	---	4.72
179	ricemixed_30...	Ready	Finished	---	r0043-asum...	69805	5075	---	7.27
178	ricemixed_30...	Ready	Finished	---	r0044-darkcal	3040	0	---	0.00
177	ricemixed_30...	Ready	Finished	---	r0045-darkcal	6602	0	---	0.00
176	ricemixed_30...	Ready	Finished	---	r0046-asum...	44533	40	---	0.09
175	ricemixed_30...	Ready	Finished	---	r0047-asum...	76613	4048	---	5.28
174	ricemixed_30...	Ready	Finished	---	r0048-asum...	15094	1057	---	7.00
173	GV_85mm_V13	Ready	Finished	---	r0049-asum...	83682	5998	---	7.17
172	GV_85mm_V13	Ready	Finished	---	r0050-asum...	76199	5676	---	7.45
171	GV_85mm_V13	Ready	Finished	---	r0051-asum...	71997	5289	---	7.35
170	GV_85mm_V13	Ready	Finished	---	r0052-asum...	100537	7733	---	7.69
169	darkcal	Ready	Finished	---	r0053-asum...	71040	4972	---	7.00
168	unmixed_rice_...	Ready	Finished	---	r0054-asum...	70404	4789	---	6.80
167	unmixed_rice_...	Ready	Finished	---	r0055-asum...	71291	5159	---	7.24
166	unmixed_rice_...	Ready	Finished	---	r0056-asum...	71301	5691	---	7.98
165	unmixed_rice_...	Ready	Finished	---	r0057-asum...	76293	5938	---	7.78

**Fig. 2.12.** A layout Graphical user Interface (GUI) of Cheetah. Dataset, Number of processed images, number of hits and hitrate can be monitored using the GUI (Blue ovals).

Apart from the functions mentioned above, Cheetah also performs the following tasks:

1. It generates virtual powder diffraction patterns of hits and non-hits separately.
2. It calculates statistics on hit rate and resolution.

### 2.4.3 Online monitoring during experiments



**Fig. 2.13.** A layout of online monitoring programs for XFELs. **(a)** Layout of OnDA at LCLS (source: Mariani et al., 2016) The left panel shows the virtual powder pattern accumulated from thousands of hits, the upper right panel shows the hit rate, and the lower right panel shows the pixel saturation rate **(b)** Layout of Online plotter at SACLA (source: Nakane et al., 2016). Solid blue line shows the hit rate, the red dots are the number of detected spots, and the blue dots are the number of saturated spots.

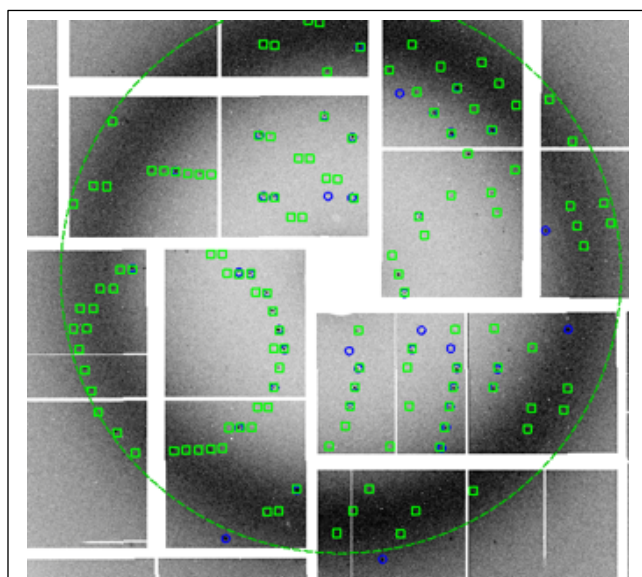
Although Cheetah can also be used during the experiment to provide online feedback, the feedback is not in real-time. There is limited availability of computer nodes for processing, the raw file is large, and the data acquisition (DAQ) system needs time to write the raw file to the required location. Statistics are available only after a run (a run is a period of time where data

are collected continuously). These data first have to be processed. This imposes a latency of a few minutes to even hours depending on the length of the run. A timely decision can't be made during the experiment.

Real-time analysis programs are developed to monitor the raw data and provide the statistics with a delay of only a few seconds (Mariani et al., 2016; Nakane et al., 2016). This helps finding either the particular batch of crystals is diffracting or not. The decision, to use a new crystal batch, can be made without wasting precious experimental time at the XFELs. OnDA (online data analysis) (Mariani et al., 2016) is used for real-time monitoring at the LCLS and the EuXFEL, while a program called 'Online Plotter' (Nakane et al., 2016) is used at SACLA (see Fig. 2.13 for the display layout of both programs).

#### 2.4.4 Indexing of Diffraction Patterns

Once the hits are determined, Miller indices and the intensities of each Bragg spots are determined. For this, the CrystFEL suite of programs (White et al., 2012) has been specially designed. As each pattern in SFX is obtained from new and randomly oriented crystal, CrystFEL considers each pattern to be independent from the others. CrystFEL requires a geometry file that contains the detector metrology such as detector distance and the size and location of each pixel.



**Fig. 2.14.** A diffraction pattern collected by CSPAD showing the peaks found by Cheetah(blue) and indexed by CrystFEL (green)

Indexing is assigning Miller indices to each Bragg spot in a diffraction pattern. The locations of intense Bragg spots are determined by Cheetah for each diffraction pattern. This list of Bragg spot positions is used now by a program called indexamajig. Indexamajig uses several fast Fourier transform (FFT) based indexing algorithms such as MOSFLM (Rossmann and Van Beek, 1999; Kabsch, 2014), DirAx (Duisenberg, 1992), XDS (Kabsch, 2010) and asdf (White et al., 2012), and XGandalf (Gevorkov et al., 2019). After indexing, Indexamajig also extracts the intensity of each (partial) Bragg spot in the diffraction pattern. The Bragg intensities, the orientation of the crystal, and unit cell parameters are output to a text file. This file is usually called a “stream file”. The stream file also has information about the raw data file names, positions of found and predicted peaks, and it also marks the saturated pixels of the diffraction pattern.

#### **2.4.5 Indexing ambiguity**

The indexing ambiguity typically arises when data are collected from crystals with higher symmetry. Unit cell parameters may have the same lengths, but the molecular orientations are different along different unit cell directions. For example, the space group of PYP crystals is  $P6_3$ . Two axes  $a$  and  $b$  have the same length and hence the crystal possesses two-fold indexing ambiguity. As a result, diffraction patterns are indexed as either  $(hkl)$  or  $(khl)$ , yet intensities  $I(hkl)$  and  $I(khl)$  are not related by symmetry and therefore are not identical. If the crystal possesses an indexing ambiguity, an additional software tool of CrystFEL called ambigator is used to resolve the ambiguity. The output file produced by indexamajig is used as an input for the ambigator. Ambigator also produces a stream file as an output.

## 2.4.6 Merging and Assessment of Data Quality

Merging is combining all of the intensities from the stream file to create a 3D reciprocal space intensity map. The intensities obtained from diffraction patterns are from many crystals of various sizes and shapes. CrystFEL uses a simple averaging approach, called Monte Carlo averaging, to merge multiple observations and to estimate the errors of the measurements. As a result, the integrated reflection intensities are determined.

As the data metric called  $R_{\text{merge}}$  used in conventional MX cannot be used for SFX data, CrystFEL uses its own figure of merit to assess the quality of data called  $R_{\text{split}}$ . It is given by

$$R_{\text{split}} = \frac{1}{\sqrt{2}} \frac{\sum |I_{\text{even}} - I_{\text{odd}}|}{0.5 \times \sum (I_{\text{even}} + I_{\text{odd}})},$$

where  $I_{\text{even}}$  and  $I_{\text{odd}}$  are the intensity merged from either even-numbered or odd-numbered patterns.

In addition, these half data can be compared using correlation coefficients, which were introduced by Karplus and Diederichs (Karplus and Diederichs, 2012) for SFX data processing. One of the correlation coefficients is called  $CC^*$  and given as

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + 2CC_{1/2}}},$$

where  $CC_{1/2}$  is a Pearson correlation coefficient between two halves of data and is calculated as:

$$CC_{1/2} = \frac{\sum_{\text{hkl}} (|I_{\text{even}}|^2 - \langle I_{\text{even}} \rangle^2) \times (|I_{\text{odd}}|^2 - \langle I_{\text{odd}} \rangle^2)}{\sqrt{\sum_{\text{hkl}} (|I_{\text{even}}|^2 - \langle I_{\text{even}} \rangle^2) \times (|I_{\text{odd}}|^2 - \langle I_{\text{odd}} \rangle^2)}}.$$

Similarly, a resolution-dependent signal-to-noise ratio (SNR),  $\langle \frac{I}{\sigma(I)} \rangle$ , is also calculated to accurately determine the resolution of the data. Both ccs can be used, however, we estimate the resolution and data quality using the  $R_{\text{split}}$  and the  $CC_{1/2}$ .

### 2.4.7 Analysis of Processed Data

After merging the data, structure factor amplitudes are calculated using a suite of programs from Collaborative Computational Project Number 4 (CCP4) (Winn et al., 2011). If the phases are already known, structure factors and the electron density are calculated using Eqn. 2.10. Otherwise, other methods as mentioned in section 2.2.6 must be used to determine the phase.

In time-resolved experiments, the structure factor amplitude is a function of both space (h, k, and l) and time (t). The observed structure factor amplitudes are usually denoted as  $|\vec{F}_{\text{obs}}^t|$  for a time delay 't' after reaction initiation and  $|\vec{F}_{\text{obs}}^{\text{dark}}|$  for the dark state. These structure factor amplitudes can be used for the calculation of difference electron density maps, omit difference maps, and extrapolated electron density maps.

#### 2.4.7.1 Difference Electron Density (DED) Maps

The goal of any time-resolved experiment is to visualize the structural difference between activated molecules and inactive molecules. Inactive molecules are not participating in the reaction. In pump-probe experiments, the inactive molecules have not absorbed a laser photon and usually belong to the 'dark' state. The activated molecules relax along one or several reaction coordinates (Cornish-bowden, 2013). The molecules visit (or populate) several intermediate states with high occupancy. A TR-SFX experiment attempts to determine the structures of these intermediate states. The structural changes between the states and the inactive state are revealed by difference electron density (DED) maps. A DED map is calculated from the measured amplitudes and the phases of a precise model of the inactive state using the difference Fourier

approximation (Cochran, 1951; Henderson and Moffat, 1971). An “ $F(t)^{\text{obs}} - F_{\text{dark}}^{\text{obs}}$ ” difference map can be calculated only if the crystal structure of the activated state is isomorphous with the inactive state. To calculate DED map, at first, a structural model is refined against observed structure factor amplitudes  $|\vec{F}_{\text{obs}}^{\text{dark}}|$  so that calculated structure factor amplitudes  $|\vec{F}_{\text{cal}}^{\text{dark}}|$  and the phases  $\phi^{\text{dark}}$  can be obtained. The observed structure factor amplitudes of both light and dark i.e.  $|\vec{F}_{\text{obs}}^{\text{t}}|$  and  $|\vec{F}_{\text{obs}}^{\text{dark}}|$  are brought to the absolute scale by scaling them to the calculated structure factor amplitudes  $|\vec{F}_{\text{cal}}^{\text{dark}}|$  (Schmidt et al., 2010). Finally, difference structure factor amplitudes are calculated using

$$|\overline{\Delta F}| = |\vec{F}_{\text{obs}}^{\text{t}}| - |\vec{F}_{\text{obs}}^{\text{dark}}| \quad 2.18$$

A weighting factor  $w$  is applied to  $|\overline{\Delta F}|$  to reduce the effect of outliers (Ursby and Bourgeois, 1997; Ren et al., 2001; Schmidt et al., 2003). It is calculated using

$$w = \frac{1}{1 + \frac{(\text{sigma}(|\overline{\Delta F}|))^2}{\langle (\text{sigma}(|\overline{\Delta F}|))^2 \rangle} + \frac{|\overline{\Delta F}|^2}{\langle |\overline{\Delta F}|^2 \rangle}}, \quad 2.19$$

where  $\text{sigma}(|\overline{\Delta F}|)$  is an error in  $|\overline{\Delta F}|$ .

Finally, the DED map is calculated using the weighted structure factor and the phase from the structure of the inactive states. It is given as

$$\Delta\rho(\text{XYZ}, t) = \frac{1}{V} \sum_{\text{hkl}} w |\overline{\Delta F}| e^{i\phi^{\text{dark}}} e^{-2\pi i(\text{hX} + \text{kY} + \text{lZ})} \quad 2.20$$

The peaks in the DED map have both negative and positive features. To be chemically meaningful, the negative features must be on atoms of the inactive state structure. They mark atoms that move during the reaction. The positive features in DED maps represent the electron densities of atoms during the reaction. To be chemically meaningful positive features cannot be



on top of the atoms of the inactive state structure. With the help of DED maps, it is possible to track even a small atomic displacement in intermediate states with small occupancies such as attachment or removal of a water molecule (Drenth and Mesters, 2007).

It is important to note that although the difference structure factors are on the absolute scale (Eqn. 3.2), a difference map is only on half of the absolute scale due to the difference Fourier approximation (Drenth and Mesters, 2007). This is important when integrating the positive (or negative) difference density features. The result must be multiplied by a factor of 2 to yield the correct number of displaced electrons.

#### **2.4.7.2 Omit Difference Maps**

DED map can only be calculated between two isomorphous datasets. Two datasets are considered isomorphous if their cell parameters have similar values. As a rule of thumb, the differences in the cell parameters should be less than 1/3 of the maximum resolution. For example, if the maximum resolution of a dataset is 2 Å, only the maximum difference of ~0.7 Å (2/3) in the cell parameter is allowed for the calculation of DED map. For non-isomorphous datasets, often omit difference map is determined for structural interpretation.

To determine an omit difference map, all the molecules or atoms of interest in the reference model are removed. Then the reference model is refined using simulated annealing (SA) algorithms (Brnger, 1991; Korostelev, Bertram and Chapman, 2002) against the observed structure factors collected from the activated crystals. The Fo-Fc map (Lamb, Kappock and Silvaggi, 2015) obtained after the refinement is called a SA omit difference map. Here “Fo” corresponds to the observed structure factors of the excited state, “Fc” is determined from the

model used for refinement. The SA omit difference map is essentially free of reference model phase bias.

### 2.4.7.3 Extrapolated Electron Density (EED) Maps

In a TRX experiment, the population transfer from the dark state to the activated states is usually small on the order of only 5-10% (Schmidt et al., 2013). A structure cannot be directly refined against the structure factors of the activated state. As a result, an additional map called extrapolated electron density map (EED) is determined. EED map is used for refining a structure of an intermediate state with the help of the difference structure factors amplitudes and the dark structure. EED structure factors are calculated using the following formula:

$$\vec{F}^{\text{Ext}} = \vec{F}_{\text{cal}}^{\text{dark}} + N\overline{\Delta F} ,$$

where  $\vec{F}_{\text{cal}}^{\text{dark}}$  are the calculated structure factors of the refined reference structure (Terwilliger and Berendzen, 1996), N is the multiplication factor and  $\overline{\Delta F}$  are difference structure factors. The value N is determined when negative electron densities in EED map just vanishes. The value of N is approximately related to the population transfer. The population transfer is given as  $PT = \frac{100}{N} \times 2$ , where the factor of two accounts for the difference map being only on half the absolute scale.

EED maps are used to determine the molecular structures of activated states. For this, a structure is fitted into EED map, and phases of  $\Delta F$  are obtained from the fitted initial model. These phases are combined with observed  $\Delta F$  and phased extrapolated structure factors are determined. The final structure is refined in reciprocal space against the phase extrapolated structure factors.

## 2.5 List of Attended Beamtimes at XFELs and my contributions

During my dissertation, I attended a large number of beamtimes at multiple XFELs worldwide (Table 2.3). To support these beamtimes, I have been involved in protein overexpression, purification, and crystallization. My main focus, though, was on data processing and data analysis.

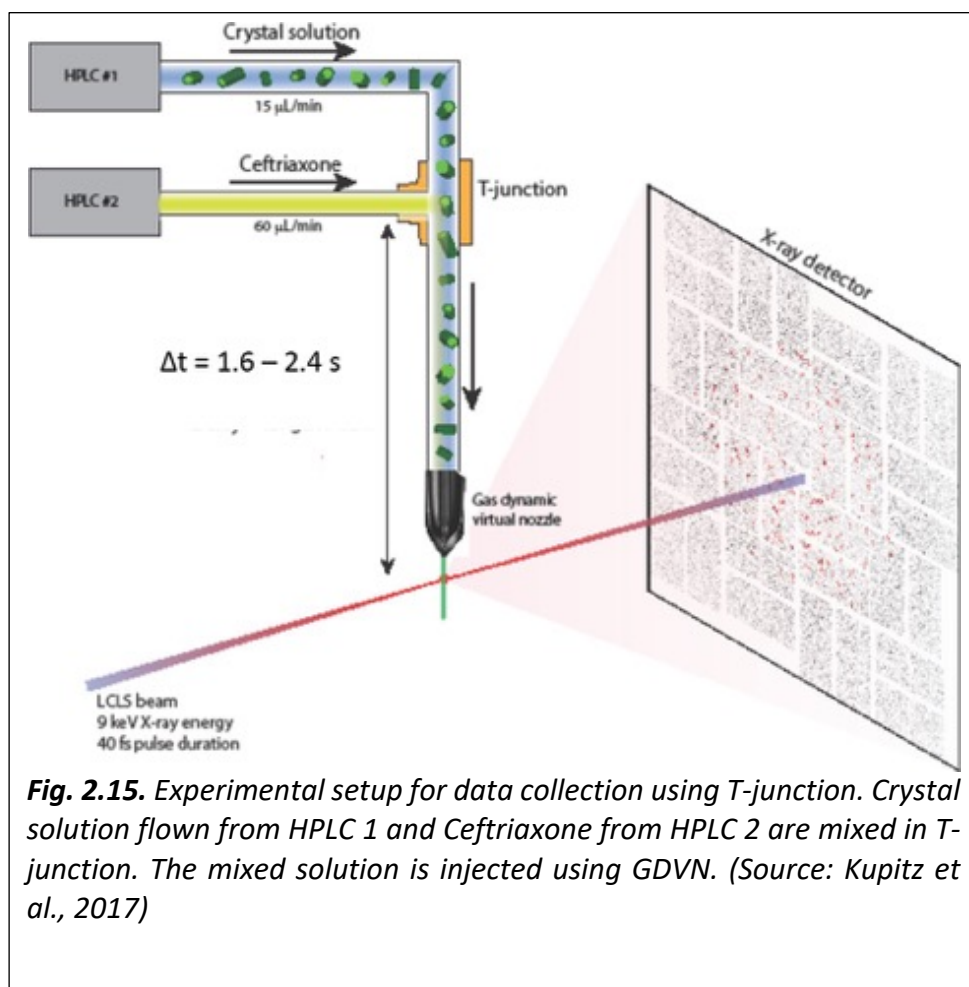
**Table 2.3.** List of beamtimes that I attended and my work in each of them.

Date	Experiment	Descriptions	XFEL	PI	My Work
December 2015	cxik1715	TR-SFX to Establish Structure-Based Enzymology at the LCLS ***	LCLS	M. Schmidt	H, I, R
January 2016	cxik4516	Signal generation in Phytochromes studied by picosecond time resolved SFX	LCLS	S. Westenhoff	H
August 2016	cxio3316	Structure determination of the full-length adenine riboswitch in the ligand-free state	LCLS	Y.-X. Wang	H
September 2016	cxin5016	TR-SFX to Establish Structure-Based Enzymology at the LCLS ***	LCLS	M. Schmidt	H, I, G, D, R
December 2016	2016B8005	TR-SFX experiments on the Z/E isomerization in bacterial Phytochromes	SACLA, Japan	M. Schmidt	H, I, G
August 2017	cxilp7315	Structure-Based Enzymology at the LCLS	LCLS	M. Schmidt	H, I, G, D, R
November 2017	cxilr3016	Z- to E-isomerization in Phytochromes	LCLS	M. Schmidt	H, I, G
March 2018	mfxls0816	The dynamics of light-driven reactions in Photosynthesis studied by fixed target TR-SFX and XES spectroscopy	LCLS	P. Fromme	XES
March 2018	2012	Serial Femtosecond MHz crystallography	EuXFEL, Germany	A. Barty	H, I, R
October 2018	2018B8008	Signal generation in Bacteriophytochromes studied by Serial Femtosecond Crystallography	SACLA, Japan	S. Westenhoff	H, I, G, D, R
February 2019	2173	Time-resolved Crystallography of Photosynthetic	EuXFEL, Germany	P. Fromme	H, I, G
March 2019	2166	Time-resolved serial Crystallography ***	EuXFEL, Germany	M. Schmidt	H, I, G, D, R
April 2019	2019A8007	Signal transduction in Bacteriophytochromes observed by Time-Resolved Serial Femtosecond Crystallography	SACLA, Japan	M. Schmidt	H, I, G, R
May 2019	2019A8063	Signal generation in Bacteriophytochromes studied by Serial Femtosecond Crystallography***	SACLA, Japan	S. Westenhoff	H, I, G, D, R
May 2019	2019A8007	Signal transduction in Bacteriophytochromes observed by Time-Resolved Serial Femtosecond Crystallography	SACLA, Japan	M. Schmidt	H, I, G, D, R
March 2020	2450	Time-Resolved Enzymology to Capture a Full Catalytic Series	EuXFEL, Germany	M. Schmidt	H, I, G, D, R

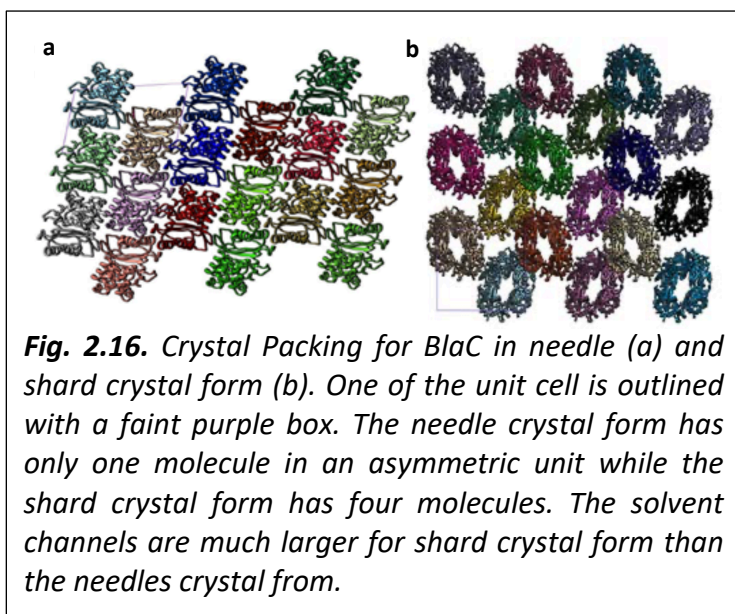
August 2020	mfxl01171	Room Temperature Structure and Inhibition of the Coronavirus SARS CoV-2 Main Protease	LCLS	M. Schmidt	H, I, G, D, R
October 2020	mfxl06818	Pushing the Time-Resolution of Mix-and-Inject Serial Crystallography to the Limit	LCLS	M. Schmidt	H, I, G

**Note:** H – Hit finding, I – Indexing, G -Geometry Refinement, D – Difference Map Calculation, R – Structure Refinement, \*\*\* Focus of this dissertation

## 2.6 Mix-and-Inject Serial Crystallography (MISC) at the LCLS

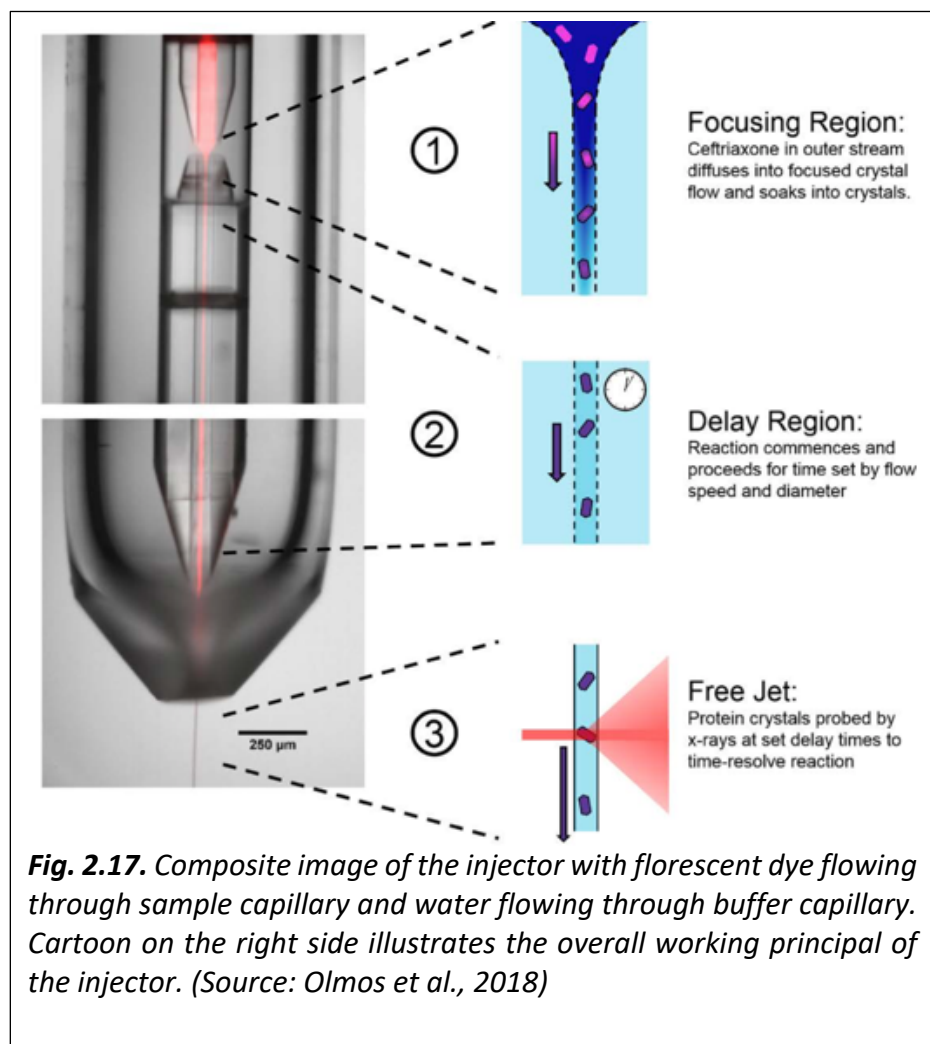


MISC (Section 2.3.4) was performed at the Coherent X-ray Imaging (CXI) instrument (Liang et al., 2015) of the LCLS in December 2015. The purpose of the experiment was to establish MISC at the XFEL and study the reaction between BlaC and ceftriaxone (Section 1.3.1) (Kupitz et al., 2017). In the experiment, BlaC crystals were mixed with substrate using a T-junction (Fig. 2.15). The T-junction was connected to a ceftriaxone solution, to protein crystals and a GDVN (Fig. 2.15). After mixing in the T-junction, the solution was injected into the X-ray interaction region after a delay of 2 seconds and probed at 120 Hz with 40 fs XFEL pulses. Diffraction patterns were collected at 120 Hz on the CSPAD (Hart et al., 2012; Herrmann et al., 2013; Blaj et al., 2015).



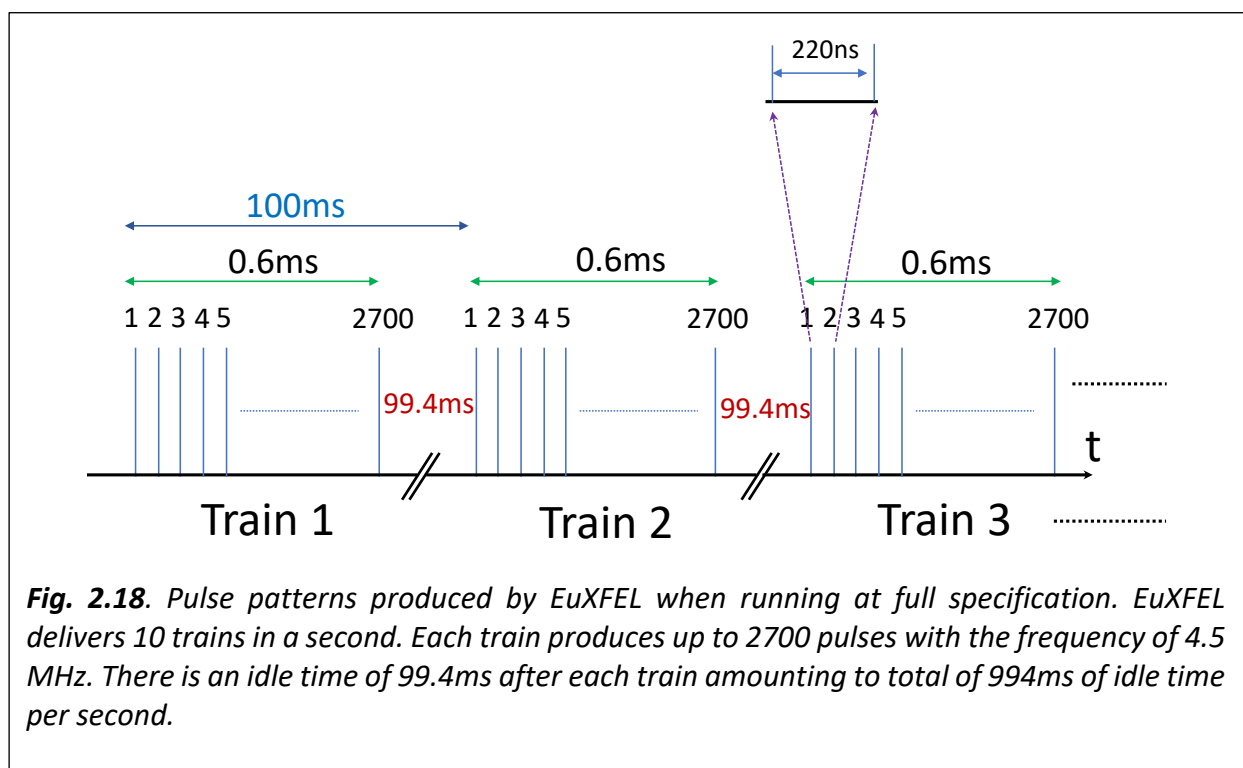
Similarly, the second MISC experiment was performed in the LCLS in September 2016 on the same substrate and enzyme. The purpose of the experiment was to determine how ceftriaxone binds to BlaC and study how BlaC modifies the antibiotic (Section 1.3.1). The injector used for the experiment was designed by the Pollack group from Cornell University (Calvey et al., 2016). The injector enables the collection of data with millisecond time resolution (Fig. 2.17. ).

Two types of crystals with either shard or needle forms were produced and used for the experiment (Fig. 2.16). Although both crystal forms are monoclinic, they have different crystal packing morphologies and hence different unit cell parameters. Datasets for the unmixed (inactive) structure and timepoints of 30ms, 100ms, 500ms, and 2 seconds were collected during the beamtime for both shard and needle crystal forms separately.



## 2.7 Pump-probe TR-SFX at the EuXFEL

EuXFEL is a newly commissioned XFEL that came online in 2017. It delivers X-ray pulses in trains at the rate of 10 Hz, with each train capable of producing 2700 X-ray pulses at 4.5 MHz (Fig. 2.18). This unique pulse pattern and high repetition rates generate several challenges for sample injection, data collection, data processing, and time-resolved experiments in general. A recent successfully accomplished SFX experiment demonstrated that SFX is feasible at the EUXFEL (Wiedorn et al., 2018). We were invited to pioneer TR-SFX experiments at the EuXFEL. A TR-SFX experiment was performed in March 2019 at the Single Particle, Clusters and Biomolecules & Serial Femtosecond Crystallography (SPB/SFX) instrument at the EuXFEL. The main purpose of

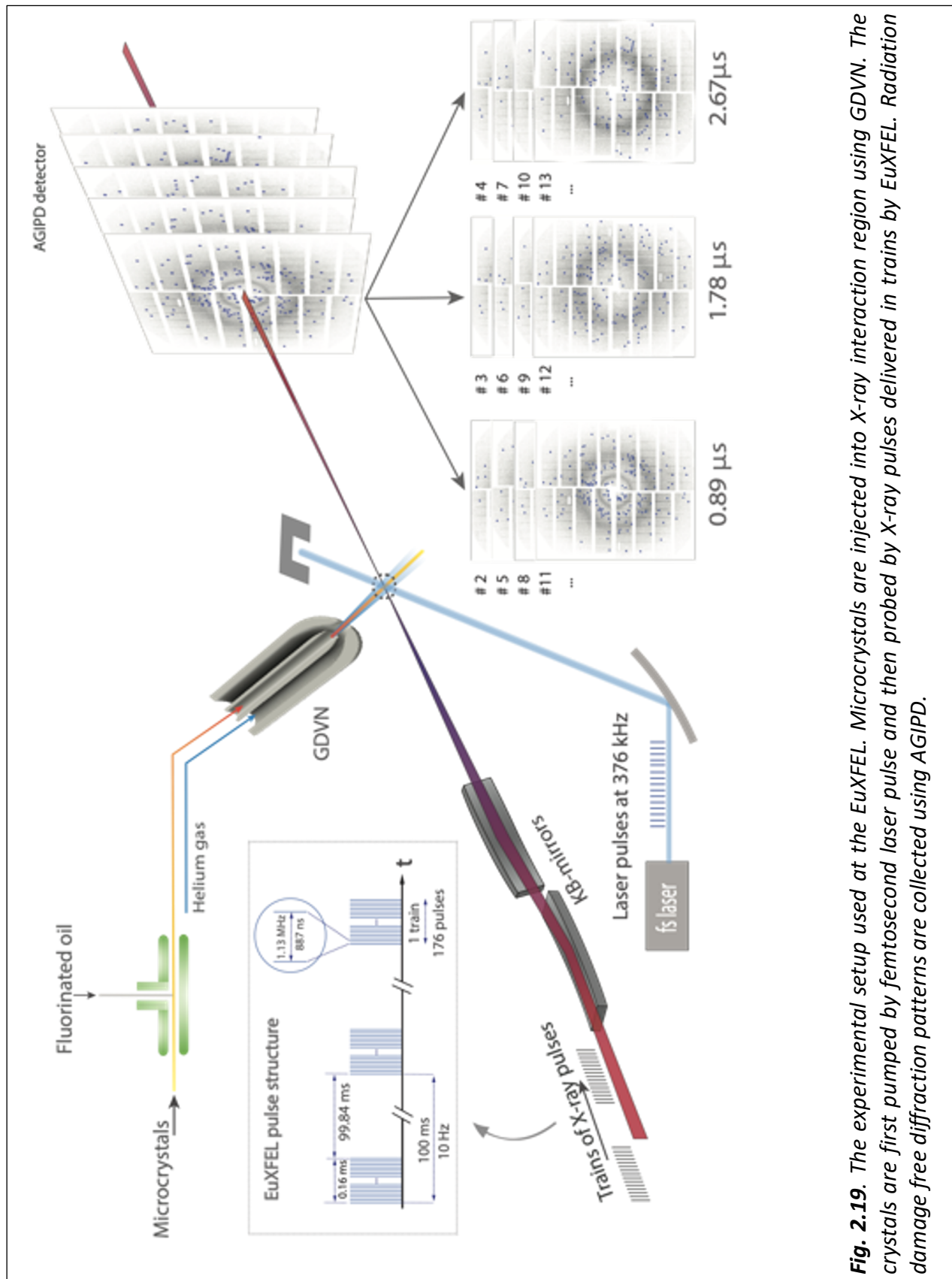


the experiment was to establish time-resolved crystallography at EuXFEL and collect time-resolved diffraction patterns at near-atomic resolution and with picosecond time resolution.

The experiment was performed to explore the picosecond regime in the photocycle of Photoactive Yellow Protein (PYP) (Fig. 1.7) after the trans-cis isomerization of the pCA chromophore. Although the structures of all the populated intermediates in the photocycle of PYP are known, the TR-SFX data on the picosecond time regime is sparse. Datasets for three different time delays i.e. 10ps, 30ps, and 100ps after photoactivation were collected and studied. These time-points were selected in such a way that they complement the data collected at a 3ps time delay at LCLS (Pande et al., 2016) and the 100ps time delay collected at Argonne National Laboratory (Jung et al., 2013).

Fig. 2.19 shows the experimental setup at the EuXFEL. PYP crystals used for the experiments were grown as described above. PYP microcrystals were injected into the X-ray interaction region with a GDVN. The jet produced by GDVN was measured to have a diameter of about 5  $\mu\text{m}$  and a flow rate of 35  $\mu\text{L}/\text{min}$ . This translates to a jet speed of approximately 30 m/s. The injected microcrystals were activated using laser pulses from a high-power femtosecond laser operating at MHz rates. A 0.5 mm thick beta barium borate (BBO) crystal was used to generate a 420 nm second harmonic from the optical laser radiation with wavelength 840 nm and 15 fs pulse duration. The pulse duration was stretched from 15 fs to 250 fs using fused silica windows and lenses. The beam size was measured to be 42  $\mu\text{m}$  Full-Width Half Maximum (FWHM) and the average laser pulse energy was about 2.3  $\mu\text{J}$ . Therefore, the laser fluence at the sample position was 1.6  $\text{mJ}/\text{mm}^2$ . The laser was synchronized to the first X-ray pulse in each train setting a certain time delay and the same timing continued for every train thereafter.





**Fig. 2.19.** The experimental setup used at the EuXFEL. Microcrystals are injected into X-ray interaction region using GdVN. The crystals are first pumped by femtosecond laser pulse and then probed by X-ray pulses delivered in trains by EuXFEL. Radiation damage free diffraction patterns are collected using AGIPD.

The activated crystals were probed using X-ray pulses delivered in the XFEL trains. The diffraction patterns are collected using an Adaptive Gain Integrating Pixel Detector (AGIPD) positioned 117.7 – 118.6 mm downstream of the sample injection region (Henrich et al., 2011; Greiffenberg, 2012).

## 2.8 Pump-probe TR-SFX at the SACLA

Two pump-probe TR-SFX experiments were carried out at SACLA, Japan in October 2018 and May 2019 in collaboration with the Westenhoff group from the University of Gothenburg, Sweden. As the Westenhoff group produced and crystallized the DrBphP CBD construct, we were involved in data processing during, and data analysis after, the experiment. The goal of the experiments was to study the photoreaction of the DrBphP CBD on the picosecond time range (Section 1.3.3).

DrBphP CBD microcrystals were mixed with nuclear grade grease (Sugahara et al., 2017) and extruded into the air using a viscous matrix injector at ambient room temperature. The microcrystals in the stream are activated using femtosecond laser pulses and then probed by X-ray pulses. The diffraction patterns are collected using a multi-port charge-coupled device (MPCCD) detector at a frame rate of 30Hz (Hatsui, 2014; Kameshima et al., 2014). Two different time-points, 1ps, and 10ps were collected during the experiment.

**Table 2.4.** Time-points collected in this dissertation for each protein.

Proteins	Timepoints				
	BlaC	Unmixed	30 ms	100 ms	500 ms
PYP	Dark	10 ps	30 ps	80 ps	Multiple $\mu$ s time-points
DrBphP-CBD	Dark	1 ps	10 ps		

## 3. Results

### 3.1 MISC at the LCLS

In the first MISC experiment, BlaC (Section 1.3.1) crystals with shard form (Section 2.6) were used. These crystals had an average width of 3-10  $\mu\text{m}$  in two dimensions and the thickness of 2-3  $\mu\text{m}$  in the third. The crystals had an approximate concentration of  $5 \times 10^{10}$  crystallites/ml. 300 mmol/L CEF (Section 1.3.1) solution was mixed with 2.0 mol/L ammonium phosphate as a stabilization buffer. Crystals were mixed with substrate and injected into the X-ray interaction region after a time delay of around 2 s.

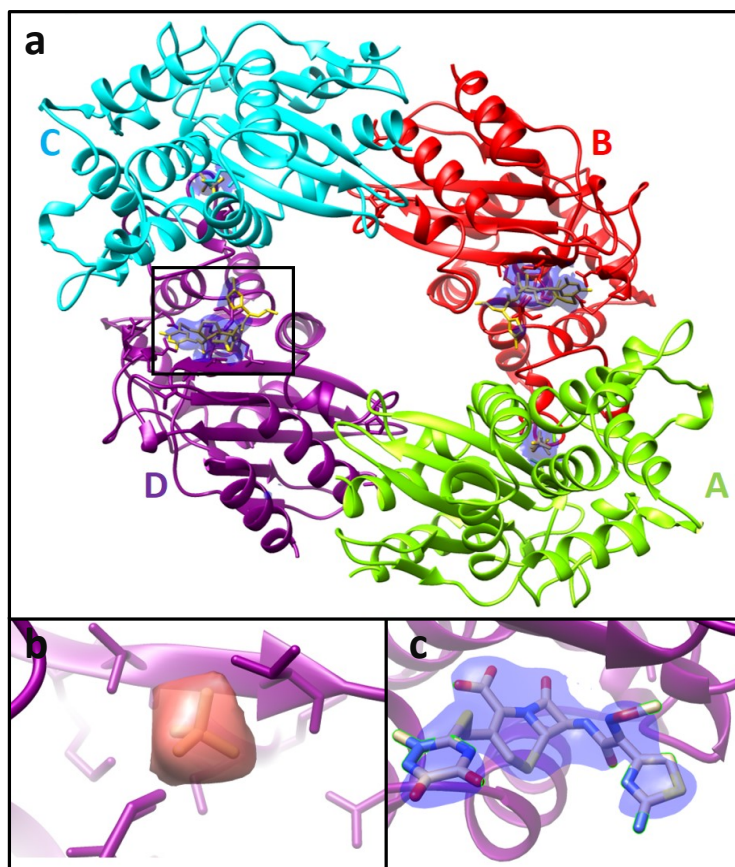
Data for mixed and unmixed were collected separately. 20,236 and 38,913 hits were determined using Cheetah (Section 2.4.2) for unmixed and mixed data respectively. The hit rate was around 1.7%. 12,853 and 22,646 images were indexed, integrated, and merged using the CrystFEL (Section 2.4.4) program suites. The further processing of data was performed using CCP4 programs. A search model was obtained from protein data bank with access code 2GDN for molecular replacement. The BlaC structure determined from the shard crystal form has 4 subunits (Fig. 2.16) in an asymmetric unit and the search model has only one subunit. So, 3 more subunits were added. REFMAC was used for model refinement. The structures were refined up to 2.8  $\text{\AA}$  and 2.4  $\text{\AA}$  resolutions for unmixed and mixed data respectively.

An SA omit difference map (Section 2.4.7.2) was determined to model the CEF into the catalytic cleft of BlaC. COOT was used to visually model the CEF. The modeled structure was then refined against the observed structure factor amplitude of mixed data. Table 3.1 shows the statistics of the data collected during the experiment.

**Table 3.1** Data collection and refinement statistics for the unmixed and mixed structures. Statistics for the highest resolution shell are shown inside the brackets. (Table modified from Kupitz et al., 2017)

	Unmixed	Mixed
Hits	20236	38913
Indexed Images	12853	22646
Resolution	2.8 Å	2.4 Å
Space Group	P1 21 1	P1 21 1
Unit Cell Dimensions (a, b, c, $\alpha$ , $\beta$ , $\gamma$ )	78.26 Å, 96.51 Å, 109.83 Å, 90°, 108.86°, 90°	77.94 Å, 95.32 Å, 111.81 Å, 90°, 109.74°, 90°
Completeness [%]	100 (99.2)	100 (100)
Multiplicity	308 (173)	343 (212)
R <sub>split</sub> [%]	16.7 (96.5)	11.9 (95.0)
CC <sub>1/2</sub> [%]	98.6 (46.1)	98.4 (37.7)
No. of reflections	36988	60598
SNR	4.36	5.32
R <sub>work</sub> / R <sub>free</sub> [%]	22.5/ 24.5	22.5/28.2
Average B value [Å <sup>2</sup> ]	73.9	73.3
Ligands	0	2
RMSD bond lengths(Å)	0.01 Å	0.01 Å
RMSD bond angles (Deg)	1.6	1.7

In a fully refined unmixed structure, phosphate molecules are observed in the active site for all 4 subunits. The SA omit difference map from mixed data shows the presence of phosphate molecules only at 2 subunits whereas the binding of CEF is observed in the other two subunits B and D (Fig. 3.1a). This shows that crystal contacts and crystal packing play significant roles during the binding of the substrate. The presence of clear electron density in the binding pockets and changes in electron density from unmixed (Fig. 3.1b) to mixed states (Fig. 3.1c) proves that MISC is indeed possible at the XFEL.



**Fig. 3.1.** (2F<sub>c</sub>-F<sub>c</sub>) electron density map at  $\sigma=1.1$  in the catalytic cleft of BlaC. **(a)** An overview of an asymmetric unit of the refined model after mixing is shown. Subunits A and C contain phosphate while B and D have a bound CEF. **(b)** Enlarged section of an active site of subunit D for the unmixed structure is shown. Red electron density shows the presence of phosphate in the active site **(c)** Enlarged section of an active site of subunit D for the mixed structure modeled with CEF (Blue electron density) is shown. (Source: Kupitz et al., 2017)

### 3.2 Enzyme intermediates captured using MISC at the LCLS

For shard crystal form, 380,993 hits were collected for 4 different time-points with an average hit rate of 7.63%. Similarly, 529,585 hits were collected for 5 different time-points with the average hitrate of 10.63% for needle crystal form (Section 2.6). Data sets for 9 time-points were collected during the second MISC beamtime conducted at the LCLS. In addition to these time-points, the 2 s time delay for the shard crystal form in the first MISC experiment (Section 3.1) was revisited. Altogether data sets for 10 timepoints – unmixed, 30 ms, 100 ms, 500 ms, and 2 s for both crystal form - were collected. Tables 3.2 and 3.3 show the statistics for the data collected for shard and needle crystal forms respectively.

SA-omit maps were calculated by using two different strategies for two forms of crystal.

They are:

- i) SA-omit map from observed structure factors for shard crystal form

A reference model obtained from the first MISC experiment (Section 3.1) was used for the initial refinement. Since difference electron density (DED) maps couldn't be calculated due to non-isomorphism between reference and the time-points data, SA-omit maps were determined (Section 2.4.7.2). For this, all the waters and the phosphate residing in the active sites of all subunits were removed. In addition, serine (Ser-70) was replaced by glycine (Gly-70) in subunits B and D and the refinements were carried out for all time delays.

**Table 3.2** Data collection and refinement statistics for the shard crystal form. Statistics for highest resolution shell are shown in brackets. (Table modified from Olmos et al., 2018)

	Reference	30 ms	100 ms	500 ms	2s
Hits	98895	35065	88413	158620	39140
Indexed Images	73170	24397	79328	134583	32201
Resolution	2.45	2.75	2.15	2.20	2.30
Space group	P21	P21	P21	P21	P21
Unit Cell [ $\text{\AA}$ , $^\circ$ ] (a, b, c and $\beta$ )	79.0 97.2 110.6 108.7	78.7 96.8 112.6 109.7	79.2 96.5 113.7 109.9	78.8 96.3 113.5 110.0	78.2 95.6 112.3 109.9
Volume [ $\text{\AA}^3$ ]	804,442	807,597	817,098	809,346	789,415
BlaC/unit cell	8	8	8	8	8
Completeness	100(100)	100(100)	100(100)	100(100)	100(100)
Multiplicity	1220.89(103.3)	526.16(142.0)	894.63(58.8)	1363.04(81.3)	329.82(59.0)
SNR	8.9(2.4)	6.4(0.9)	7.1(1.0)	8.3(0.9)	5.4(1.1)
R <sub>split</sub> [%]	9.8(209.4)	14.2(121.1)	11.18(111.0)	9.7(126.3)	11.9(104.1)
CC <sub>1/2</sub> [%]	99.4(41.1)	98.6(34.5)	99.4(37.5)	99.6(31.0)	96.8(35.4)
Refinement					
R <sub>cryst</sub> /R <sub>free</sub> [%]	19.2/24.4	19.3/25.0	20.9/23.9	21.9/25.0	23.5/26.6
* <sup>B</sup> CEF/E-CFO* <sup>a</sup>	0/0	91/23 <sup>b</sup>	57/32	40/36*	58/25
* <sup>D</sup> CEF/E-CFO* <sup>a</sup>	0/0	89/24 <sup>b</sup>	54/40	38/44	51/31
Stacking	no	yes	yes	yes	yes
H <sub>2</sub> O	315	143	499	431	399
Average B value [ $\text{\AA}^2$ ]	48.2	51.7	42.3	37.3	36.2
Ligands	0	2+2(stacking)	2+2(stacking)	2+2(stacking)	2+2(stacking)
RMSD bond lengths( $\text{\AA}$ )	0.008	0.010	0.008	0.008	0.008
RMSD bond angles (Deg)	1.10	1.72	1.66	1.67	1.74
PO4	4	2	2	2	2

<sup>a</sup> occupancy of full length, intact CEF to covalently bound, open E-CFO\*, which has lost R, after refinement, numbers are rough estimates.

<sup>b</sup> if full occupancy with CEF and E-CFO\* is not reached, the protein structure represents a mixture of two structures, one with ligand bound, and one with no ligand bound

\*<sup>B</sup> for subunit B

\*<sup>D</sup> for subunit D

\* addition of OH instead of the double bond  $\Delta$



**Table 3.3** Data collection and Refinement statistics for the needle crystal form. Statistics for the highest resolution shell are shown in brackets. (Table modified from Olmos et al. 2018)

	Reference	30 ms	100 ms	500 ms	2s
Hits	171314	64507	115223	141935	36606
Indexed Images	111466	34590	87580	87058	23278
Resolution	1.8	1.9	1.8	1.9	2.05
Space group	P21	P21	P21	P21	P21
Unit Cell [ $\text{\AA}$ , $^\circ$ ] (a, b, c and $\beta$ )	39.6 41.6 69.3 104.8	39.5 41.6 69.3 104.8	39.6 41.6 69.3 104.9	39.6 41.7 69.5 104.9	39.6 41.7 69.5 104.9
Volume [ $\text{\AA}^3$ ]	110,375	110,096	110,323	110,908	110,908
Completeness	100(100)	100(100)	100(100)	100(100)	100(100)
Multiplicity	985.3(54.5)	329.8(26.8)	831.3(89.0)	806.0(36.5)	238.0(27.3)
SNR	9.6(1.2)	5.8(0.8)	9.6(1.6)	8.6(0.9)	5.1(1.1)
$R_{\text{split}}$ [%]	6.6(97.0)	12.2(136.3)	6.6(72.5)	8.8(129.1)	14.0(105.9)
$CC_{1/2}$ [%]	99.7(39.1)	99.4(40.4)	99.7(55.1)	99.7(30.2)	99.13(38.8)
Refinement					
$R_{\text{cryst}}/R_{\text{free}}$ [%]	21.5/24.5	20.7/26.2	23.0/26.7	21.7/26.4	20.0/25.0
$N^c$	- na -	9	9	6	5
CEF/E-CFO <sup>*a</sup>	0/0	59/0	51/35	43/53	71/0
Stacking	no	no	no	no	no
H <sub>2</sub> O	167	203	154	104	175
Average B value [ $\text{\AA}^2$ ]	34.7	16.9	10.5	15.7	18.3
Ligands	0	1	1	1	1
RMSD bond lengths ( $\text{\AA}$ )	0.008	0.007	0.007	0.003	0.008
RMSD bond angles (Deg)	1.06	1.57	1.74	1.49	1.57

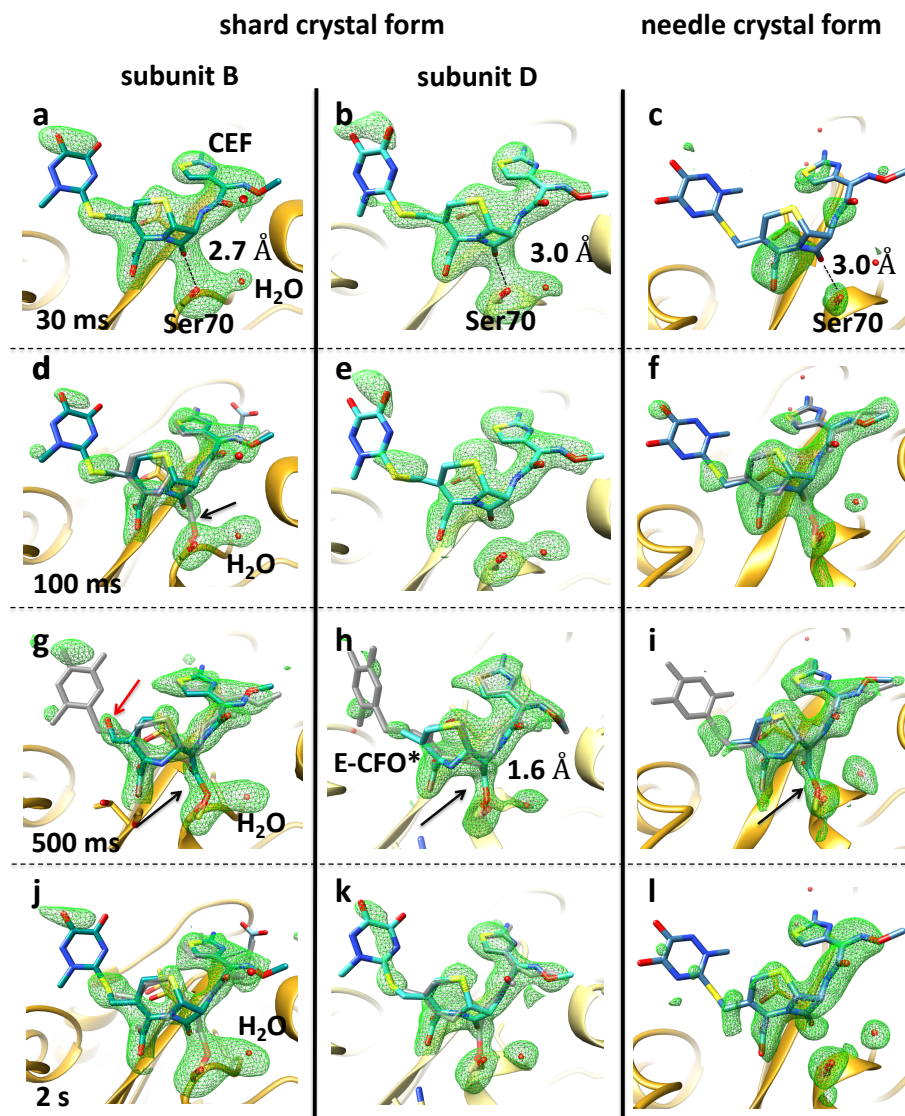
<sup>c</sup>If N does not extrapolate to 100 % occupancy, a fraction of reference structure is present. This is ignored in the refinement.

na: not applicable

ii) SA-omit maps from extrapolated structure factors for the needle crystal form:

Subunit D from the reference model obtained from the earlier MISC experiment (Section 3.1) was extracted for the initial refinement of the reference model of the needle crystal form. Isomorphous DED maps were calculated for all time-points by using a method described in section 2.4.7.1. For modeling the substrate ligand into the BlaC structure, conventional electron density maps were calculated by extrapolating the ligand occupancy to 1. The extrapolated structure factors for all the time-points were calculated using a method mentioned in section 2.4.7.2. Finally, SA-omit maps were refined from extrapolated structure factors using the same method as used for the shard crystal form (Section 3.2(i)).

Figure 3.2 shows the details of these maps in the active site of BlaC. Strong electron density is observed in the active site for 30ms after mixing with the substrate. It proves that the diffusion into the crystals was successful. At this time delay, the formation of a non-covalently bound enzyme-substrate (ES) complex is seen (Fig. 3.2 a, b, c and Fig. 1.6). The occupancy of CEF is full at 30ms. At 100 ms, the ES complex still prevails as the major component ( ~70% occupancy, see Table 3.3). An  $\beta$ -lactam ring opens and covalently binds to SER-70. This intermediate can be seen as a minor fraction ( ~30% occupancy) ) (Fig. 3.2 d, e, f). At 500 ms, high occupancy of an intermediate that is covalently bound to the BlaC called E-CFO\* is observed. Splitting of the leaving group (R) is observed at the same time delay (Fig. 3.2 g, h, i). Finally, at 2 s, binding sites are once again occupied mainly by the full-length CEF with a minor contribution of E-CFO\*.



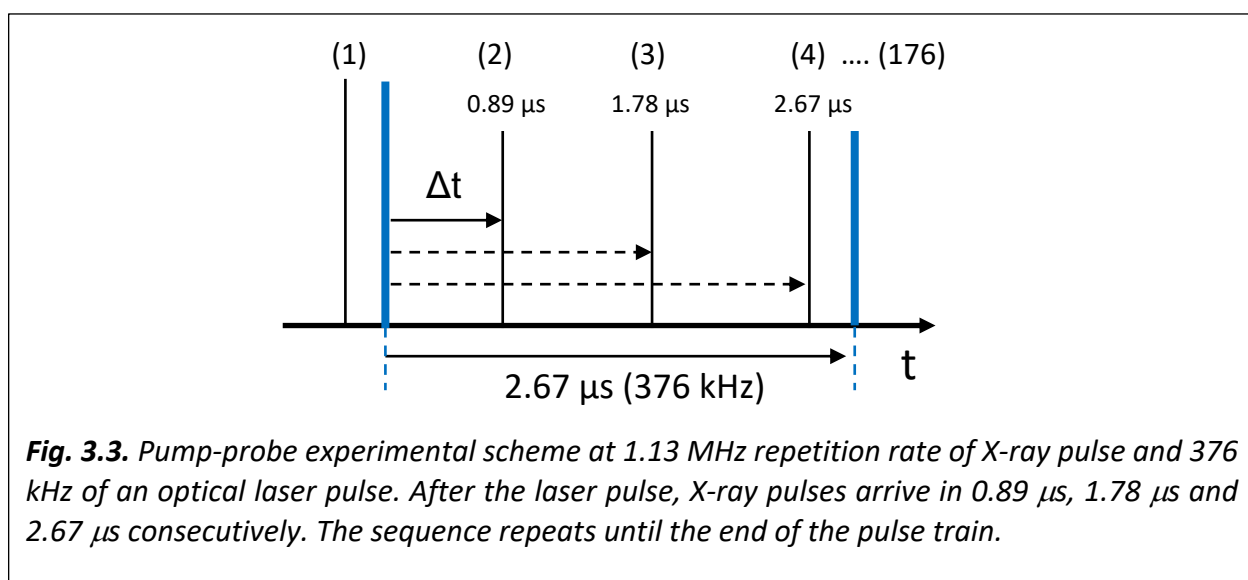
**Fig. 3.2.** Ceftriaxone density (contoured at  $2.5 \sigma$ ) in the active site in shard and needle crystal forms at various times delays. The main species are displayed in blue, the minor species in gray. First two columns: shard crystal form, SA-omit maps (green). Third column: needle crystal form, SA omit maps calculated using extrapolated structure factors. Time delays are arranged from the top (30 ms) to bottom (2 s). (**a, b, c**) The formation of ES complex at 30 ms is observed. The full-length CEF model (blue) is displayed. (**d, e, f**) Early phases of the formation of a covalently bound CEF adduct at 100 ms. The full-length CEF model (blue) is displayed together with the minor E-CFO\* species (gray). E-CFO\* is not observed yet in the subunit D of shard crystal form (panel e). (**g, h, i**) Formation of E-CFO\* (blue) at 500 ms with small contamination of full-length CEF (gray). The red arrow points to electron density that may favor the interpretation by an OH group. (**j, k, l**) A mixture of the non-covalently bound, full-length CEF, and E-CFO\* in the shard crystal form (both subunits) at 2 s. The electron density in the needle crystal form favors only the full-length CEF species. (Figures from Olmos et al., 2018)

### 3.3 Pump-probe TR-SFX at the EuXFEL

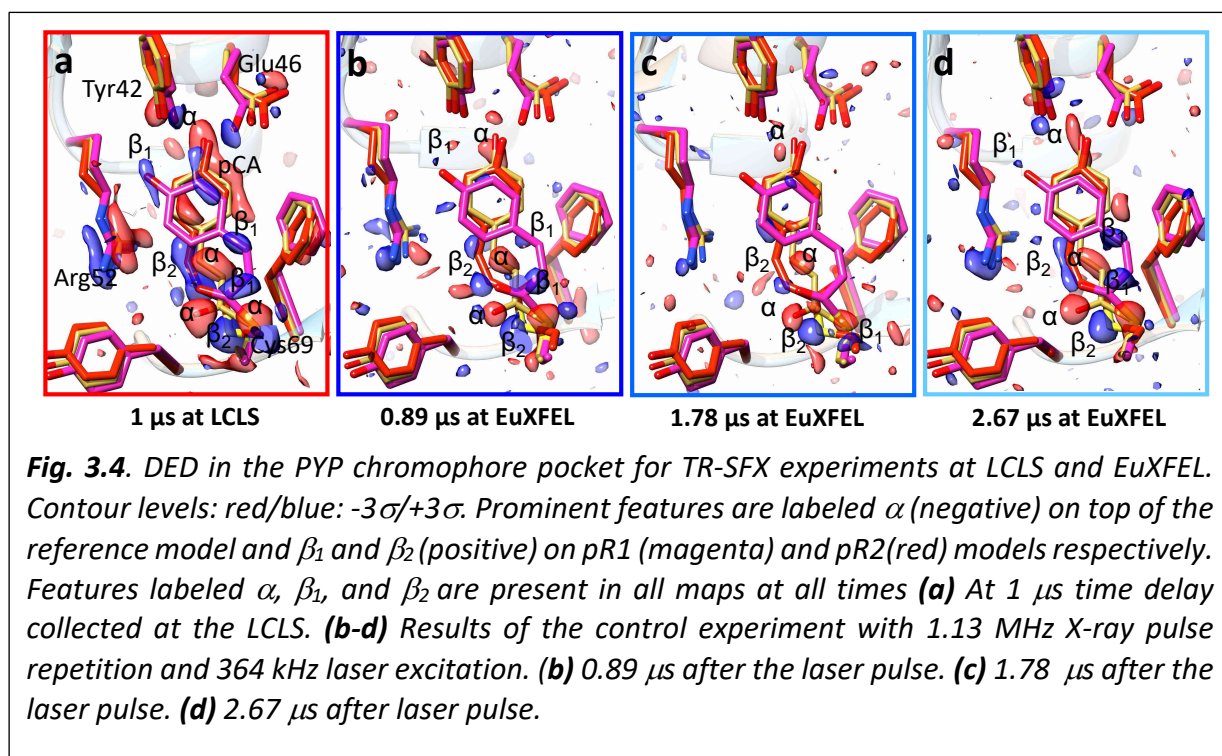
Pump-probe TR-SFX experiment performed on PYP at the EuXFEL generated a large amount of data. A total of 1,165,616 hits were identified by Cheetah (Section 2.4.2) with the average hit rate of 2.53 %. Out of these, 583,730 were indexable. Since the PYP crystal has an indexing ambiguity, the ambiguity was solved by using the CrystFEL program called ambigator (Section 2.4.5). 16 time-points were collected during four 12 hours shift of beam time. DED maps (Section 2.4.7.1) were calculated for interpretation of these time-points.

The EuXFEL during our beam time was delivering 176 pulses per train with a frequency of 1.13 MHz (Section 2.7 and Fig. 2.10). This amounts to 1760 pulses in a second. At first, a few minutes of SFX data without any activation of PYP crystals were collected (Pure dark, Table 3.4). Then, two types of pump-probe schemes were used to collect TR-SFX data.

#### 3.3.1 Control Experiment at 1.13 MHz



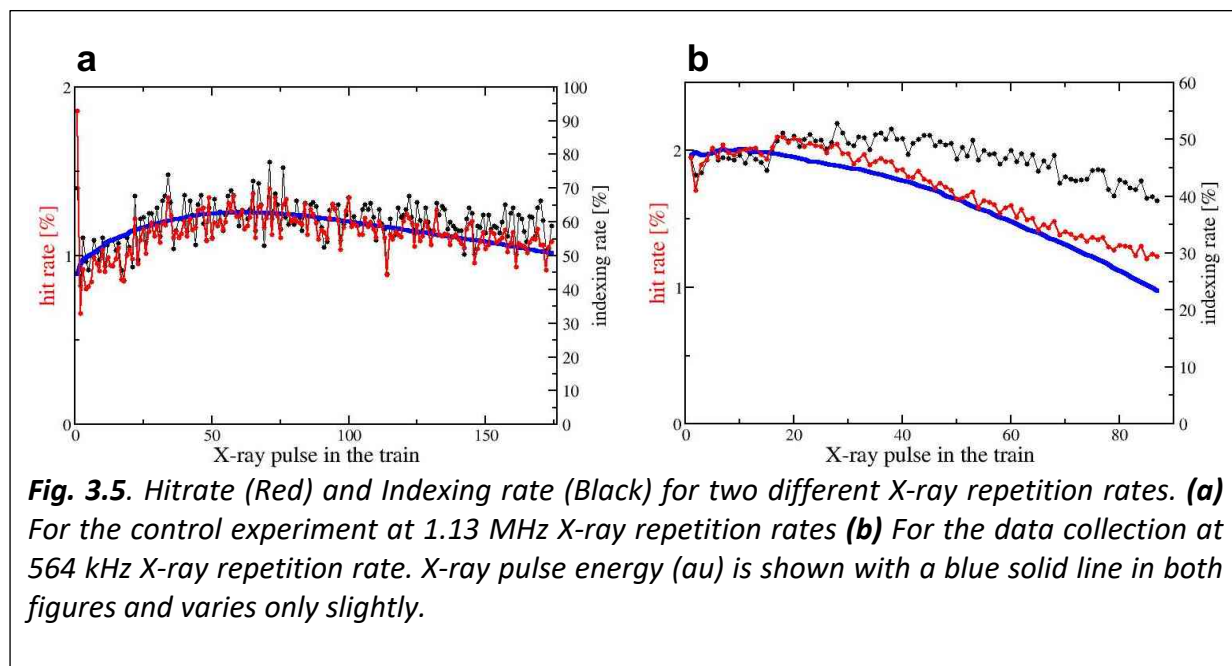
The timing scheme for the first pump-probe experiment is shown in Fig. 3.3. In this scheme, we exposed crystals to the optical laser at 376 kHz. The optical laser was synchronized to the first X-ray pulse. The optical laser initiated the reaction at 2.3 ns after the pulse 1. Accordingly, pulses 2, 3, and 4 probed the reaction after 0.89  $\mu$ s, 1.78  $\mu$ s, and 2.67  $\mu$ s respectively.



With this sequence, every third pulse probed the reaction with the same time delay and,  $176/3 \cong 58$  laser activations were achieved per pulse train.

The data were separated into different time-points according to the pulse identity (ID). Data collection statistics of all time-points collected at 1.13 MHz are shown in table 3.4. DED maps (Section 2.4.7.1) of each timepoint with reference to the pure dark were calculated. With a jet speed of 30m/s and a laser focus of 42  $\mu$ m (FWHM), the activated volume of a sample should exit the X-ray interaction region within  $\frac{42}{30} \times \frac{100}{67} = 2 \mu$ s. However, similar DED features are seen for all three-time-points (Fig. 3.4). The DED map for 2.67  $\mu$ s (Fig. 3.4d), which essentially should

be free of signal, also has similar features. All of these maps contain a mixture of PYP intermediate pR1 and pR2 that occurs in the early  $\mu\text{s}$  time range. The DED maps of all three-time-points – 0.89  $\mu\text{s}$ , 1.78  $\mu\text{s}$  and 2.67  $\mu\text{s}$ - completely resemble that of 1  $\mu\text{s}$  time delay collected at LCLS (Tenboer et al., 2014). This shows that all three pulses after activation are affected by the optical laser. In addition to this, the plot of hit rate vs pulse ID shows an abrupt decay from 2% to 1% in hit rate between the first two pulses of all trains (Fig. 3.5a). This verifies that each X-ray pulse is not probing new fresh and pristine crystal. The maximum jet velocity achievable with our dense PYP slurry is not enough to reliably replace the sample at the X-ray interaction region (Pandey et al., 2019). Consequently, both optical laser and X-ray frequencies are too high for this data collection scheme. As a result, X-ray pulse repetition rates and the optical laser frequency were reduced to 564 kHz and 141 kHz respectively.

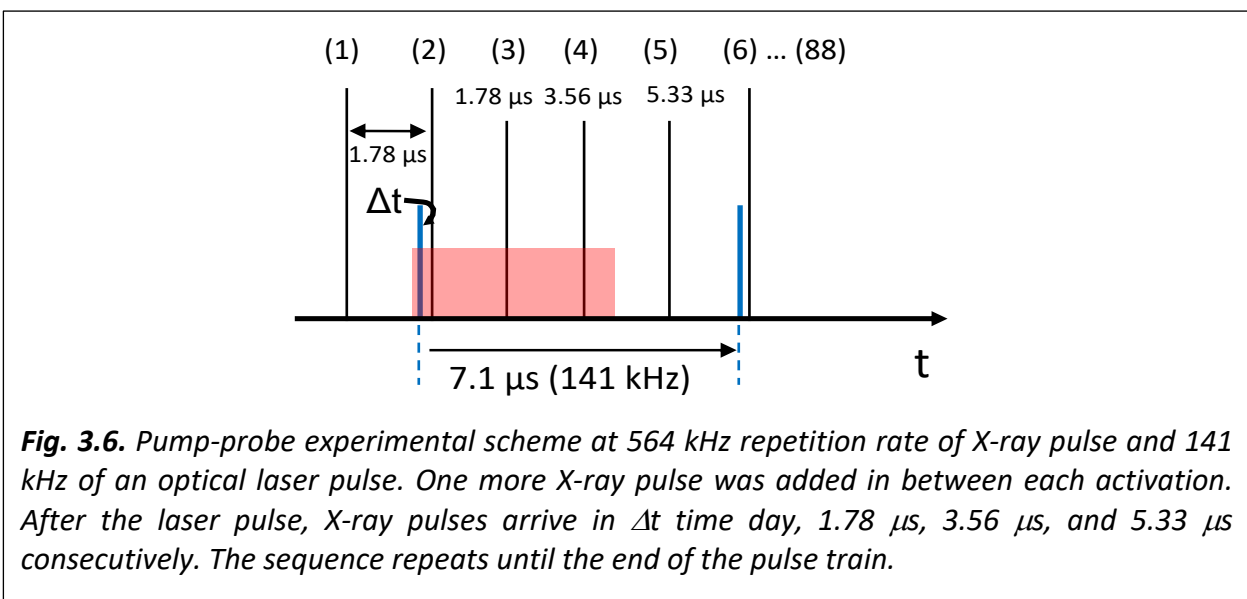


**Table 3.4** Statistics of data collected at 1.13 MHz X-ray pulse and 376 kHz optical laser repetition rates. Highest resolution shell statistics are shown inside the bracket.

<b>Control experiments</b>				
Repetition rates X-ray/laser	1.13 MHz/376 kHz			
Temperature	285 K			
<b>Data collection</b>				
Space group	P6 <sub>3</sub> (173)			
Cell dimensions a, b, c (Å) $\alpha, \beta, \gamma$ (°)	66.9, 66.9, 40.8 90, 90, 120			
Resolution (Å)	1.71 (1.71 - 1.74) Å			
	<b>0.89 <math>\mu</math>s</b>	<b>1.78 <math>\mu</math>s</b>	<b>2.67 <math>\mu</math>s</b>	<b>pure dark</b>
Number of hits	13,642	13,734	14,183	5,725
Hit/indexing rate [%]	1.1/60.0	1.1/60.0	1.2/60.0	0.6/85.7
Reflections observed	992,449	1,010,839	973,205	662,072
No of unique reflections	11,502	11,510	11,497	11,504
R <sub>split</sub> (%)	14.9 (76.8)	14.7 (76.8)	14.6 (70.6)	18.1 (111.1)
CC1/2 (%)	97.9 (50.0)	98.0 (51.4)	97.9 (48.6)	96.1 (33.8)
Completeness (%)	99.5 (98.7)	99.6 (98.7)	99.5 (98.4)	99.5 (100.0)
Redundancy	86.3 (13.9)	87.8 (13.2)	84.6 (12.7)	57.6 (19.0)

### 3.3.2 Picosecond time delay data collection at 564 kHz

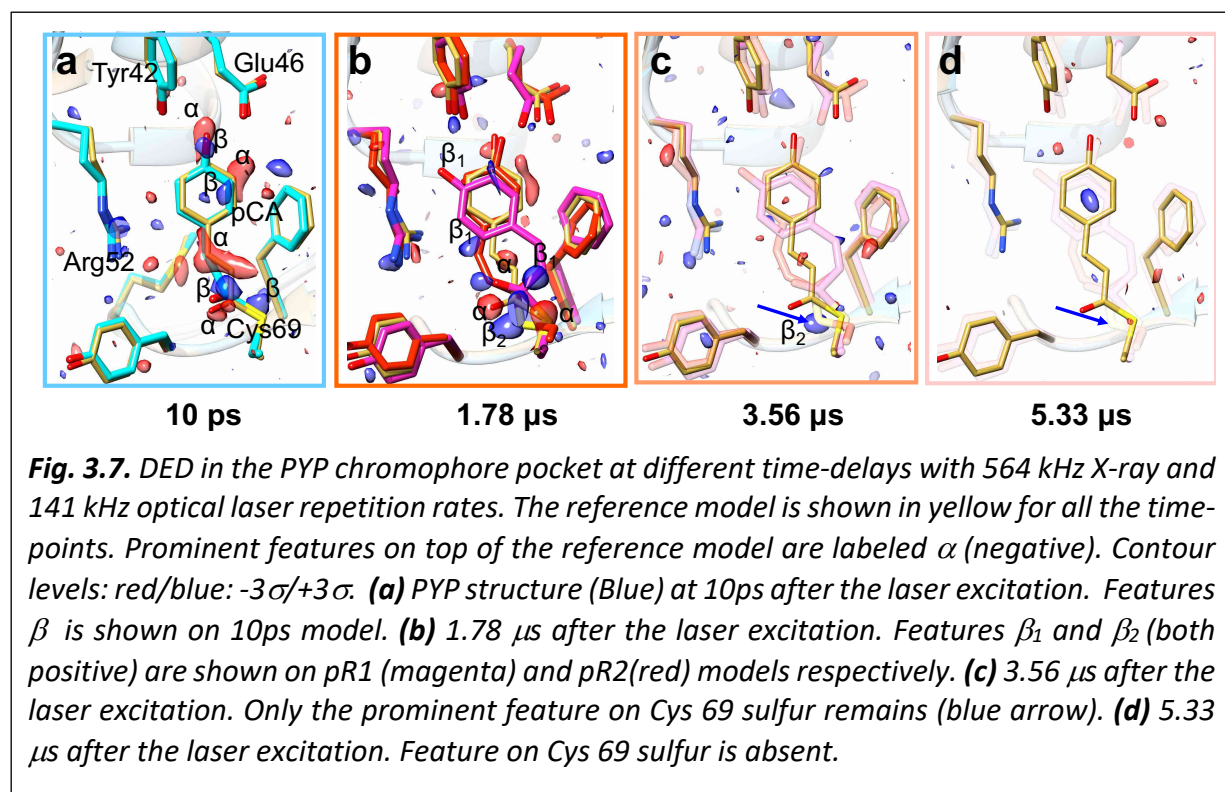
With the new X-ray repetition rate, the hit rate across the entire train remains essentially constant (Fig. 3.5b). This shows that the sample is sufficiently refreshed in the X-ray interaction region before the next X-ray pulse arrives at 564 kHz. With the new frequency of optical laser pulse, every fourth pulse is activated. The first pulse of the optical laser was synchronized at 10ps before the second X-ray pulse (Fig. 3.6). Consequently, every 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> pulse probe the reaction after the time delays of 1.78  $\mu$ s, 3.56  $\mu$ s, and 5.33  $\mu$ s respectively. As the X-ray repetition rate was reduced to 564 kHz, only 88 pulses in a train were available. This amounts to  $88/4 \cong 22$  laser activations per train.



DED maps for each time-points with reference to the pure dark were calculated (Figure 3.7). At 10ps time delay, a DED map with strong features similar to that of 3 ps collected at LCLS was observed (Pande et al., 2016) (Fig. 3.7a). The DED map obtained from the subsequent X-ray pulse at 1.78  $\mu$ s is completely different than that of 10 ps and shows the same features as 1.78



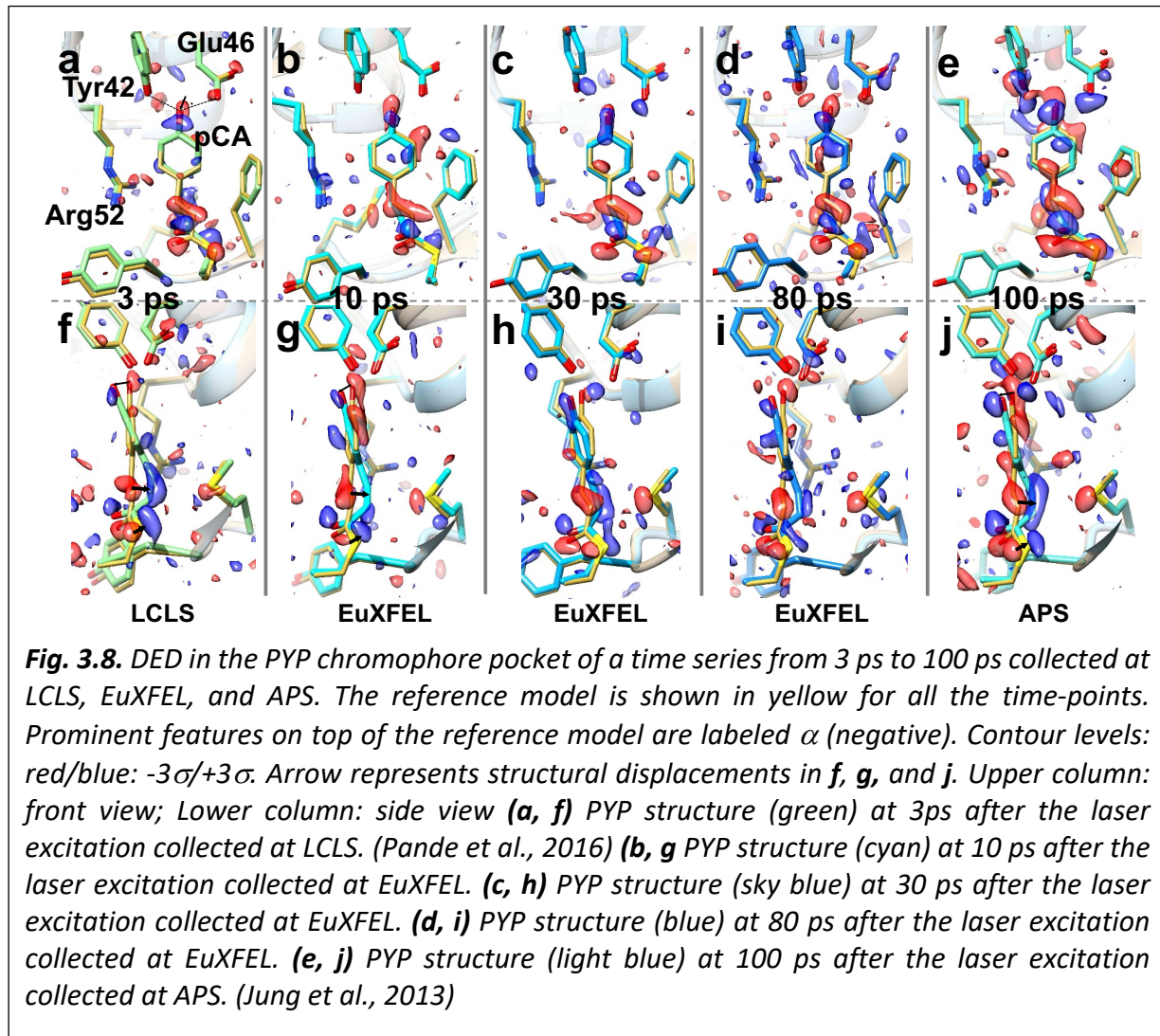
$\mu$ s collected during the control experiment. The delay 1.78  $\mu$ s is described by a previously mentioned mixture of two intermediates pR1 and pR2. At 3.56  $\mu$ s delay, the microsecond DED signal decays to a spurious positive feature as shown in Fig. 3.7c. The feature is caused due to displacement of Cys 69 sulfur at the 3.56  $\mu$ s time delay. This feature completely vanishes at 5.33  $\mu$ s and the X-ray pulse is completely free from the contamination of the optical laser pulse. It means a short time after 3.56  $\mu$ s, the jet volume excited by laser completely leaves the X-ray interaction region.



**Table 3.5** Statistics of all timepoints collected with 564 kHz X-ray pulse and 141 kHz optical laser pulse repetition rates. Highest resolution data inside the bracket.

<b>time delays</b>				
Repetition rates X-ray/Laser	564 kHz/141 kHz			
Resolution	1.6 (1.6 - 1.63) Å			
Temperature	285 K			
Space group	P6 <sub>3</sub> (173)			
Unit cell	a = 66.9 Å b = 66.9 Å c = 40.8 Å $\alpha=90^\circ$ $\beta=90^\circ$ $\gamma=120^\circ$			
	<b>10 ps</b>	<b>dark1 (1.78 <math>\mu</math>s)</b>	<b>dark2 (3.56 <math>\mu</math>s)</b>	<b>dark3 (5.33 <math>\mu</math>s)</b>
Hits	93,130	91,184	92,365	88,373
Hit/indexing rate [%]	2.8/43.5	2.8/43.4	2.8/43.4	2.7/43.7
reflections observed	4,016,763	3,929,272	4,017,291	3,883,477
Unique reflections	14,124	14,142	14,141	14,157
Redundancy	284.4 (8.5)	277.8 (7.2)	284.1 (8.2)	274.3 (8.3)
Completeness (%)	99.3 (91.9)	99.4 (93.7)	99.4 (93.3)	99.5 (95.6)
R-split (%)	6.9 (64.0)	7.0 (63.9)	7.0 (60.0)	7.3 (67.1)
CC1/2 (%)	99.6 (60.3)	99.6 (60.0)	99.6 (58.4)	99.5 (58.4)
	<b>30 ps</b>	<b>dark1 (1.78 <math>\mu</math>s)</b>	<b>dark2 (3.56 <math>\mu</math>s)</b>	<b>dark3 (5.33 <math>\mu</math>s)</b>
Hits	81,066	79,580	80,027	77,608
Hit/indexing rate [%]	1.8/50.7	1.8/50.8	1.8/50.9	1.8/51.5
reflections observed	3,948,268	3,892,799	3,921,154	3,853,579
Unique reflections	13,727	13,722	13,726	13,725

Redundancy	287.6 (17.9)	283.7 (17.2)	285.7 (17.7)	280.8 (18.2)
Completeness (%)	98.1 (99.6)	98.1 (99.4)	98.1 (99.4)	98.1 (99.8)
R-split	5.6 (40.2)	5.8 (43.5)	5.8 (38.8)	6.0 (40.9)
CC1/2	99.6 (76.1)	99.6 (75.4)	99.6 (80.6)	99.5 (79.8)
	<b>80 ps</b>	<b>dark1 (1.78 <math>\mu</math>s)</b>	<b>dark2 (3.56 <math>\mu</math>s)</b>	<b>dark3 (5.33 <math>\mu</math>s)</b>
Hits	30,860	29,264	28,332	28,235
Hit/indexing rate [%]	1.1/60.4	1.0/60.1	1.0/62.2	1.0/60.3
reflections observed	2,148,177	2,020,504	2,052,291	1,969,452
Unique reflections	13,711	13,700	13,705	13,703
*Redundancy	156.7 (15.1)	147.5 (14.1)	149.7 (15.1)	143.7 (14.7)
Completeness (%)	98.0 (97.3)	97.9 (97.1)	98.0 (97.3)	97.9 (97.5)
R-split	8.88 (49.9)	9.58 (49.8)	9.62 (49.5)	9.73 (50.3)
CC1/2	99.1 (68.4)	98.9 (70.8)	98.9 (70.5)	98.8 (67.0)



Using the same scheme as for 10ps, we collected datasets at 30 ps and 80 ps time delays as well. Data reduction statistics of all time-points collected with 564 kHz X-ray pulse repetition rates are shown in table 3.5. All of these datasets resulted in DED maps with interpretable features at the near-atomic resolution. Along with these time-points, the 3ps delay dataset collected at LCLS and 100ps delay dataset collected at APS were revisited. Figure 3.8 shows the

DED maps for all of these time-points. The molecular structures of all the time-points are obtained by using method as described in section 2.4.7.3.

Fig. 3.8 f-j shows the side view of the pCA chromophore at various time delays. These figures show the twisted cis configuration of the chromophore. The torsion angle  $\phi_{\text{tail}}$  (Fig. 1.7b) of chromophore tail was observed as 39°, 51°, 54°, 40°, and 30° for 3ps, 10ps, 30ps, 80ps, and 100ps respectively. This shows that that the torsional angle at 3ps increases at 10ps and 30ps and relaxes through 80 ps to the final value at 100 ps.

The number of excited molecules for 10, 30, and 80 ps time delays is estimated to be ~ 7%. With femtosecond optical laser pulse, the maximum obtainable primary photoexcitation yield is around 20% when excited at absorption maximum (Hutchison and van Thor, 2017). The yield was diminished because the excitation was achieved at 420 nm rather than the central absorption peak at 450 nm. Nevertheless, diffraction data produced interpretable DED maps in our case, because the laser penetration depth matched the crystal size.

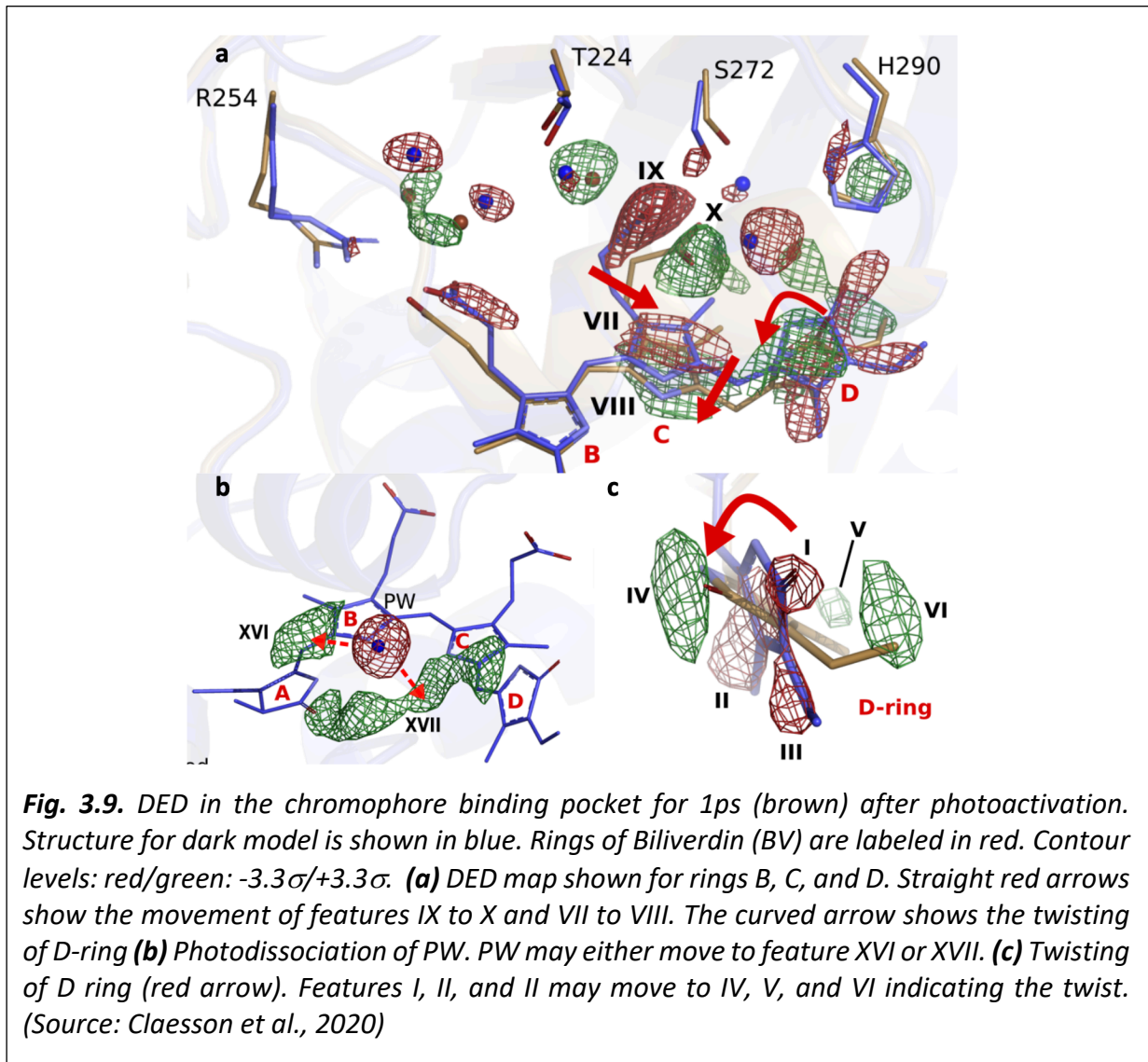
### 3.4 Primary structural photo-response of Phytochrome captured using pump-probe TR-SFX at SACLA

The photoreaction of the CBD construct of DrBphP was investigated at 1ps and 10ps time delays after activation (Section 1.3.3 and 2.8). A total of 351920 hits were collected with an average hit rate of 60%. Cheetah adapted for SACLA (Section 2.4.2) was used for the hit finding (Nakane et al., 2016). The datasets had resolution up to 2 Å and resulted in DED maps with an exceptionally strong signal. The data statistics are shown in table 3.6.

**Table 3.6** Statistics of all timepoints collected at SACLA. Highest resolution data inside the bracket.

Timepoints	Dark	1ps	10ps
Space Group	P212121	P212121	P212121
Cell dimensions (a, b, c)			
a, b, c (Å)	54.98 116.69 117.86	54.98 116.69 117.86	54.98 116.69 117.86
$\alpha, \beta, \gamma$ (°)	90.0 90.0 90.0	90.0 90.0 90.0	90.0 90.0 90.0
Data resolution overall (Å)	45.77–2.07 (2.10–2.07)	41.46–2.21 (2.25–2.21)	45.77–2.14 (2.17–2.14)
Rsplit (%)	5.79 (120.05)	10.59 (114.64)	5.70 (121.86)
SNR ( $I/\sigma(I)$ )	9.21 (0.83)	6.10 (0.88)	10.11 (0.99)
CC (1/2)	0.99 (0.33)	0.98 (0.38)	0.99 (0.344)
Completeness (%)	100 (100)	100 (100)	100 (100)
Multiplicity	461.35 (65.9)	106.11 (34.3)	347.36 (62.1)
Number of hits	149074	42853	159997
Number of indexed patterns	70726	21150	70335
Indexing rate (%)	47.44	49.35	43.96

Number of total reflections	24017763	5310179	17823530
Number of unique reflections	52060	39316	43279



A search model for molecular replacement was retrieved from the protein data bank (Berman et al., 2000) with the access code 5K5B. The model 5K5B has a single subunit. So, a second subunit was added in the asymmetric unit for accuracy in refinement. DED maps (Section

2.4.7.1) were calculated for the interpretation of the data. Extrapolated structure factors were calculated for determining the light structures (Section 2.4.7.3). DED maps for both time-points are similar. However, the map at 10 ps was less interpretable due to weak electron density features than the 1ps. As a result, only the 1ps time-point is studied.

The DED map of 1ps shows many strong features in its chromophore region. Important signals like the twist of D-ring (Fig. 3.9c) and photodissociation of pyrrole water (PW) are observed (Fig. 3.9b). Similarly, we observed strong negative density features on the atoms of the D-ring (Fig. 3.9c) of the BV and positive density at both faces of the ring. These features indicate that the ring twist from around  $20^\circ$  in a dark to  $60^\circ$  (monomer A) and  $90^\circ$  (monomer B) at 1ps after the photoexcitation. Following the twisting of the ring, we observe the translation of C-ring by approximately  $0.69\text{\AA}$  and the C-ring propionate chain detaches from its conserved anchoring residues Ser272 and Ser274 (Fig. 3.9a) (Features IX and X) (Claesson *et al.*, 2020). Strong difference electron density around the A-ring of BV shows the movement of the A-ring (away from PW). The coordinated structural changes in the binding pocket are observed which result in the liberation of the chromophore.



## 4. Discussions

Here, results from our MISC and pump-probe TR-SFX experiments are discussed. Their significance is first pointed out separately. In the end, they are summarized together to connect MISC and pump-probe TR-SFX experiments and are linked to future opportunities at XFELs.

### 4.1 MISC at XFELs

A mix-and-inject study is fundamentally a diffusion experiment where the reaction is initiated by diffusion of a substrate into a crystal. For optimum reaction initiation, one molecule of ligand per active site of an enzyme is required. This 1:1 ratio of substrate to catalytic cleft is called stoichiometric concentration (SC). SC is the minimum concentration required to establish essentially full occupancy of the ligand in the catalytic cleft under optimal condition (Schmidt, 2020). As the concentration of protein is on the order of 15-25 mmol/L and higher in crystals (Schmidt et al., 2013; Olmos et al., 2018), the SC becomes very high as well. However, if the substrate concentration is higher than the enzyme concentration, SC is reached much faster than the diffusion time.

Assuming the diffusion coefficient of  $2.3 \times 10^{-6} \text{ cm}^2/\text{s}$  for CEF in water, the characteristic diffusion times into the center of the  $10 \times 10 \times 3 \text{ }\mu\text{m}^3$  shards and the  $5 \times 2 \times 2 \text{ }\mu\text{m}^3$  needles would be a few milliseconds. The diffusion times in crystals may be very different and much slower than in the solution. At the characteristic diffusion times, about 63% of the CEF concentration outside the crystal is reached in the crystal center (Olmos *et al.*, 2018). In the second MISC experiment (Section 3.1), BlaC crystal suspensions were mixed typically 1:4 with about 200-300 mmol/L CEF. Although CEF concentration is slightly diluted after mixing, it is much higher than the

concentration of BlaC molecules in the crystal (16 mmol/L for Shard crystal form and 30 mmol/L for needles crystal form). The SC of CEF is reached at a time  $t$  much faster than the diffusion time  $\tau_D$ . This time  $t$  can be estimated as  $t = \tau_D \cdot f$ , where  $f = -\ln(a - s/C_{out})$ ,  $s$  is the concentration of BlaC in the crystals and  $C_{out}$  is the outside concentration.  $f$  is  $\sim 0.1$  for shards and 0.2 for needles. As a result, with 200mmol/L initial concentration of CEF, SC is reached at 0.36 ms and 0.17 ms for shards and needles crystals forms respectively (Olmos et al., 2018).

The characteristic diffusion time of the substrate into the centers of crystals can vary due to many reasons such as differences in crystal packing, crystal size, and crystal morphology (Orville, 2018). The crystal form with less solvent content and narrower intra-lattice channels will take longer to diffuse than the crystal form with larger solvent content and larger channels (Geremia et al., 2006). The solvent volume was estimated on the order of 59% for shards crystal form and 28% for needles crystal form. Fig. 2.16 shows the substantial differences in the solvent channel sizes for two crystal forms. The pore size in the shard crystal form is more than 4 times greater than that of the needle crystal form (Fig. 2.16). This facilitates the diffusion of a large molecule such as CEF in shards crystals and produces a strong electron density signal in the catalytic cleft. So, evaluating more than one kind of crystal form may increase the possibility of a successful MISC experiment. Similarly, to initiate a reaction within a crystal effectively and produce a strong electron density signal in the catalytic cleft, the crystal has to be as small as possible, the concentration of substrate outside crystal should be as high as possible and the diffusion coefficient should be as large as possible (Schmidt, 2020). To measure the diffusion time in protein crystals is a formidable challenge and might be possible with the MISC technique. For this, occupancies in active centers must be related to concentrations of ligand delivered into the

crystals by diffusion. Experiments with time-delays much faster than 30 ms are underway to achieve this.

## 4.2 Structure dynamics of pCA chromophore after photoexcitation

The first TR-SFX experiment on PYP was conducted with Coherent X-ray imaging instrument at the LCLS in 2014 (Tenboer et al., 2014). The experiment produced DED maps with an exceptionally good signal that showed that the TR-SFX is feasible at XFEL. Consequently, another experiment on PYP was conducted at the same instrument and facility (Pande et al., 2016). This experiment achieved a time resolution of 140 fs to cover the fs time scale up to 3ps. A 1.6 Å molecular movie shows the trans-cis isomerization of central pCA chromophore (Fig. 1.7). The isomerization occurs approximately 600fs after the photoexcitation and proceeds through a so-called conical intersection. The conical intersection is the joint between the electronically excited stated potential energy surface (PES) and electronic ground state PES. PYP is in electronically excited state between 100fs and 400fs, and the electronic ground state is recovered after 700fs (Pandey, Poudyal and Malla, 2020).

**Table 4.1:** Geometry of pCA chromophore after refinement

	Dark	3ps	10ps	30ps	80ps	100ps
Hydrogen bond pCA to Glu-46 (Å)	2.55	3.30	2.82	2.87	2.86	2.79
Hydrogen bond pCA to Tyr- 42 (Å)	2.58	2.57	2.65	2.62	2.64	2.39
$\phi_{\text{tail}} (^{\circ})$	172	39	51	54	40	30

As the *trans* to *cis* isomerization progresses, a large rotation about the C2=C3 double bond of pCA chromophore (Fig. 1.7b) occurs at ~600fs. The structure then relaxes to the ground state

PES after 700 fs and the torsional angle decreases to  $\varphi_{\text{tor}} = 39^\circ$  at 3ps. With X-ray data collected at the EuXFEL, the ps range between 3 ps and 100 ps was explored (Section 2.7 and 3.3). The torsional angle increases at 10ps ( $51^\circ$ ) and 30ps ( $54^\circ$ ) and relaxes through 80ps ( $40^\circ$ ) to a final value at 100ps ( $30^\circ$ ) (Table 4.1 and Fig. 3.8). The torsional relaxations up to 3ps occur together with an increase of hydrogen bond distance between pCA-O<sub>4</sub> and Glu-O<sub>e</sub>. After 80ps, the hydrogen bonds relax to shorter distances and approach the distances as observed in the dark structure (Table 4.1). These structures of PYP obtained over the ps time range provide direct structural evidence showing that an energetically highly strained structure reorders and slowly relaxes over a long time. A nonexponential, nonergodic, and ultrafast relaxation from a high energy state towards a longer-lived thermal reservoir, characterized by intermediate  $\tau$  is directly observed (Fig 1.7).

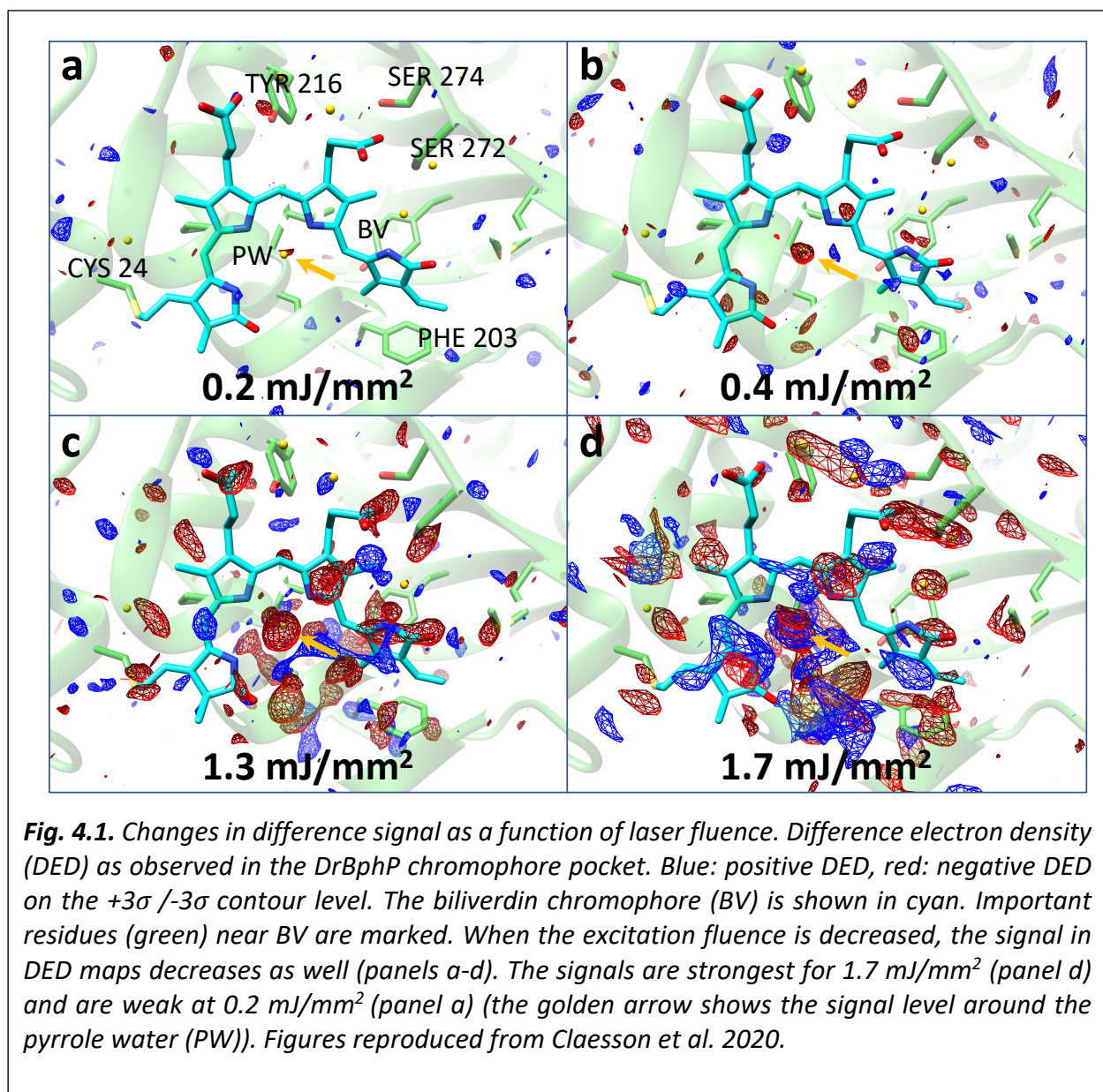
### 4.3 Primary structural photoresponse of phytochrome

After 1ps photoexcitation, the structure of CBD construct of DRBphP shows changes of BV chromophore(Fig. 1.8) and the surrounding residues such as the twist of D-ring, displacement of C-ring, and associated changes of water network which connects D-ring and C-ring propionate and His-290 (Section 3.4 and Fig. 3.9) (Claesson *et al.*, 2020). Furthermore, significant changes are observed around A-ring, Asp-207, and PW. These changes are retained even at 10ps, however at a lower population. These extensive and coordinated structural changes around the binding pocket show the liberation of the chromophore from the protein scaffold. We propose that these changes are necessary for the conformational rearrangement during photoconversion of bacterial phytochrome to Pfr state.

The D-ring of the bilin chromophore isomerizes around the around C15-C16 bond from Z in Pr to E in Pfr state (Burgie et al., 2014; Takala et al., 2014). Based on anticipated steric clashes with the C-ring methyl group, it has been proposed that the D-ring rotates counter-clockwise in plant and cyanobacterial phytochromes, but clockwise in bacterial phytochromes (Rockwell et al., 2009). However, for bacterial DrBPhP, we observe that the D-ring is rotated counter-clockwise by tens of degrees at 1ps time delay. This is strongly supported by the DED map (Fig. 3.9 c). Two positive peaks (V and VI, Fig. 3.9c) in the DED map indicate the new positions of the vinyl and methyl group of D-ring which verifies the counter-clockwise rotation. However, we cannot fully exclude the fact that the direction of rotation might have been influenced due to the truncation of the phytochrome construct or the crystal packing. NMR studies have reported conformational heterogeneity in the chromophore binding pocket of phytochromes in solution (Song et al., 2011, 2018; Lim et al., 2018). Crystallization could select one of the conformations, which may result in a preferred rotation of D-ring in the counter-clockwise direction. More experiments, perhaps another crystal form is necessary to shine a light on this behavior.

In a pump-probe TR-SFX, it is very important to activate a large fraction of the molecules without damaging the crystal. A reaction must be initiated rapidly, uniformly, and nondestructively in the crystal with enough laser power (Moffat, 2001). Low laser power will prevent sufficient photo-initiation, whereas high power deposits energy and may damage the crystal (Hutchison et al., 2016). Therefore, it is necessary to investigate the reaction beforehand by spectroscopy or by pump power titration to determine the appropriate photoexcitation regime.

A power titration was conducted during the experiment with the CBD construct of DrBphP at SACLA (See 3.4)(Claesson et al., 2020). Fluences of 1.7 mJ/mm<sup>2</sup>, 1.3 mJ/mm<sup>2</sup>, 0.4 mJ/mm<sup>2</sup>, 0.2 mJ/mm<sup>2</sup> were used and DED maps were calculated for each energy. Although prominent DED features are still present at similar positions in all fluences (Fig. 4.1), the DED signal becomes highest at the fluence 1.7 mJ/mm<sup>2</sup> and smallest at 0.2 mJ/mm<sup>2</sup>. This shows that the higher laser fluences do not alter the result of this experiment and rather, positively contribute to stronger DED maps that can be interpreted more easily (Pandey, Poudyal and Malla, 2020).



Several factors can influence the laser power reaching the sample position. For instance, a grease jet may scatter the light, thereby decreasing the light reaching the sample. The reaction may not initiate completely and the whole beamtime may end as a failure. Pump power titration can be an important method to determine the laser fluence in pump-probe TR-SFX. Its most important advantage is that it can be conducted with the same experimental setup as the actual experiment. As a result, a pump power that gives rise to the highest and interpretable signal in DED maps can be directly used to perform the actual experiment. Although one can argue that power titration takes a period of important beamtime, however, with the advent of high repetition XFEL, the time consumed by performing power titration won't be significant than the advantages provided by this method.

#### **4.4 TR-SFX at the high repetition XFEL, EuXFEL**

High repetition rates XFELs enable the collection of a large number of diffraction patterns at much faster rates. When the final design specification is reached, the EuXFEL can deliver 27000 pulses/s (Table 2.3 and Fig. 2.18) . Although only 3500 patterns/s can be stored and streamed out by AGIPD, this value is still more than 16 times greater than the rate diffraction patterns are produced at the LCLS. Assuming a hit rate of 5% and an indexing rate of 60%, a data set containing 25000 indexed patterns can be collected only in 4 minutes (Schmidt, 2020). If technological development permits the use of all 27000 pulses/s, a dataset can be collected within a minute. In the future, the increased data rate at the EuXFEL may support the collection of a large amount of data within very limited experimental time. This opens the door to the deployment of enhanced analysis methods such as singular value decomposition (SVD) to extract

macromolecular structures and their dynamics in the crystalline ensemble (Schmidt et al., 2003). SVD is used to deconvolute the mixture of one or more intermediates and obtain structure factors of time-independent intermediates. This has already been achieved on slower timescales that were covered by several TRX experiments collected over multiple days of beamtimes at synchrotrons (Ihee *et al.*, 2005; Schmidt *et al.*, 2013; Pandey *et al.*, 2019). However, with EuXFEL, a larger number of datasets that cover the entire reaction can be collected only in 1 shift. This allows a deeper understanding of relationships between structural dynamics and their functions efficiently.

In pump-probe experiments, optical laser pulses are synchronized to the X-ray pulses. Time delays are determined by temporal differences between frequencies of pump pulses and probe pulses. Extra caution should be taken to avoid light contamination as mentioned in section 3.3.1. However, MISC experiments at high repetition XFELs are less complicating than the pump-probe experiments. The time delays in MISC can be fixed by changing the flow rate and the distance traveled by the microcrystal's suspension. The experiment basically becomes SFX where all the X-ray pulses contribute to the same time delay and the high repetition pulses are used efficiently.

#### **4.5 Pump-probe and MISC experiments at XFELs**

Pump-probe TR-SFX can be used only if a photo-reactive compound called chromophore is present in the biological macromolecule. For the molecules without chromophore, a different technique has to be used. Several experiments have been performed where non-photoactive molecules are mixed with caged and light activatable substrates (Section 2.3.4). Although this method has provided insights on many enzymes (Schlichting et al., 1990; Goeldner and Givens,



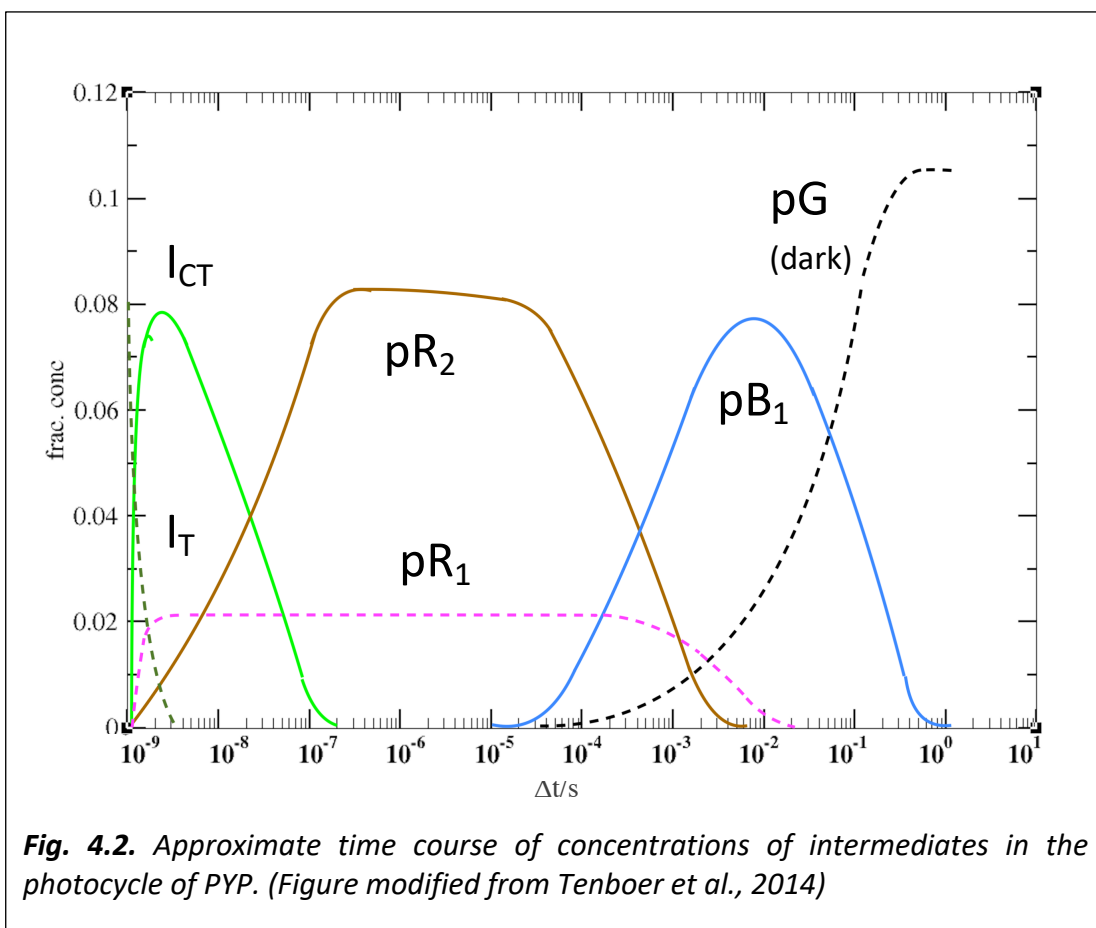
2005), the unavailability of the suitable caged substrate has limited the studies of many proteins. This limitation can be overcome using MISC, which doesn't require caged substrate at all. Therefore, with pump-probe TR-SFX and MISC together, the dynamics of both photoactive and non-photoactive molecules can be studied.

#### **4.5.1 From oscillations on ultrafast timescales to transient intermediates**

Crystallography is an ensemble method that averages over a large number of molecules. With crystallography, it is only possible to observe the dynamics of the ensemble rather than that of a single molecule. At room temperature and at thermal equilibrium, protein molecules fluctuate constantly through different conformational substates (Austin *et al.*, 1975). It is important to note two types of motions: oscillations, and relaxation processes (Frauenfelder, Sligar and Wolynes, 1991; Parak and Frauenfelder, 1993). In thermal equilibrium, short-lived higher-energy states are created (with a probability given by the Boltzmann factor) for example by collisions with water molecules. The corresponding structures are then relaxing back towards the energy minimum. Relaxations also take place when non-equilibrium states are created by a chemical reaction or by using a laser pump pulse. The molecules relax towards thermal equilibrium, most likely using the same diffusive-like modes of motion as in thermal equilibrium.

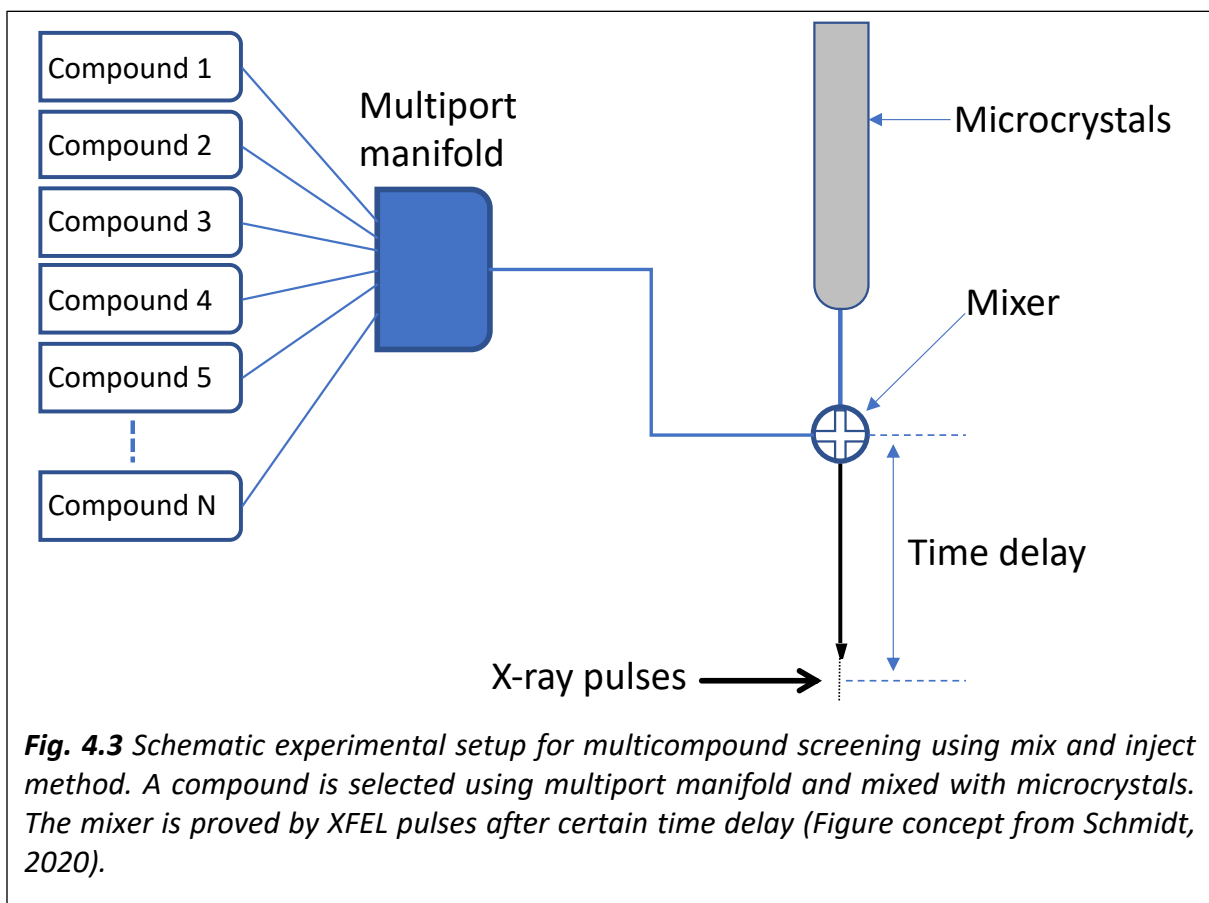
Oscillations and relaxations may be on very different time scales. Dynamically, proteins show a complex behavior (Parak and Frauenfelder, 1993). On the ultrafast time scale (or at cryogenic temperatures), the protein dynamics are governed by fast oscillations that decay fast and de-phase quickly. A once established synchronicity between oscillating molecules decays. In our experiments, we investigated the protein dynamics on longer pico-second timescales where the fast oscillations should have been all kinetically de-phased. However, relaxations happen on

much longer characteristic time scales up to even nanoseconds (Parak, Knapp and Kucheida, 1982). On timescales up to a few 100 ps, global structural relaxations should still be correlated through the ensemble. Relaxations of protein moieties such as  $\alpha$ -helices were investigated earlier (Parak et al., 2006; Schmidt et al., 2009). They are also identified in protein folding experiments, where structures relax into a so-called folding funnel down a rugged energy landscape (Onuchic et al., 1996; Wolynes, Luthey-Schulten and Onuchic, 1996). A similar scenario is observed for PYP (Section 1.3.2 and Fig. 1.7). A smooth structural transition without distinct intermediates is observed until  $\sim 100$ ps after reaction initiation (Fig. 3.8). The ensemble of molecules apparently relaxes in synchronicity through the 10ps, 30ps, and 80ps time-points (Fig. 3.8, Table 4.1). Finally, the first true intermediate  $I_T$  (Fig. 1.7 a) is observed around 1 ns.



On longer timescales, even the relaxations de-phase and become uncorrelated in the ensemble. Then the dynamics of the crystalline ensemble is governed by simple chemical kinetics (Steinfeld, Francisco and Hase, 1998). The reaction can be described by a chemical, kinetic mechanism with rate coefficients and intermediate states (Fig. 4.2). Caused by thermal excitations, members of the crystalline ensemble may overcome barriers of activation and advance to the next intermediate as also observed in PYP (Ihee *et al.*, 2005; Jung *et al.*, 2013; Tenboer *et al.*, 2014). As relaxations are not synchronous, and the occupancies of molecules on top of the barriers of activation are very small, the trajectory from one intermediate to the next is impossible to determine from the crystalline ensemble. This has also been observed in the mix-and-inject experiments on BlaC where the reaction dynamics can be described by time-dependent occupancy variations of the enzyme-substrate complex and the covalently bound acyl complex (Section 3.2 and Fig. 3.2 ). From a more elevated perspective, our observations on PYP and BlaC are explained by diffusion-like polymer dynamics on fast time scales and by conventional chemical kinetics on longer times scales. This fine balance between an unstructured polymer and a molecule with a rigid structure enables and promotes the functions of the proteins investigated here, and most likely the functions of all proteins.

#### 4.5.2 Rapid ligand screening to support structure-based drug design



Rapid data collection at high repetition XFELs has created new opportunities in structure-based time-resolved crystallography. With amazingly shortening of the data collection time, a large number of datasets with many time-points of a chemical reaction can be collected. This can help in rapid ligand screening. As shown in figure 4.3, an experiment can be conceived where hundreds of compounds can be tested within a shift. Many small chemicals, drugs, or drug fragments can be tested by simply mixing with the microcrystals (Schmidt, 2020). The experiment can be simplified by using a T-junction for mixing as shown by Kupitz et al., 2017 (Fig. 2.15). As a result, MISC with high repetition XFELs can become an important tool for structure-based drug

design. They can contribute to the knowledge that can save many lives by directly observing molecules of biomedical importance in action or by probing a drug target with a multitude of active compounds.

#### **4.5.3 From TR-SFX to time-resolved single particles imaging**

With TR-SFX, structures of intermediates, the reaction pathways, and rate coefficients connecting reaction intermediates can be determined. However, time-resolved changes of some photosensitive proteins such as phytochromes, photosynthetic reaction center, still need to be revealed. Proteins such as phytochromes (Fig. 1.8) have a conformational change of tens of Å between light and dark-adapted states. It has to be determined whether these large conformational changes that occur during a reaction are compatible in the crystalline lattice (Pandey, Poudyal and Malla, 2020).

In crystallography, molecules may crystallize in different conformation due to structural heterogeneity and may follow different reaction pathways. There is also a danger that the average structure hides important functions and allosteric molecular mechanisms (Kruschel and Zagrovic, 2009; Pandey, Poudyal and Malla, 2020). This makes difficulties in accurately understanding the reaction mechanism (Song et al., 2011, 2018; Lim et al., 2018). Observing one molecule at a time may provide a solution to this and avoid information loss caused by ensemble averaging. Intense X-ray pulses produced by XFELs can produce single-particle diffraction patterns by imaging single biomolecules (Hosseinizadeh et al., 2017). With a sufficiently large number of single-particle diffraction patterns, it is possible to reconstruct the electron density at a moderate atomic resolution (Poudyal, Schmidt and Schwander, 2020). And with high repetition XFELs, producing a sufficient number of diffraction patterns is less of a problem.

## 5. References

- Alper, J. (1994) 'Boning up: Newly isolated proteins heal bad breaks', *Science*. doi: 10.1126/science.8278805.
- Anderson, S., Crosson, S. and Moffat, K. (2004) 'Short hydrogen bonds in photoactive yellow protein', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444490400616X.
- Auldridge, M. E. and Forest, K. T. (2011) 'Bacterial phytochromes: More than meets the light', *Critical Reviews in Biochemistry and Molecular Biology*. doi: 10.3109/10409238.2010.546389.
- Austin, R. H. *et al.* (1975) 'Dynamics of Ligand Binding to Myoglobin', *Biochemistry*. doi: 10.1021/bi00695a021.
- Bae, G. and Choi, G. (2008) 'Decoding of Light Signals by Plant Phytochromes and Their Interacting Proteins', *Annual Review of Plant Biology*. doi: 10.1146/annurev.arplant.59.032607.092859.
- Barends, T. R. M. *et al.* (2015) 'Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation', *Science*. doi: 10.1126/science.aac5492.
- Barty, A. *et al.* (2012) 'Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements', *Nature Photonics*. doi: 10.1038/nphoton.2011.297.
- Barty, A. *et al.* (2014) 'Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data', *Journal of Applied Crystallography*. doi: 10.1107/S1600576714007626.
- Berman, H. M. *et al.* (2000) 'The Protein Data Bank', *Nucleic Acids Research*. doi:

10.1093/nar/28.1.235.

Berman, H. M. *et al.* (2002) 'The protein data bank', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444902003451.

Björn, L. O. (2015) 'Photoactive proteins', in *Photobiology: The Science of Light and Life, Third Edition*. doi: 10.1007/978-1-4939-1468-5\_11.

Blaj, G. *et al.* (2015) 'X-ray detectors at the Linac Coherent Light Source', *Journal of Synchrotron Radiation*. doi: 10.1107/S1600577515005317.

Blake, C. C. F. *et al.* (1965) 'Structure of hen egg-white lysozyme: A three-dimensional Fourier synthesis at 2 Å resolution', *Nature*, 206(4986), pp. 757–761. doi: 10.1038/206757a0.

Blake, C. C. F. and Phillips, D. C. (1962) 'Biological Effects of Ionizing Radiation at the Molecular Level', *Vienna: International Atomic Energy Agency*.

Bourgeois, D. and Weik, M. (2009) 'Kinetic protein crystallography: A tool to watch proteins in action', *Crystallography Reviews*. doi: 10.1080/08893110802604868.

Boutet, S. *et al.* (2012) 'High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography', *Science*. doi: 10.1126/science.1217737.

Boyd, D. B. and Lunn, W. H. W. (1979) 'Electronic Structures of Cephalosporins and Penicillins.

9. Departure of a Leaving Group in Cephalosporins', *Journal of Medicinal Chemistry*. doi: 10.1021/jm00193a006.

Bragg, W. H., Bragg, W. L. (1949) 'The crystalline state.', *A General survey*, 1.

Bragg, W. L. (1912) 'The specular reflection of x-rays [3]', *Nature*, p. 410. doi: 10.1038/090410b0.

Bragg, W. L. (1913) 'The Structure of Some Crystals as Indicated by Their Diffraction of X-rays',

*Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. The Royal Society, 89(610), pp. 248–277. doi: 10.1098/rspa.1913.0083.

Bragg, W. L. (1914) 'The Analysis of Crystals by the X-ray Spectrometer', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. The Royal Society, 89(613), pp. 468–489. doi: 10.1098/rspa.1914.0015.

Brand, H. S., Ligtenberg, A. J. M. and Veerman, E. C. I. (2014) 'Saliva and wound healing', *Monographs in Oral Science*. doi: 10.1159/000358784.

Brnger, A. (1991) 'Simulated Annealing In Crystallography', *Annual Review of Physical Chemistry*. doi: 10.1146/annurev.physchem.42.1.197.

Brodersen, D. E. *et al.* (2000) 'The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B, on the 30S ribosomal subunit', *Cell*. Cell Press, 103(7), pp. 1143–1154. doi: 10.1016/S0092-8674(00)00216-6.

Burgie, E. S. *et al.* (2014) 'Crystallographic and electron microscopic analyses of a bacterial phytochrome reveal local and global rearrangements during photoconversion', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M114.571661.

Calvey, G. D. *et al.* (2016) 'Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers', *Structural Dynamics*. doi: 10.1063/1.4961971.

Chapman, H. N. *et al.* (2006) 'Femtosecond diffractive imaging with a soft-X-ray free-electron laser', *Nature Physics*. doi: 10.1038/nphys461.

Chapman, H. N. *et al.* (2011) 'Femtosecond X-ray protein nanocrystallography', *Nature*. doi: 10.1038/nature09750.

Claesson, E. *et al.* (2020) 'The primary structural photoresponse of phytochrome proteins



captured by a femtosecond x-ray laser', *eLife*. doi: 10.7554/eLife.53514.

Cochran, W. (1951) 'Some properties of the (F o – F c)-synthesis', *Acta Crystallographica*. International Union of Crystallography, 4(5), pp. 408–411. doi: 10.1107/s0365110x51001355.

Cornish-bowden, A. (2013) *Fundamentals of Enzyme Kinetics - Front Matter, Methods for the Study of Marine Benthos*. doi: 10.1002/9781118542392.fmatter.

Davis, S. J., Vener, A. V. and Vierstra, R. D. (1999) 'Bacteriophytochromes: Phytochrome-like photoreceptors from nonphotosynthetic eubacteria', *Science*. doi: 10.1126/science.286.5449.2517.

DePonte, D. P. *et al.* (2008) 'Gas dynamic virtual nozzle for generation of microscopic droplet streams', *Journal of Physics D: Applied Physics*. doi: 10.1088/0022-3727/41/19/195505.

Doye, J. P. K. and Poon, W. C. K. (2006) 'Protein crystallization in vivo', *Current Opinion in Colloid and Interface Science*. doi: 10.1016/j.cocis.2005.10.002.

Drenth, J. and Mesters, J. (2007) *Principles of protein X-ray crystallography: Third edition, Principles of Protein X-Ray Crystallography: Third Edition*. doi: 10.1007/0-387-33746-6.

Dubée, V. *et al.* (2012) 'Inactivation of Mycobacterium tuberculosis L,D-transpeptidase Ldt Mt1 by carbapenems and cephalosporins', *Antimicrobial Agents and Chemotherapy*. doi: 10.1128/AAC.00665-12.

Duisenberg, A. J. M. (1992) 'Indexing in single-crystal diffractometry with an obstinate list of reflections', *Journal of Applied Crystallography*. doi: 10.1107/S0021889891010634.

Duszenko, M. *et al.* (2015) 'In vivo protein crystallization in combination with highly brilliant radiation sources offers novel opportunities for the structural analysis of post-translationally modified eukaryotic proteins', *Acta Crystallographica Section:F Structural Biology*

*Communications*. doi: 10.1107/S2053230X15011450.

Fiedler, E. *et al.* (2002) 'Snapshot of a key intermediate in enzymatic thiamin catalysis: Crystal structure of the  $\alpha$ -carbanion of ( $\alpha,\beta$ -dihydroxyethyl)-thiamin diphosphate in the active site of transketolase from *Saccharomyces cerevisiae*', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.022510999.

Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991) 'The energy landscapes and motions of proteins', *Science*. doi: 10.1126/science.1749933.

Friedrich, W., Knipping, P. and Laue, M. (1913) 'Interferenzerscheinungen bei Röntgenstrahlen', *Annalen der Physik*, 346(10), pp. 971–988. doi: 10.1002/andp.19133461004.

Gan, F. *et al.* (2014) 'Extensive remodeling of a cyanobacterial photosynthetic apparatus in far-red light', *Science*. doi: 10.1126/science.1256963.

Garman, E. F. (2010) 'Radiation damage in macromolecular crystallography: What is it and why should we care?', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444910008656.

Geremia, S. *et al.* (2006) 'Simulation of diffusion time of small molecules in protein crystals', *Structure*. doi: 10.1016/j.str.2005.12.007.

Gevorkov, Y. *et al.* (2019) 'XGANDALF - Extended gradient descent algorithm for lattice finding', *Acta Crystallographica Section A: Foundations and Advances*. doi: 10.1107/S2053273319010593.

Goeldner, M. and Givens, R. S. (2005) *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules*, *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules*. doi: 10.1002/3527605592.

Greiffenberg, D. (2012) 'The AGIPD detector for the European XFEL', *Journal of Instrumentation*. doi: 10.1088/1748-0221/7/01/C01103.

Hahn, T. (2006) *International Tables for Crystallography*.

Hart, P. *et al.* (2012) 'The CSPAD megapixel x-ray camera at LCLS', in *X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Applications*. doi: 10.1117/12.930924.

Hatsui, T. (2014) 'Developments of X-ray Imaging Detectors at SACLA/SPring-8: Current Status and Future Outlook', *Synchrotron Radiation News*. doi: 10.1080/08940886.2014.930805.

Helliwell, J. R. (1985) 'Protein crystallography with synchrotron radiation', *Journal of Molecular Structure*. doi: 10.1016/0022-2860(85)85023-7.

Helliwell, J. R. (1988) 'Protein crystal perfection and the nature of radiation damage', *Journal of Crystal Growth*. doi: 10.1016/0022-0248(88)90322-3.

Helliwell, J. R. *et al.* (2002) 'Time-resolved and static-ensemble structural chemistry of hydroxymethylbilane synthase', in *Faraday Discussions*. doi: 10.1039/b201331b.

Henderson, R. and Moffat, J. K. (1971) 'The difference Fourier technique in protein crystallography: errors and their treatment', *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*. doi: 10.1107/s0567740871004060.

Henrich, B. *et al.* (2011) 'The adaptive gain integrating pixel detector AGIPD a detector for the European XFEL', in *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. doi: 10.1016/j.nima.2010.06.107.

Herrmann, S. *et al.* (2013) 'CSPAD-140k: A versatile detector for LCLS experiments', *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors*

*and Associated Equipment*. doi: 10.1016/j.nima.2013.01.057.

Hosseinzadeh, A. *et al.* (2017) 'Conformational landscape of a virus by single-particle X-ray scattering', *Nature Methods*. doi: 10.1038/nmeth.4395.

Huang, Z. and Kim, K. J. (2007) 'Review of x-ray free-electron laser theory', *Physical Review Special Topics - Accelerators and Beams*. doi: 10.1103/PhysRevSTAB.10.034801.

Hughes, J. (2013) 'Phytochrome Cytoplasmic Signaling', *Annual Review of Plant Biology*. doi: 10.1146/annurev-arplant-050312-120045.

Hunter, M. S. *et al.* (2014) 'Fixed-target protein serial microcrystallography with an X-ray free electron laser', *Scientific Reports*. doi: 10.1038/srep06026.

Hutchison, C. D. M. *et al.* (2016) 'Photocycle populations with femtosecond excitation of crystalline photoactive yellow protein', *Chemical Physics Letters*. doi: 10.1016/j.cplett.2016.04.087.

Hutchison, C. D. M. and van Thor, J. J. (2017) 'Populations and coherence in femtosecond time resolved X-ray crystallography of the photoactive yellow protein', *International Reviews in Physical Chemistry*. doi: 10.1080/0144235X.2017.1276726.

Ihee, H. *et al.* (2005) 'Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds', *Proc. Nat. Acad. Sci.* doi: 10.1073/pnas.0409035102.

Jason Tenboer, Shibom Basu, Nadia Zatsepin, K. P. *et al.* (2014) 'Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein', *Science*.

Johnson, L. N. and Phillips, D. C. (1965) 'Structure of some crystalline lysozyme-inhibitor complexes determined by X-ray analysis at 6 Å resolution', *Nature*. doi: 10.1038/206761a0.

Jung, Y. O. *et al.* (2013) 'Volume-conserving trans-cis isomerization pathways in photoactive

yellow protein visualized by picosecond X-ray crystallography', *Nature Chemistry*. doi: 10.1038/nchem.1565.

Kabsch, W. (2010) 'Integration, scaling, space-group assignment and post-refinement', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444909047374.

Kabsch, W. (2014) 'Processing of X-ray snapshots from crystals in random orientations', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S1399004714013534.

Kameshima, T. *et al.* (2014) 'Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments', *Review of Scientific Instruments*. doi: 10.1063/1.4867668.

Karplus, P. A. and Diederichs, K. (2012) 'Linking crystallographic model and data quality', *Science*. doi: 10.1126/science.1218231.

Kendrew, J. C. *et al.* (1958) 'A three-dimensional model of the myoglobin molecule obtained by x-ray analysis', *Nature*, 181(4610), pp. 662–666. doi: 10.1038/181662a0.

Kendrew, J. C. (1958) 'Architecture of a protein molecule', *Nature*, 182(4638), pp. 764–767. doi: 10.1038/182764a0.

Kern, J. *et al.* (2013) 'Simultaneous femtosecond x-ray spectroscopy and diffraction of photosystem II at room temperature', *Science*. doi: 10.1126/science.1234273.

Kirian, R. A. *et al.* (2011) 'Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals', *Acta Crystallographica Section A: Foundations of Crystallography*. doi: 10.1107/S0108767310050981.

Kong, K. F., Schneper, L. and Mathee, K. (2010) 'Beta-lactam antibiotics: From antibiosis to resistance and bacteriology', *APMIS*. doi: 10.1111/j.1600-0463.2009.02563.x.

Koopmann, R. *et al.* (2012) 'In vivo protein crystallization opens new routes in structural biology', *Nature Methods*. doi: 10.1038/nmeth.1859.

Korostelev, A., Bertram, R. and Chapman, M. S. (2002) 'Simulated-annealing real-space refinement as a tool in model building', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444902003402.

Kort, R. *et al.* (1996) 'The xanthopsins: a new family of eubacterial blue-light photoreceptors.', *The EMBO Journal*. doi: 10.1002/j.1460-2075.1996.tb00685.x.

Kruschel, D. and Zagrovic, B. (2009) 'Conformational averaging in structural biology: Issues, challenges and computational solutions', *Molecular BioSystems*. doi: 10.1039/b917186j.

Kupitz, C., Grotjohann, I., *et al.* (2014) 'Microcrystallization techniques for serial femtosecond crystallography using photosystem II from *Thermosynechococcus elongatus* as a model system', *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi: 10.1098/rstb.2013.0316.

Kupitz, C., Basu, S., *et al.* (2014) 'Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser', *Nature*. doi: 10.1038/nature13453.

Kupitz, C. *et al.* (2017) 'Structural enzymology using X-ray free electron lasers', *Structural Dynamics*, 4(4). doi: 10.1063/1.4972069.

Lamb, A. L., Kappock, T. J. and Silvaggi, N. R. (2015) 'You are lost without a map: Navigating the sea of protein structures', *Biochimica et Biophysica Acta - Proteins and Proteomics*. doi: 10.1016/j.bbapap.2014.12.021.

Lewis, K. (2013) 'Platforms for antibiotic discovery', *Nature Reviews Drug Discovery*. doi: 10.1038/nrd3975.

- Liang, M. *et al.* (2015) 'The Coherent X-ray Imaging instrument at the Linac Coherent Light Source', *Journal of Synchrotron Radiation*. doi: 10.1107/S160057751500449X.
- Lim, S. *et al.* (2018) 'Correlating structural and photochemical heterogeneity in cyanobacteriochrome NpR6012g4', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1720682115.
- Lomb, L. *et al.* (2011) 'Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser', *Physical Review B - Condensed Matter and Materials Physics*. doi: 10.1103/PhysRevB.84.214111.
- Mancuso, A. P. *et al.* (2019) 'The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european XFEL: Initial installation', *Journal of Synchrotron Radiation*. doi: 10.1107/S1600577519003308.
- Mariani, V. *et al.* (2016) 'OnDA: Online data analysis and feedback for serial X-ray imaging', *Journal of Applied Crystallography*. doi: 10.1107/S1600576716007469.
- Martin-Garcia, J. M. *et al.* (2019) 'High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source', *IUCrJ*. doi: 10.1107/S205225251900263X.
- McPherson, A. and Gavira, J. A. (2014) 'Introduction to protein crystallization', *Acta Crystallographica Section F: Structural Biology Communications*. doi: 10.1107/S2053230X13033141.
- Meents, A. *et al.* (2017) 'Pink-beam serial crystallography', *Nature Communications*. doi: 10.1038/s41467-017-01417-3.
- Meyer, T. E. (1985) 'Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira*

halophila', *BBA - Bioenergetics*. doi: 10.1016/0005-2728(85)90094-5.

Meyer, T. E. *et al.* (1987) 'Properties of a Water-Soluble, Yellow Protein Isolated from a Halophilic Phototrophic Bacterium That Has Photochemical Activity Analogous to Sensory Rhodopsin', *Biochemistry*. doi: 10.1021/bi00376a012.

Miller, M. T. *et al.* (2002) 'The catalytic cycle of  $\beta$ -lactam synthetase observed by x-ray crystallographic snapshots', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.232361199.

Moffat, K. (1989) 'Time-resolved macromolecular crystallography.', *Annual review of biophysics and biophysical chemistry*. doi: 10.1146/annurev.bb.18.060189.001521.

Moffat, K. (1998) 'Time-Resolved Crystallography', *Acta Crystallographica Section A: Foundations of Crystallography*. doi: 10.1107/S0108767398010605.

Moffat, K. (2001) 'Time-resolved biochemical crystallography: A mechanistic perspective', *Chemical Reviews*. doi: 10.1021/cr990039q.

Moffat, K., Szebenyi, D. and Bilderback, D. (1984) 'X-ray Laue diffraction from protein crystals', *Science*. doi: 10.1126/science.223.4643.1423.

Nakane, T. *et al.* (2016) 'Data processing pipeline for serial femtosecond crystallography at SACLA', *Journal of Applied Crystallography*. doi: 10.1107/S1600576716005720.

Nango, E. *et al.* (2016) 'A three-dimensional movie of structural changes in bacteriorhodopsin', *Science*, 354(6319), pp. 1552–1557. doi: 10.1126/science.aah3497.

Nave, C. (1995) 'Radiation damage in protein crystallography', *Radiation Physics and Chemistry*. doi: 10.1016/0969-806X(95)92800-E.

Neutze, R. *et al.* (2000) 'Potential for biomolecular imaging with femtosecond X-ray pulses',



*Nature*. doi: 10.1038/35021099.

Nienhaus, K. *et al.* (2005) 'Ligand migration and protein fluctuations in myoglobin mutant L29W', *Biochemistry*. doi: 10.1021/bi047513t.

Olmos, J. L. *et al.* (2018) 'Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography', *BMC Biology*, 16(1). doi: 10.1186/s12915-018-0524-5.

Onuchic, J. N. *et al.* (1996) 'Protein folding funnels: The nature of the transition state ensemble', *Folding and Design*. doi: 10.1016/S1359-0278(96)00060-0.

Orville, A. M. (2018) 'Entering an era of dynamic structural biology.', *BMC Biology*. doi: 10.1186/s12915-018-0533-4.

Owen, R. L., Rudiño-Piñera, E. and Garman, E. F. (2006) 'Experimental determination of the radiation dose limit for cryocooled protein crystals', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0600973103.

Page, M. G. P. (2004) 'Cephalosporins in clinical development', *Expert Opinion on Investigational Drugs*. doi: 10.1517/13543784.13.8.973.

Palzkill, T. (2013) 'Metallo- $\beta$ -lactamase structure and function', *Annals of the New York Academy of Sciences*. doi: 10.1111/j.1749-6632.2012.06796.x.

Pande, K. *et al.* (2016) 'Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein', *Science*. doi: 10.1126/science.aad5081.

Pandey, S. *et al.* (2019) 'Time-resolved serial femtosecond crystallography at the European XFEL', *Nature Methods*. doi: 10.1038/s41592-019-0628-z.

Pandey, S., Poudyal, I. and Malla, T. N. (2020) 'Pump-probe time-resolved serial femtosecond crystallography at x-ray free electron lasers', *Crystals*. doi: 10.3390/cryst10070628.

Parak, F. and Frauenfelder, H. (1993) 'Protein dynamics', *Physica A: Statistical Mechanics and its Applications*. doi: 10.1016/0378-4371(93)90431-3.

Parak, F. G. *et al.* (2006) 'Protein dynamics on different timescales', *Journal of Non-Crystalline Solids*. doi: 10.1016/j.jnoncrysol.2006.01.106.

Parak, F., Knapp, E. W. and Kucheida, D. (1982) 'Protein dynamics. Mössbauer spectroscopy on deoxymyoglobin crystals', *Journal of Molecular Biology*. doi: 10.1016/0022-2836(82)90285-6.

Pellequer, J. L. *et al.* (1998) 'Photoactive yellow protein: A structural prototype for the three-dimensional fold of the PAS domain superfamily', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.95.11.5884.

Perez-Llarena, F. and Bou, G. (2009) 'β-Lactamase Inhibitors: The Story so Far', *Current Medicinal Chemistry*. doi: 10.2174/092986709789104957.

Perutz, M. F. *et al.* (1960) 'Structure of Hæmoglobin: A three-dimensional fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis', *Nature*, pp. 416–422. doi: 10.1038/185416a0.

Pothineni, S. B. *et al.* (2014) 'Tightly integrated single- and multi-crystal data collection strategy calculation and parallelized data processing in JBLulce beamline control system', *Journal of Applied Crystallography*. doi: 10.1107/S1600576714022730.

Poudyal, I., Schmidt, M. and Schwander, P. (2020) 'Single-particle imaging by x-ray free-electron lasers - How many snapshots are needed?', *Structural Dynamics*. doi: 10.1063/1.5144516.

Ravelli, R. B. and Garman, E. F. (2006) 'Radiation damage in macromolecular cryocrystallography', *Current Opinion in Structural Biology*. doi: 10.1016/j.sbi.2006.08.001.

Redecke, L. *et al.* (2013) 'Natively inhibited trypanosoma brucei cathepsin B structure determined by using an x-ray laser', *Science*. doi: 10.1126/science.1229663.

- Ren, Z. *et al.* (1999) 'Laue crystallography: Coming of age', *Journal of Synchrotron Radiation*. doi: 10.1107/S0909049599006366.
- Ren, Z. *et al.* (2001) '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*. doi: 10.1021/bi0107142.
- Ren, Z. and Moffat, K. (1994) 'Laue Crystallography for Studying Rapid Reactions', *Journal of Synchrotron Radiation*. doi: 10.1107/s0909049594006692.
- Rockwell, N. C. *et al.* (2009) 'Distinct classes of red/far-red photochemistry within the phytochrome superfamily', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0902370106.
- Roedig, P. *et al.* (2016) 'Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering', *Journal of Applied Crystallography*. doi: 10.1107/S1600576716006348.
- Röntgen, W. C. (1898) 'Ueber eine neue Art von Strahlen', *Annalen der Physik*, 300(1), pp. 12–17. doi: 10.1002/andp.18983000103.
- Rossmann, M. G. and Van Beek, C. G. (1999) 'Data processing', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444999008379.
- Salemme, F. R. (1972) 'A free interface diffusion technique for the crystallization of proteins for X-ray crystallography', *Archives of Biochemistry and Biophysics*. doi: 10.1016/0003-9861(72)90530-9.
- Schlichting, I. *et al.* (1990) 'Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis', *Nature*. doi: 10.1038/345309a0.

- Schmidt, M. *et al.* (2003) 'Application of singular value decomposition to the analysis of time-resolved macromolecular x-ray data', *Biophysical Journal*. doi: 10.1016/S0006-3495(03)75018-8.
- Schmidt, M. *et al.* (2004) 'Protein kinetics: Structures of intermediates and reaction mechanism from time-resolved x-ray data', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0305983101.
- Schmidt, M., Nienhaus, K., *et al.* (2005) 'Ligand migration pathway and protein dynamics in myoglobin: A time-resolved crystallographic study on L29W MbCO', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0504932102.
- Schmidt, M., Ihee, H., *et al.* (2005) 'Protein-ligand interaction probed by time-resolved crystallography.', *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1385/1-59259-912-5:115.
- Schmidt, M. *et al.* (2007) 'Structural basis for the photochemistry of  $\alpha$ -phycoerythrocyanin', *Biochemistry*. doi: 10.1021/bi061844j.
- Schmidt, M. *et al.* (2009) 'Protein dynamics of a  $\beta$ -sheet protein', *European Biophysics Journal*. doi: 10.1007/s00249-009-0427-z.
- Schmidt, M. *et al.* (2010) 'Five-dimensional crystallography', *Acta Crystallographica Section A: Foundations of Crystallography*. doi: 10.1107/S0108767309054166.
- Schmidt, M. (2013) 'Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography', *Advances in Condensed Matter Physics*. doi: 10.1155/2013/167276.
- Schmidt, M. *et al.* (2013) 'Protein energy landscapes determined by five-dimensional crystallography', *Acta Crystallographica Section D: Biological Crystallography*. doi:

10.1107/S0907444913025997.

Schmidt, M. (2019) 'Time-resolved macromolecular crystallography at pulsed X-ray sources', *International Journal of Molecular Sciences*. doi: 10.3390/ijms20061401.

Schmidt, M. (2020) 'Reaction initiation in enzyme crystals by diffusion of substrate', *Crystals*. doi: 10.3390/cryst10020116.

Solem, J. C. (1986) 'Imaging biological specimens with high-intensity soft x rays', *Journal of the Optical Society of America B*. doi: 10.1364/josab.3.001551.

Song, C. *et al.* (2011) 'Two ground state isoforms and a chromophore D-ring photoflip triggering extensive intramolecular changes in a canonical phytochrome', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1013377108.

Song, C. *et al.* (2018) '3D structures of plant phytochrome a as Pr and Pfr from solid-state NMR: Implications for molecular function', *Frontiers in Plant Science*. doi: 10.3389/fpls.2018.00498.

Steinfeld, J. I., Francisco, J. S. and Hase, W. L. (1998) *Chemical Kinetics and Dynamics*.

Suga, M. *et al.* (2017) 'Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL', *Nature*. doi: 10.1038/nature21400.

Sugahara, M. *et al.* (2017) 'Hydroxyethyl cellulose matrix applied to serial crystallography', *Scientific Reports*. doi: 10.1038/s41598-017-00761-0.

Takala, H. *et al.* (2014) 'Signal amplification and transduction in phytochrome photosensors', *Nature*. doi: 10.1038/nature13310.

Takala, H. *et al.* (2016) 'Light-induced structural changes in a monomeric bacteriophytochrome', *Structural Dynamics*. doi: 10.1063/1.4961911.

Taylor, B. L. and Zhulin, I. B. (1999) 'PAS domains: internal sensors of oxygen, redox potential,

and light.', *Microbiology and molecular biology reviews : MMBR*.

Taylor, G. (2003) 'The phase problem', in *Acta Crystallographica - Section D Biological Crystallography*. doi: 10.1107/S0907444903017815.

Tenboer, J. *et al.* (2014) 'Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein', *Science*. doi: 10.1126/science.1259357.

Terwilliger, T. C. and Berendzen, J. (1996) 'Bayesian difference refinement', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444996006725.

Tremblay, L. W., Fan, F. and Blanchard, J. S. (2010) 'Biochemical and structural characterization of mycobacterium tuberculosis  $\beta$ -lactamase with the carbapenems ertapenem and doripenem', *Biochemistry*. doi: 10.1021/bi100232q.

Tripathi, S. *et al.* (2012) 'PH dependence of the photoactive yellow protein photocycle investigated by time-resolved crystallography', *Biophysical Journal*. doi: 10.1016/j.bpj.2011.11.4021.

Tschentscher, T. *et al.* (2017) 'Photon Beam Transport and Scientific Instruments at the European XFEL', *Applied Sciences*. doi: 10.3390/app7060592.

Ursby, T. and Bourgeois, D. (1997) 'Improved Estimation of Structure-Factor Difference Amplitudes from Poorly Accurate Data', *Acta Crystallographica Section A: Foundations of Crystallography*. doi: 10.1107/S0108767397004522.

Weierstall, U., Spence, J. C. H. and Doak, R. B. (2012) 'Injector for scattering measurements on fully solvated biospecies', *Review of Scientific Instruments*. doi: 10.1063/1.3693040.

White, T. A. *et al.* (2012) 'CrystFEL: A software suite for snapshot serial crystallography', *Journal of Applied Crystallography*. doi: 10.1107/S0021889812002312.

Wiedorn, M. O. *et al.* (2018) 'Megahertz serial crystallography', *Nature Communications*, 9(1), p. 4025. doi: 10.1038/s41467-018-06156-7.

Wilson, A. J. C. (1942) 'Determination of absolute from relative X-ray intensity data [3]', *Nature*. doi: 10.1038/150152a0.

Winick, H. *et al.* (1981) 'Wiggler and undulator magnets', *Physics Today*. doi: 10.1063/1.2914568.

Winn, M. D. *et al.* (2011) 'Overview of the CCP4 suite and current developments', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444910045749.

Woitowich, N. C. *et al.* (2018) 'Structural basis for light control of cell development revealed by crystal structures of a myxobacterial phytochrome', *IUCrJ*, 5, pp. 619–634. doi: 10.1107/S2052252518010631.

Wolynes, P. G., Luthey-Schulten, Z. and Onuchic, J. N. (1996) 'Fast-folding experiments and the topography of protein folding energy landscapes', *Chemistry and Biology*. doi: 10.1016/S1074-5521(96)90090-3.

Yamashita, A. *et al.* (2003) 'Capturing enzyme structure prior to reaction initiation: Tropinone reductase-II-substrate complexes', *Biochemistry*. doi: 10.1021/bi0272712.

Young, I. D. *et al.* (2016) 'Structure of photosystem II and substrate binding at room temperature', *Nature*. doi: 10.1038/nature20161.

Zeng, X. and Lin, J. (2013) 'Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria', *Frontiers in Microbiology*. doi: 10.3389/fmicb.2013.00128.

# Curriculum Vitae

## Suraj Pandey

Place of Birth: Sindhupalchok, Nepal

### Education

University of Wisconsin – Milwaukee (2014 -2020)

**Doctor of Philosophy (Ph.D.), Physics**

Dissertation: Pump-probe and mix and inject experiments at X-ray free electron lasers

Patan Multiple Campus, Tribhuvan University, Nepal (2009-2013)

**Master's Degree, Physics**

Tri-Chandra College, Tribhuvan University, Nepal (2005-2009)

**Bachelor's Degree, Physics**

### Awards and Fellowships

1. **R1 Distinguished Graduate Fellowship Award, 2020**
2. **Graduate Student Excellence Fellowship Award, 2018**
3. **Physics Department Research Award with Lichtman Research Fund, 2018**
4. **Research Excellence Award, 2018 Spring, 2017 Spring, 2016 Spring, 2015 Fall and Spring, 2014 Fall)**
5. **Chancellor's Graduate Student Award, 2020 Fall, 2019 Fall and Spring, 2018 Fall and Spring, 2017 Fall and Spring, 2016 Fall and Spring, 2015 Fall and Spring, 2014 Fall**



## Publications

1. “Pump-probe time resolved serial femtosecond crystallography at X-ray Free Electron Lasers”, **S. Pandey** et al., Crystals 10, 628 (2020)
2. “The primary structural photoresponse of phytochrome proteins captured by a femtosecond X-ray laser”, E. Claesson, W. Y. Wahlgren, H. Takala, **S. Pandey** et al., Elife 9, e53514 (2020)
3. “Time Resolved Serial Femtosecond Crystallography at XFEL”, **S. Pandey** et al. Nat Methods 17,73-78 (2020)
4. “High-resolution crystal structures of a Myxobacterial phytochrome at cryo and room temperature”, J.C. Sanchez, M. Carrillo, **S. Pandey**, et al. Structural Dynamics 6, 2019
5. “Megahertz Serial Crystallography”, M.O. Wiedorn et al. (Including **S. Pandey**) Nature Communications 9, 4025 (2018)
6. “Enzyme Intermediates Captured ‘on-the-fly’ by Mix-and-Inject Serial Crystallography”, J.L. Olmos<sup>#</sup>, **S. Pandey**<sup>#</sup>, et al. BMC Biology 16:59 (2018) [**#Equally Contributed**]
7. “Structural Basis for Light Control of Cell Development Revealed by Crystal Structures of a Myxobacterial Phytochrome”, N. C. Woitowich, et al. (including **S. Pandey**) IUCrJ 5, 619-634 (2018)
8. “Structural Enzymology Using X-ray Free Electron Lasers”, C. Kupitz, J. L. Olmos, M. Holl, L. Tremblay, K. Pande, **S. Pandey** et al. Structural Dynamics 4, 044003 (2017)
9. “The Room Temperature Crystal Structure of a Bacterial Phytochrome Determined by Serial Femtosecond Crystallography”, P. Edlund, et al. (including **S. Pandey**), Scientific Reports 6, 35279 (2016)

## Talks and Poster Presentations

1. “Time-resolved Serial Femtosecond Crystallography at EuXFEL”, at the NSF sponsored BioXFEL STC 7<sup>th</sup> Annual International Conference, 28-30 January 2020; Puerto Rico [**invited TALK**]

2. "Time-resolved Serial Femtosecond Crystallography", at GRADs meeting, University of Wisconsin at Milwaukee; 11/01/2019 **[TALK]**
3. "Time-resolved Serial Femtosecond Crystallography using MHz X-ray Pulse Rates at the European XFEL", at the Nature Conference titled "Functional dynamics – visualizing molecules in action"; 6-8 November 2019; Tempe, AZ **[POSTER]**
4. "Structural basis for light control of cell development revealed by crystal structures of a myxobacterial Phytochrome", at the NSF sponsored BioXFEL STC 6<sup>th</sup> Annual International Conference; 11-15 February 2019; San Diego, CA **[POSTER]**
5. "Capturing Enzymatic Intermediates 'on-the-fly' via Mix-and-Inject Serial Crystallography", at the NSF sponsored BioXFEL STC 5<sup>th</sup> Annual Site Visit; 15-17 May 2018; Buffalo, NY **[POSTER]**
6. "Capturing Enzymatic Intermediates 'on-the-fly' via Mix-and-Inject Serial Crystallography", at the NSF sponsored BioXFEL STC 5<sup>th</sup> Annual International Conference; 13-16 February 2018; New Orleans, LA (Co-presented with J.L. Olmos of Rice University) **[POSTER]**
7. "Time-resolved Serial Crystallography", at GRADs meeting, University of Wisconsin at Milwaukee; 12/15/2017 **[TALK]**
8. "X-ray Free Electron Laser (XFEL) in protein Crystallography", at the NSF sponsored BioXFEL STC 4<sup>th</sup> Annual International Conference; 10-12 January 2017; Las Vegas, NV (Co-presented with Ishwor Poudyal of UW-Milwaukee) **[POSTER]**