

December 2020

# Macromolecular Structure Determination at X-Ray Free Electron Lasers from Single-Particle Imaging to Time-Resolved X-Ray Crystallography

Ishwor Poudyal  
University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>



Part of the [Biophysics Commons](#), and the [Physics Commons](#)

---

## Recommended Citation

Poudyal, Ishwor, "Macromolecular Structure Determination at X-Ray Free Electron Lasers from Single-Particle Imaging to Time-Resolved X-Ray Crystallography" (2020). *Theses and Dissertations*. 2580.  
<https://dc.uwm.edu/etd/2580>

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact [open-access@uwm.edu](mailto:open-access@uwm.edu).

MACROMOLECULAR STRUCTURE DETERMINATION AT  
X-RAY FREE ELECTRON LASERS FROM  
SINGLE-PARTICLE IMAGING TO TIME-RESOLVED  
X-RAY CRYSTALLOGRAPHY

by

Ishwor Poudyal

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

## MACROMOLECULAR STRUCTURE DETERMINATION AT X-RAY FREE ELECTRON LASERS FROM SINGLE-PARTICLE IMAGING TO TIME-RESOLVED X-RAY CRYSTALLOGRAPHY

by

Ishwor Poudyal

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

X-ray free-electron lasers (XFELs) open the possibility of obtaining diffraction information from a single biological macromolecule. This is because XFELs can generate extremely intense X-ray pulses which are so short that diffraction data can be collected before the sample is destroyed. By collecting a sufficient number of single-particle diffraction patterns from many tilts of a molecule relative to the X-ray beam, the three-dimensional electron density can be reconstructed ab-initio. The resolution and therefore the information content of the data will ultimately depend largely on the number of patterns collected at the experiment. We estimate the number of diffraction patterns required to reconstruct the electron density at a targeted spatial resolution. This estimate is verified by simulations for realistic X-ray fluences, repetition rates, and experimental conditions available at modern XFELs. Employing the bacterial phytochrome as a model system we demonstrate that sub-nanometer resolution is within reach.

Gold nanoparticles (AuNPs) and their conjugation to biological samples have numerous potential applications. Attaching biological molecules to highly scattering AuNPs may provide a powerful way to identify and resolve single molecules. Here, we have shown the weakly scattering part, the thiol coating can be resolved in the presence of a strong scatterer gold from a thiol-decorated gold nanoparticle using the single-particle imaging (SPI) method. One of the scientific cases for building XFELs is the structure determination of single macromolecules and macromolecular complexes at atomic resolution. We propose a promising route of obtaining a sub-nanometer resolution to visualize the atomic details of biological macromolecules using SPI with XFELs.

Biological processes are highly dynamic. Static structures may provide limited insight into protein function, but they offer no information on protein dynamics. For understanding how proteins function, one must investigate structural changes as they happen. The advent of the first XFELs has enabled time-resolved serial femtosecond crystallography (TR-SFX) experiments. The reaction is initiated by light excitation or by the rapid mixing of the microcrystals of functionally active biomolecules. The structural changes are probed after a certain time-delay by the *fs* X-ray pulses. The European XFEL (EuXFEL), with its unique pulse structure, opens a new era for TR-SFX in the study of the dynamics of biological systems. At EuXFEL, X-rays arrive in pulse trains at 10 Hz. When the design specifications are reached, it can produce up to 2,700 pulses with up to 4.5 MHz pulse repetition in a train. At 4.5 MHz, each pulse train is 600  $\mu$ s long with nearly 99.4 ms gaps between the trains. The ultrashort pulse length of the EuXFEL gives the unique possibility to study the dynamics of very short-lived complexes and transient states of dynamic molecules.

We have conducted a first time-resolved experiment using PYP as a model system with a complex MHz X-ray pulse structure at the EuXFEL. This successful time-resolved experiment opens the door for future time-resolved experiments at the EuXFEL. With the knowledge and experience gained from TR-SFX experiments, the possibility of a similar experiment with time-resolved SPI in the future is also discussed.

© Copyright by Ishwor Poudyal, 2020  
All Rights Reserved

# TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>ii</b>
<b>LIST OF FIGURES.....</b>	<b>ix</b>
<b>LIST OF TABLES.....</b>	<b>xii</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>xiii</b>
<b>1. Introduction .....</b>	<b>1</b>
1.1 Structure Determination of Biomolecules .....	1
1.1.1 Historical Perspectives .....	1
1.1.2 X-ray Crystallography .....	3
1.1.3 Coherent Lensless X-ray Imaging .....	5
1.2 From Single-particle Structure Determination to Time-resolved X-ray Crystallography...6	
1.3 Model Systems .....	8
1.3.1 Phytochrome.....	8
1.3.2 Thiol-decorated Gold Nanoparticles .....	10
1.3.3 Photoactive Yellow Protein (PYP).....	11
<b>2. Materials and Methods .....</b>	<b>13</b>
2.1 Free-electron Laser (FEL) Radiation.....	13
2.2 Radiation Damage and Femtosecond X-ray Pulses.....	13
2.3 Interaction of X-rays with Matter .....	14
2.3.1 Scattering by Atoms .....	15
2.3.2 Diffraction from a Molecule.....	16
2.3.3 Diffraction from a Crystal .....	17
2.3.4 Laue and Bragg Diffraction Conditions .....	19
2.3.5 Reciprocal Lattice and Ewald Construction .....	20
2.3.6 Debye-Waller Factor .....	22

2.4	Diffraction Patterns .....	24
2.4.1	Mapping the Detector onto the Ewald Sphere.....	25
2.4.2	Difference Between Single-Particle and Crystal Diffraction Patterns .....	26
2.4.3	Simulation of Diffraction Patterns.....	27
2.4.4	Random Orientations Using Quaternions.....	29
2.5	Number of Snapshots .....	31
2.5.1	Background.....	31
2.5.2	Mathematical Formulation .....	32
2.5.3	Estimating the Average Number of Photons .....	36
2.6	Single-particle Reconstruction .....	38
2.6.1	Orientation Determination.....	39
2.6.2	3D Merging .....	40
2.6.3	Phase Retrieval .....	41
2.7	Serial Femtosecond Crystallography.....	51
2.7.2	X-ray Detectors .....	53
2.7.3	Processing of XFEL Data.....	54
2.7.4	Experiments at XFELs .....	61
<b>3.</b>	<b>Results.....</b>	<b>67</b>
3.1	Simulation Results of a Phytochrome Molecule .....	67
3.1.1	Simulation Strategy .....	67
3.1.2	Simulated Diffraction Patterns .....	68
3.1.3	Merged 3D Diffraction Volume .....	72
3.1.4	Phasing and its Validation .....	73
3.1.5	Low- and High-Resolution Approximation.....	76
3.1.6	Effect of Background .....	77
3.1.7	Modeled Pfr Structure .....	79
3.2	Simulation Results of a Thiol Decorated Gold Nanoparticles. ....	81



3.2.1	Simulation Strategy .....	81
3.2.2	Orientation Determination and Iterative Phasing .....	81
3.2.3	Difference Fourier Technique .....	85
3.3	Experimental Results.....	87
3.3.1	Time-Resolved experiment at EuXFEL .....	87
<b>4.</b>	<b>Discussions.....</b>	<b>93</b>
4.1	Future of SPI.....	93
4.2	Reference-enhanced SPI.....	94
4.3	Mapping Conformational Spectrum .....	95
4.4	An Outlook on SFX.....	97
4.5	Photoactive Proteins: <i>cis-trans</i> Isomerizations .....	99
4.6	Optimum Power for Pump-Probe TR-SFX .....	101
4.7	SPI and Crystallographic Data Analysis: A Comparison.....	102
4.8	Future of Time-Resolved Experiments .....	106
<b>5.</b>	<b>References .....</b>	<b>109</b>
<b>6.</b>	<b>Appendices .....</b>	<b>121</b>
Appendix A:	Justification for Calculating Joint Probability .....	121
Appendix B:	Python Code to Estimate the Exact Number of Snapshots.....	121
Appendix C:	Random Phasor Approximation .....	122
	<b>Curriculum Vitae.....</b>	<b>125</b>

# LIST OF FIGURES

Figure 1.1	Length scale of important biological objects.....	1
Figure 1.2	Schematic for an optical microscope and X-ray imaging.....	2
Figure 1.3	Pr and Pfr structures of phytochromes.....	9
Figure 1.4	Thiol monolayer protected Gold nanoparticles.....	11
Figure 1.5	PYP structure and its photocycle.....	12
Figure 2.1	Graphical representation of a scattering vector.....	16
Figure 2.2	Scattering factors for selected atoms.....	17
Figure 2.3	Representation of a unit cell and two-dimensional crystal lattices.....	19
Figure 2.4	The Ewald construction.....	22
Figure 2.5	The atomic form factor of a Carbon atom multiplied by a Debye-Waller factor.	23
Figure 2.6	Mapping of pixels in an area detector to points in reciprocal space.....	24
Figure 2.7	Elastic scattering in forward geometry.....	28
Figure 2.8	Distribution of a uniformly rotation on the unit 2-sphere.....	31
Figure 2.9	Probability $P$ and correction factor $C_f$ .....	35
Figure 2.10	Plot of $nS$ as a function of resolution.....	38
Figure 2.11	Illustration of the importance of phase in far-field diffraction.....	44
Figure 2.12	Block diagram of the error-reduction method.....	47
Figure 2.13	Reconstruction of Einstein image from Fourier moduli data.....	49
Figure 2.14	Schematic representation of an SFX setup.....	52

Figure 2.15	Theoretical estimate of the hit rate using Poisson statistics.....	56
Figure 2.16	Representative of a peak-finding and indexing results from PYP crystals.....	58
Figure 2.17	Plot of 2D central section of the reciprocal lattice, before and after solving the indexing ambiguity.....	61
Figure 2.18	Setup of an MHz TR-SFX experiment at EuXFEL.....	63
Figure 2.19	PYP microcrystals.....	64
Figure 2.20	X-ray pulse train structure and laser excitation.....	66
Figure 3.1	Simulated diffraction patterns of a phytochrome molecule.....	69
Figure 3.2	Noise-free diffraction patterns from multiple particles.....	70
Figure 3.3	Noise-free diffraction patterns of a single-molecule and a crystal.....	71
Figure 3.4	Merged 3D diffraction volume of a phytochrome molecule.....	72
Figure 3.5	Electron density obtained by iterative phasing and its validation by FSC.....	75
Figure 3.6	Average number of photons per pixel as a function of resolution.....	77
Figure 3.7	Effect of different types of backgrounds on the resolution.....	78
Figure 3.8	Density difference of the full-length Pr and proposed full-length Pfr state.....	80
Figure 3.9	Simulated pattern from a thiol-decorated gold nanoparticle.....	83
Figure 3.10	Merged 3D diffraction volume of thiol-decorated gold nanoparticle.....	84
Figure 3.11	Electron density and validation of resolution using FSC.....	84
Figure 3.12	Reconstructed electron density of a thiol decorated gold nanoparticle using the phases of the gold cluster only.....	86
Figure 3.13	TR-SFX experiment conducted at the EuXFEL.....	90
Figure 3.14	Hit and indexing rates.....	91

Figure 3.15	DED in the chromophore pocket of PYP at different points.....	92
Figure 4.1	Manifold obtained by embedding 45,000 simulated diffraction patterns.....	97
Figure 4.2	Statistics on the deposited structures on PDB solved using XFELs to date.....	99
Figure 4.3	Indexing rate as a function of the number of Bragg peaks.....	105

# LIST OF TABLES

Table 2.1	Participated experiments at the worldwide XFELs.....	62
Table 3.1	Estimated number of snapshots to reach SNR=1.0.....	71
Table 3.2	Statistics of the control data collected with 1.13 MHz X-ray pulses.....	88
Table 3.3	Statistics of the data collected with 564 kHz X-ray pulses.....	89

# ACKNOWLEDGEMENTS

I want to express my deepest gratitude to my advisor, Professor Marius Schmidt, for his guidance and support throughout my PhD research. He has shown me the value of clear communication and trained me to ask big questions throughout my time in the lab. I am deeply thankful to him for providing me the opportunity to participate in multiple exciting projects. His optimism and trust encouraged me to explore my interests. Without his persistent help, this dissertation would not have been possible. I am also glad he made sure that I graduated in time. Also, I am very grateful and indebted to Professor Peter Schwander for his feedback and support on the single-particle project. His guidance, support, and motivation enabled me to complete that project. His expertise with X-ray diffraction physics and insights into a wide range of topics have benefitted me greatly. I am grateful that Marius and Peter were always patient and encouraging, helping me to work through the difficult times. The time-resolved experiments at XFELs were made possible with the help of collaborators from BioXFEL; I am indebted to them.

I am grateful to my doctoral committee members, Professor Peter Schwander, Distinguished Professor Abbas Ourmazd, Professor Valerica Raicu, and Professor Ionel Popa for taking the time to serve on my committee and providing invaluable comments on my dissertation.

Early on in my graduate studies here at UWM, I received excellent mentorship and guidance from Dr. Jason Tenboer and Dr. Christopher Kupitz, who helped me extensively in the Schmidt lab and making sure I was familiar with the lab techniques, in a short time. I thank my “lab mates,” Suraj Pandey, Tek Narsingh Malla, and Tyler Norwood for their help, support, and encouragement. Suraj

and I worked together on different projects from pump-probe to mix-and-inject time-resolved serial crystallography experiments. Thank you, Suraj, for your hard work and fearlessness. Also, many thanks to my colleagues Netra, Dhruba, Narayan, Manish, and Dammar. You were always available if I needed advice, both work and non-work related, coffee, and lunch break discussion. I also thank my fellow graduate students for their company and cheerfulness I needed at crucial moments in my graduate career.

My gratitude extends to the staff in the UWM Physics department, whose support has been invaluable. I take this opportunity to thank my professors from my high school, BSc, and MSc degrees who helped me shape the way I think and how I see science. I am thankful to my friends from my BSc, and MSc program, many of whom share this journey of grad school in a faraway land, for the long-distance mutual backing and encouragement during the new experience abroad.

I am forever indebted to my parents, brothers, sisters, uncle, and aunt for their unwavering love and encouragement. I dedicate this degree to my family for their unwavering love and faith in my ability to realize this goal.

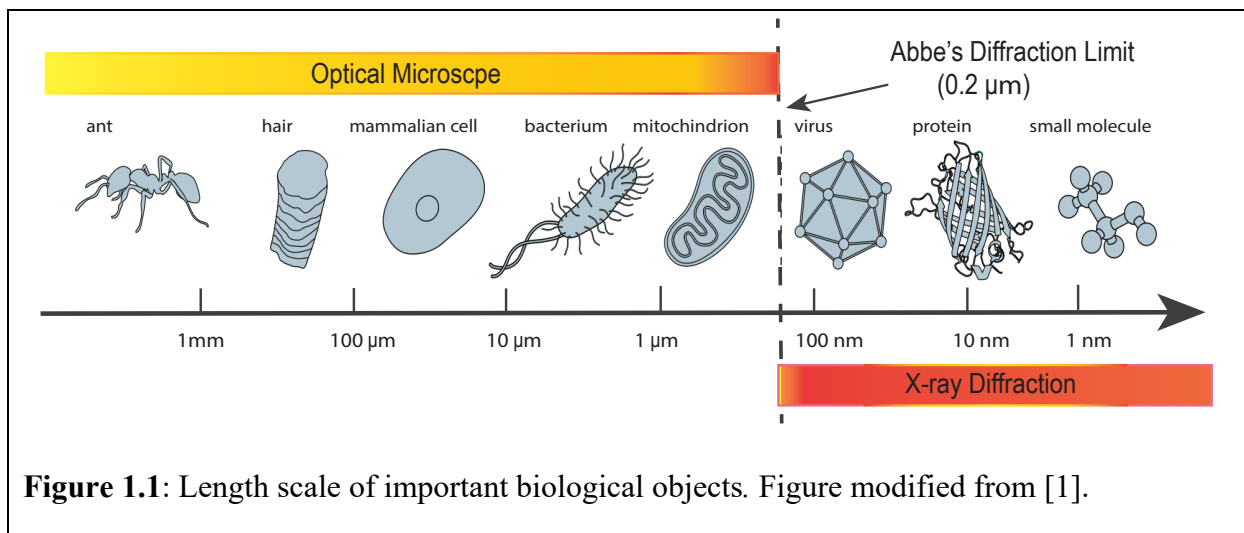
Last but not the least, I am grateful to my loving wife, Sabita, without her inspiration and enthusiasm none of this could have happened. Her contribution to this work cannot be overstated. Being with you is the most beautiful adventure.

# 1. Introduction

## 1.1 Structure Determination of Biomolecules

### 1.1.1 Historical Perspectives

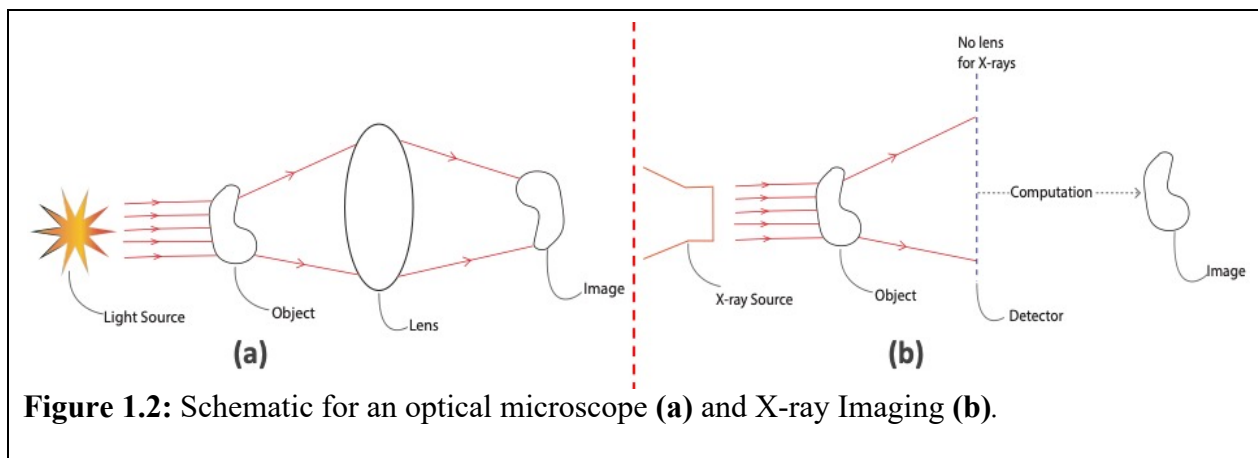
The goal of structural biology is to obtain three-dimensional models of a biological molecule and be able to understand its function. The biological function involves changes in structure; thus, a single static structure is not enough. Optical microscopy is the most widely used technique for structure determination in physical, material, and biological sciences. The invention of the light microscope in the late 16<sup>th</sup> century made it possible to examine parts of the cell invisible to the naked eye. Studying living organisms under the optical microscope opened a new world of microbiology. For a long time, optical microscopy was the only technique to study small biological objects and understand their functions.





In 1873, the microscopist Ernst Abbe came up with an equation related to the microscope's resolution and the wavelength of the light, commonly known as the “diffraction limit” for optical microscopy [2]. This limit states that it is impossible to resolve elements of a structure that are closer to each other than half the wavelength ( $\lambda$ ) in the lateral ( $x, y$ ) plane, independent of the optics being used. Fig. 1.1 shows the length scale of important biological objects with the Abbe diffraction limit at  $200\text{ nm}$  which displays the maximum resolution achievable using optical microscopy (Fig. 1.2). The utility of light microscopy is governed by the use of visible light, which limits the resolution by its wavelength. The shorter the wavelength of the radiation, the better is the resolution. It is necessary to use the light of wavelength smaller than the size of a feature we want to see in an image.

When the structure of a biomolecule is determined, the  $x,y,z$  coordinates of all atoms in the molecule are known. The typical distance between centers of two bonded atoms in a molecule is of the order of  $1.5\text{ \AA}$ . The wavelength of X-rays is short enough to resolve individual atoms of a molecule. X-ray diffraction is used for the past 70 years to determine protein structures and those of other biologically important macromolecules.



**Figure 1.2:** Schematic for an optical microscope (a) and X-ray Imaging (b).

### 1.1.2 X-ray Crystallography

The discovery of X-ray radiation by Wilhelm Röntgen in 1895 opened the door for many important applications in medicinal and scientific research. For this discovery, Rontgen was awarded the Nobel prize in Physics in 1901, the first year this prize was awarded. In 1912, the German physicist Max von Laue and his colleagues showed that a crystal, which consists of periodic arrangements of molecules, functions as a three-dimensional diffraction grating for X-ray radiation. They obtained the first X-ray photograph of a diffraction pattern from a zinc-blende (ZnS) crystal [3]. The next important development was to exploit the information from X-ray photographs to determine the atomic structures of crystalline matter. Sir William Lawrence Bragg, collaborating with his father, Sir William Henry Bragg pioneered this work and determined the atomic structures of crystalline sodium chloride and diamond in 1913. For this work, they were awarded the Nobel prize in Physics in 1915. In 1934, J. Desmond Bernal and Dorothy Crowfoot, for the first time, observed X-ray diffraction from hydrated pepsin protein crystals [4]. They kept the crystals hydrated in a capillary tube sealed at both ends during the X-ray diffraction experiment. After this, crystal structures of important biochemical substances like cholesterol, penicillin, vitamin B12, and insulin were obtained [5]. The discovery of the twisted-ladder structure of DNA, the famous double helix, by Watson and Crick in 1953 marked the milestone in the history of science. The first macromolecular crystal structures, hemoglobin [6] and myoglobin [7], were solved with great effort later in the 1950s and 1960s. These pioneering studies marked what has been called the “big-bang” of structural biology [8]. After these breakthroughs, several other milestones were reached which have jointly shaped the field of X-ray crystallography of biological macromolecules [9].

The impact of X-ray crystallography on the biological sciences has been immense. After more than a century of progression, X-ray crystallography has become a routine method. Out of 160,000 protein structures deposited in the Protein Data Bank (PDB) [10], over 140,000 structures of these are determined by X-ray crystallography, making it the most widely used technique. Another technique like cryo-electron microscopy (Cryo-EM) is also on a rapid move. The number of cryo-EM images uploaded to the Electron Microscopy Data Bank has increased from just 8 in 2002 to more than 9,000 this year. With better electron detectors and image processing techniques along with improvement in the lenses, it makes possible to achieve atomic resolution using cryo-EM [11, 12].

X-ray diffraction is based on the scattering of X-ray photons by atoms. When illuminated with a sufficiently coherent X-ray beam, waves scattered from individual atoms in the molecule will interfere with each other, producing a diffraction pattern on an X-ray sensitive detector placed far away. Many diffraction patterns in different orientations are needed to obtain a structure. Intensities are quantitatively extracted from a diffraction pattern. Since the interaction of X-rays with matter is weak, crystals are used to amplify the intensities in specific directions. Structure factor amplitudes can be directly determined from intensities. However, the phase of the structure factor is not experimentally accessible and needs to be retrieved from additional methods [13]. Once amplitudes and phases are known, electron density maps are calculated and interpreted by a structural model.

The resolution of a crystal structure is largely determined by how well the molecules are ordered in a crystal. The main obstacle to routine high-resolution structure determination in crystallography

has been the “crystallization bottleneck,” the task of producing well-diffracting crystals. Crystals are very often difficult to grow, in particular, if the molecules are large and flexible and admit different conformations. An effective way is to determine the structure of a biomolecule without the need for crystallization. Cryo-EM has been the most widely used technique for the analysis of single particles. To determine the structure, one needs to freeze the samples at cryogenic temperatures. But biological functions are dynamic and are not compatible with cryogenic conditions. Modern X-ray sources with their unparalleled intensity make it possible to image single molecules at an ambient temperature. Single-particle X-ray imaging has emerged as an alternative approach for 3D reconstructions of non-crystalline objects.

### 1.1.3 Coherent Lensless X-ray Imaging

The difficulty in designing good quality lenses for X-ray has led to the development of lens-less imaging, known as Coherent Diffractive Imaging (CDI). This method enables the structure determination of noncrystalline specimens and nanocrystals.

In 1952, David Sayre [14] applied Shannon’s sampling theory [15] to crystal diffraction. He pointed out that if scattering could be detected between Bragg reflections, there is adequate information to solve the phase problem of diffraction. This became the basis for an alternative method for structure determination called coherent lens-less single-particle imaging (SPI) inspired by X-ray crystallography. In this technique, diffraction patterns from thousands of copies of the molecule are recorded and computationally assembled in three dimensions before phasing to reconstruct the electron density. It is also sometimes described as “*crystallography without crystals*” or “*optical microscopy without lenses.*”

## 1.2 From Single-particle Structure Determination to Time-resolved X-ray Crystallography

Understanding protein function requires both the knowledge of the three-dimensional static structure and the conformational changes that take place when the molecule reacts. During a reaction, the atomic coordinates of a molecule are changing. Even in their resting state, at thermal equilibrium, structures fluctuate between numerous conformational sub-states [16–18]. The structural variability is difficult to extract from crystallographic data. Crystallography is an ensemble method as it averages over a large number of molecules in different conformations. There is the danger that the average structure hides the important functional and allosteric molecular mechanisms [19, 20]. The ability to observe one molecule at a time could provide a solution to this. Modern pulsed X-ray sources generate intense X-rays with very short pulses which make it possible to image single biomolecules. Time-resolved SPI could compensate for the information loss. However, SPI experiments have so far only achieved resolutions in the regime of a few tens of nanometers [21, 22], a far cry from the near-atomic resolution required.

This dissertation is partly aimed at changing this view. For this, single-particle reconstruction from smaller biological molecules needs to be tied to time-resolved investigations. For this, lessons from time-resolved crystallography to capture short-lived intermediate states in a biological reaction can be used. This dissertation includes:

1. A mathematical formulation is developed to determine the number of diffraction patterns required to reconstruct a reliable electron-density of targeted resolution at the desired signal-to-noise ratio (SNR) value. The mathematical formulation is tested using simulated

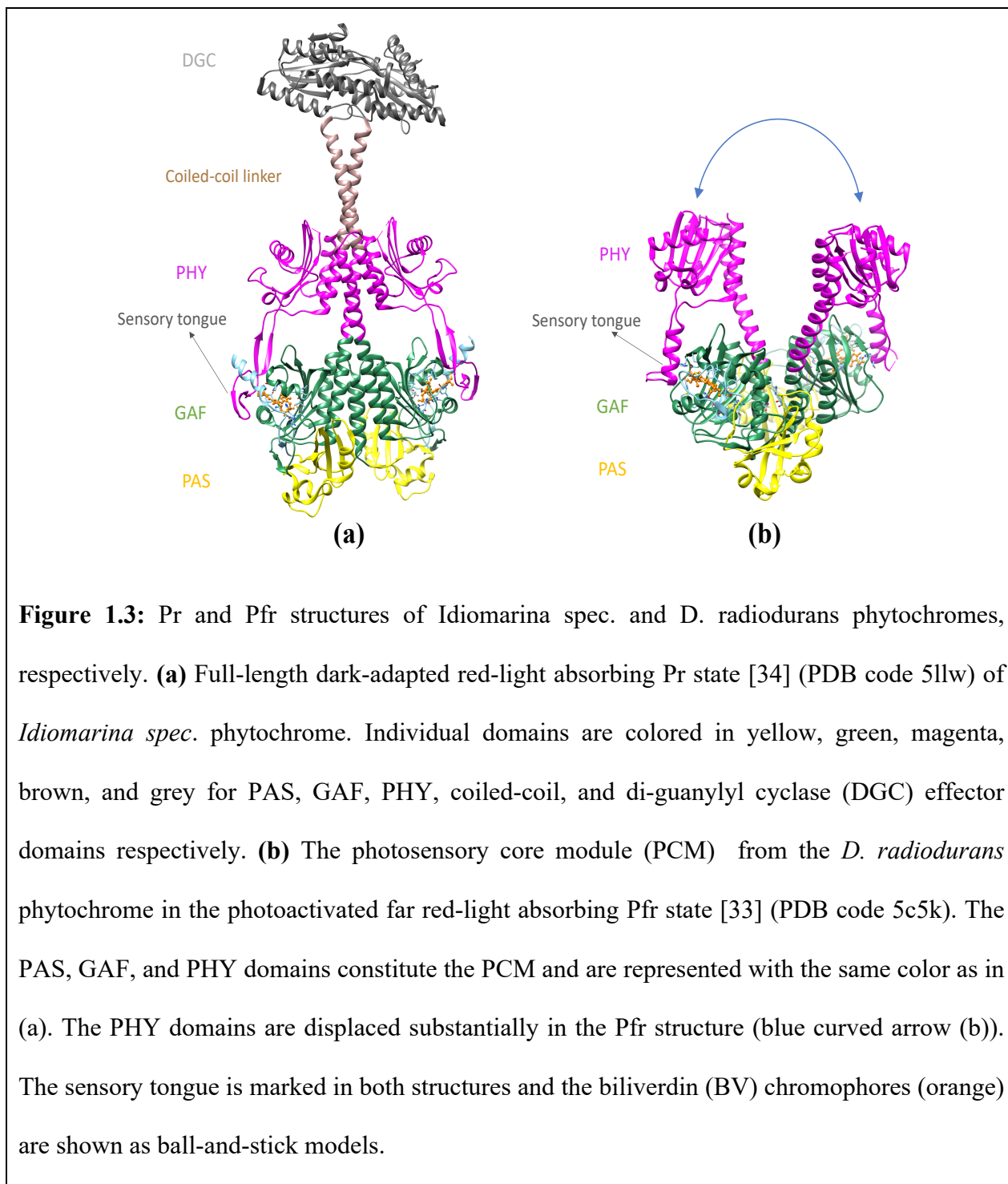
single-particle snapshots of an intact phytochrome molecule in its dark-adapted red-light absorbing (Pr) state.

2. A time-resolved single-particle experiment is simulated using the phytochrome Pr state and the far-red light-absorbing Pfr state (see section 1.3.1 for details of phytochrome structure). The Pfr state of any intact (full length) phytochrome is not known yet. But the crystal structure of a smaller sized phytochrome construct consisting only of PAS, GAF, and PHY domains (see below for a description) is known in the Pfr state that can be used to guess the corresponding full-length phytochrome Pfr structure.
3. So far SPI techniques at X-ray Free-Electron Lasers (XFELs) were applied to large biological assemblies, primarily viruses [21–23]. Whether the application of SPI to smaller particles is possible, is not known. One of the focus of this dissertation is to estimate by simulation how the SPI approach could be applied to smaller biological molecules and eventually reach a near-atomic resolution. For this, thiol-decorated gold nanoparticles are chosen as a model system.
4. A pioneering time-resolved serial crystallography experiment at the European XFEL (EuXFEL) with photoactive yellow protein as a model system was conducted. We hope that ultimately a similar experiment can be conducted with time-resolved SPI in the future. Similarities and differences will be discussed.

## 1.3 Model Systems

### 1.3.1 Phytochrome

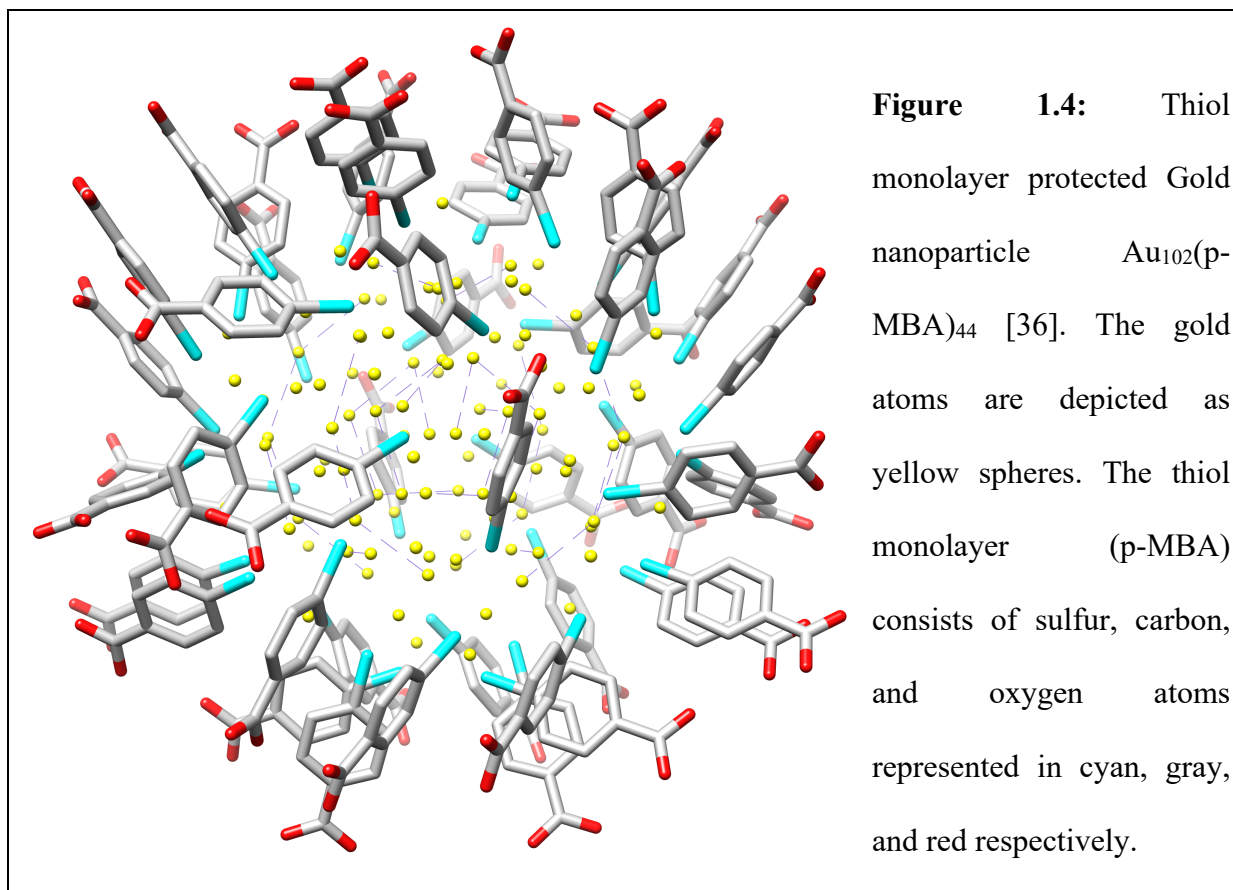
Phytochromes are red-light photoreceptors characterized in plants, fungi, and bacteria [24, 25] and undergo large structural changes after red light absorption. The full-length, functional bacterial phytochromes (BphPs) consist of multiple domains. The PAS (Period ARNT Sim), GAF (cGMP phosphodiesterase/adenylyl cyclase/FhIA), and PHY (phytochrome-specific) domains form the photosensory core module (PCM) [26–30]. An effector domain has enzymatic activity which is covalently linked to the PHY domain. The PHY domain has a tongue-like extension which contacts the GAF domain to seal the biliverdin (BV) chromophore pocket [27, 31] as shown in Fig. 1.3. Upon photoexcitation, phytochromes interconvert between a dark-adapted red-light absorbing state, Pr, and a photoactivated far red-light absorbing state, Pfr. The sensory tongue probes the configuration of the BV chromophore and transmits the signal to the PHY domain. The structure of the tongue undergoes substantial changes between the Pr and Pfr states [32, 33]. In Pr state, the tongue assumes a loop to  $\beta$ -strand conformation, whereas, in the Pfr state, it assumes a loop to  $\alpha$ -helix conformation. Accordingly, large-scale conformational changes with amino acid displacements across several tens of Å between the Pr and Pfr states are required [33–35]. However, the molecular details of the structural changes during the Pr to Pfr transition and their long-range effects on the effector domains are not well understood. Such changes may not be accommodated by the crystal lattice. In contrast, with SPI these structural transitions may be observed, which provides the major motivation to perform the simulations reported further down.





### 1.3.2 Thiol-decorated Gold Nanoparticles

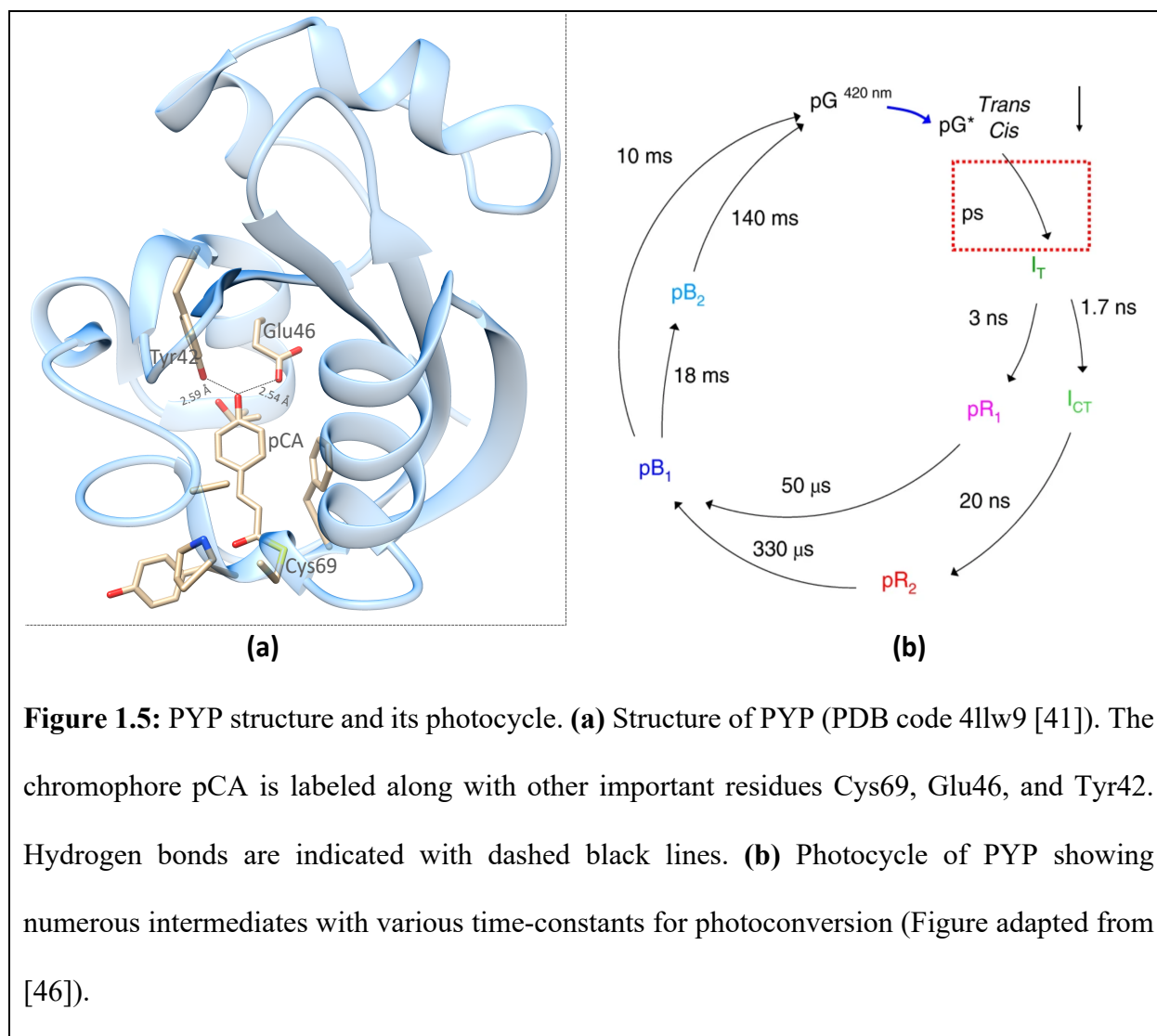
Thiol-decorated gold nanoparticles (TDGN) consist of 102 gold atoms surrounded by 44 p-MBAs (*p*-mercaptobenzoic acid)  $\text{Au}_{102}(\text{p-MBA})_{44}$  as shown in Fig. 1.4. The high-resolution structure of thiol-gold has been obtained by X-ray crystallography [36]. Our goal is to resolve a weakly scattering part, the thiol coating, in the presence of a strong scatterer (gold) using the SPI method. The structure of  $\text{Au}_{102}(\text{p-MBA})_{44}$  is formed by a central symmetric  $\text{Au}_{79}$  gold core with a protective gold-thiolate layer of composition  $\text{Au}_{23}(\text{p-MBA})_{44}$  [36]. The gold-sulfur bonds are very stable which enable the binding of the biomolecules on the surface of gold clusters. The interactions between p-MBA molecules also stabilize the thiol monolayer. The thiol alone consists of 44 p-MBAs and scatters only a few photons. This makes it difficult to retrieve a three-dimensional electron density at low-resolution. Strongly scattering gold nanoparticles could be helpful to identify and orient single-particle snapshots. Attaching a weakly scattering organic molecule to a strongly scattering nanoparticle offers a previously unexplored route of imaging biomolecular systems by the SPI technique.



### 1.3.3 Photoactive Yellow Protein (PYP)

PYP was first discovered in *Halorhodospira halophila* [37], a purple sulfur bacterium. It absorbs blue-light and undergoes photocycle consisting of several short-lived intermediates, namely  $I_T$ ,  $pR_1$ ,  $pR_2$ ,  $pB_1$ , and  $pB_2$  as shown in Fig. 1.5 (a) [38–41]. The time-constants for the photoconversion from one intermediate to another range from femtoseconds to seconds. The photocycle of PYP has been extensively investigated on all time scales [42–44]. The chromophore in PYP is *para*-coumaric acid (pCA) [45], covalently bonded to a Cys69 residue. The pCA head forms two short hydrogen bonds with Glu46 and Tyr42 as shown in Fig. 1.5 (a).

PYP is a small water-soluble protein with only 125 residues and has a molecular weight of 14 kDa. It has been a suitable model system used for the development of numerous data collection techniques. Time-resolved serial femtosecond crystallographic (TR-SFX) experiments (detail about this method is discussed in section 2.6) using PYP as a model system were conducted at the Linac Coherent Light Source (LCLS) facility with ultrafast time resolution [41, 44]. This demonstrates that TR-SFX is able to quantify structural changes between various reaction intermediates. In this dissertation, we investigate the picosecond time range in the photocycle of PYP to pioneer TR-SFX at the EuXFEL.



**Figure 1.5:** PYP structure and its photocycle. **(a)** Structure of PYP (PDB code 4llw9 [41]). The chromophore pCA is labeled along with other important residues Cys69, Glu46, and Tyr42. Hydrogen bonds are indicated with dashed black lines. **(b)** Photocycle of PYP showing numerous intermediates with various time-constants for photoconversion (Figure adapted from [46]).

## 2. Materials and Methods

### 2.1 Free-electron Laser (FEL) Radiation

The FEL concept was introduced by John Madey, in 1971 [47]. The idea of extending radiation from an FEL to shorter wavelengths, in the X-ray regime, were explored [48]. FELs produced X-ray wavelengths, typically ranging from 0.01 to 10 *nm* [48, 49]. XFELs produce spatially and temporally highly coherent X-ray pulses with a duration of tens of femtoseconds and  $10^{12}$ - $10^{13}$  hard X-ray photons per pulse. XFELs produce beams that are more than 10 orders of magnitude brighter than the most powerful synchrotron sources [50]. The intense pulses produced from the XFELs are explained by the principle of self-amplified spontaneous emission (SASE) [47]. SASE emission originates from spontaneous undulator radiation and is a stochastic process. SASE FEL radiation is characterized by shot-to-shot fluctuations in the wavelength spectrum and the FEL pulse energy. The randomness, in the SASE process, becomes imprinted on the X-ray pulses, such that each one of them has a slightly different intensity and different average photon energy. The first hard XFEL, the Linac Coherent Light Source (LCLS) at Stanford Linear Accelerator Center (SLAC), California was commissioned in 2009. Thereafter, five XFELs have come into operation (as of the year 2020) and two more are still under construction.

### 2.2 Radiation Damage and Femtosecond X-ray Pulses

The inelastic cross-section of a molecule for X-rays is very large compared to that of the elastic scattering cross-section [52]. Consequently, a significant amount of energy is deposited in the sample which causes damage, but this process is not instantaneous. Since radiation damage

accumulates more slowly at low temperature, crystals are conventionally held at cryogenic temperature during data collection using synchrotrons [53]. Cryo-cooling diminishes secondary damage induced by reactions of free radicals generated by energetic electrons. However, the primary radiation damage can only be mitigated by a reduction of exposure time or by the attenuation of X-ray beam intensity which ultimately limits the resolution of the collected data.

The idea of flash imaging to overcome the degradation of X-ray images of living cells was first suggested by Solem *et al.* [54]. A detailed molecular dynamics analysis [55] gave the first insight that atomic resolution of a single biological molecule could be possible using femtosecond pulses from an XFEL source. The simulation shows that the radiation damage could be ignored for short pulse duration (such as  $< 30$  fs). Since diffraction is instantaneous and damage requires some time to evolve, femtosecond X-ray pulses from an XFEL produce essentially damage-free diffraction patterns. This is commonly known as the “diffraction-before-destruction” principle [55, 56]. This principle was first experimentally demonstrated using soft X-rays with a pulse duration of 25 fs containing  $10^{12}$  photons [57]. As a result of the high number of photons in the pulse, photoelectric absorption deposits sufficient energy in the sample (a microstructure milled through a silicon nitride membrane) to bring it up to a temperature of 60,000 K and destroy it completely. However, the extremely short pulse duration allows the collection of the data before the sample is blown apart.

### 2.3 Interaction of X-rays with Matter

X-rays interaction with matter is based on the first-order Born approximation, also known as “kinematic” or “single-scattering” approximation [58]. Within the first-order Born

approximation, the diffracted wave for an object illuminated by a plane wave in the far-field is the Fourier transform of the object. This approximation is of fundamental importance to reconstruct the electron-density from intensities extracted from diffraction patterns.

### 2.3.1 Scattering by Atoms

The *atomic scattering factor* can be written as

$$f(\mathbf{q}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{q} \cdot \mathbf{r}) d\mathbf{r}. \quad (2.1)$$

The scattering amplitude  $f$  is given by the Fourier transform of the electron density  $\rho$ . The integration is over the entire space  $\mathbf{r}$  and  $\mathbf{q}$  represents the scattering vector as shown in Fig. 2.1. The *atomic scattering factor* describes the scattering efficiency of a given atom in a given direction. For any atom scattering in the forward direction,  $f$  is normalized to the total number of electrons in an atom. With the increase of the scattering angle, the waves scattered by individual electrons become more and more out of phase which decreases  $f$ . For crystallographic software packages, the atomic scattering factors are calculated from empirical approximations based on *Hartree-Fock* wave functions. For each atom, the atomic form factor is represented by a sum of Gaussians as [13]

$$\begin{aligned} f(q) &= \sum_{i=1}^4 a_i \cdot \exp\left(-\frac{b_i |q|^2}{4}\right) + c \\ &= \sum_{i=1}^4 a_i \cdot \exp\left(-b_i \left(\frac{\sin\theta}{\lambda}\right)^2\right) + c. \end{aligned} \quad (2.2)$$

The parameters,  $a_i$ ,  $b_i$ , and  $c$  are called the *Cromer-Mann coefficients*. They are listed in the International Tables for Crystallography [59]. A plot of atomic form factors as a function of

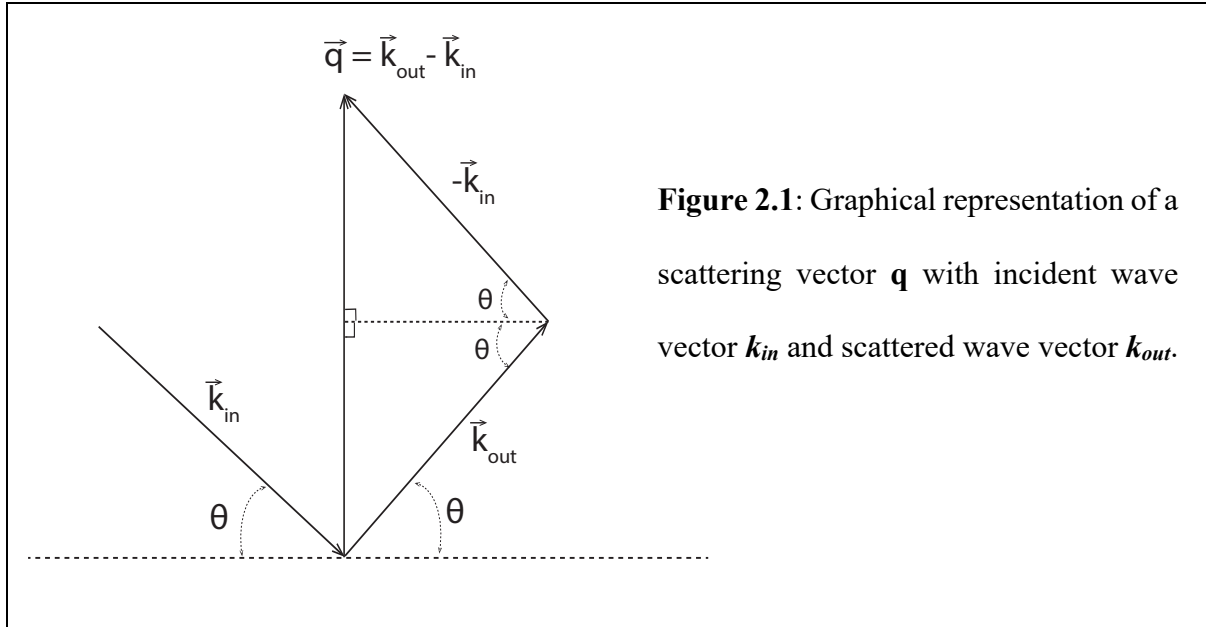
$\sin\theta/\lambda$  is shown in Fig. 2.2.  $\theta$  represents half of the scattering angle i.e. half of the angle between the incident and the scattered wave vector as shown in Fig. 2.1.  $\lambda$  represents the wavelength of the incoming X-ray radiation.

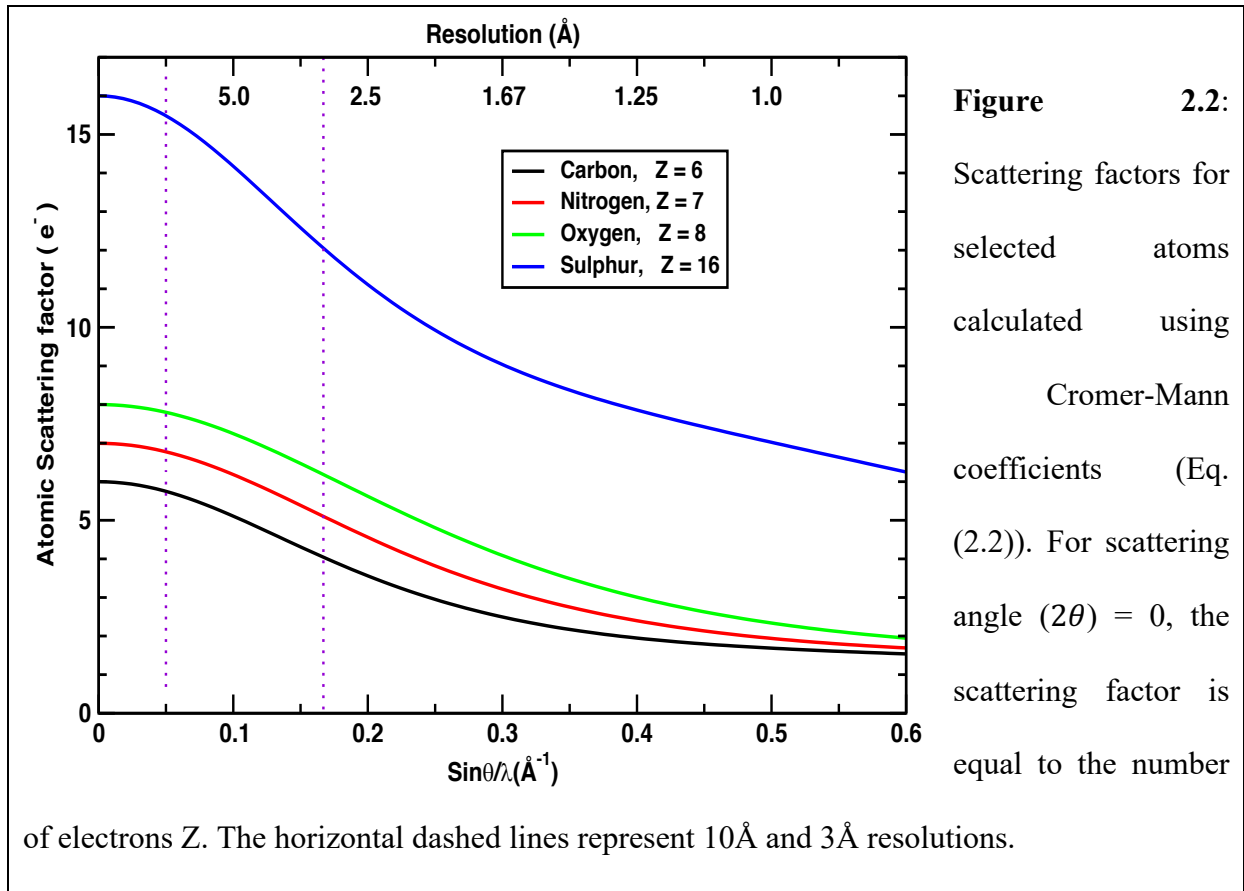
### 2.3.2 Diffraction from a Molecule

Molecules consist of a number of atoms that are bound. Diffraction of X-rays by a molecule is given by the structure factor and calculated as

$$F^M(\mathbf{q}) = \sum_{j=1}^N f_j(\mathbf{q}) \exp(2\pi i \mathbf{q} \cdot \mathbf{r}_j) . \quad (2.3)$$

The sum is taken over all atoms  $N$  at positions  $\mathbf{r}_j$  scattering with their atomic form factor  $f_j$ .





### 2.3.3 Diffraction from a Crystal

A crystal consists of a three-dimensional translationally periodic arrangement of atoms each of which diffract with their atomic form factor. The structure factor of the crystal is the addition of diffraction from all atoms in the crystal with the correct phases given by

$$F^C(\mathbf{q}) = \sum_{m=1}^M f_m(\mathbf{q}) \exp(2\pi i \mathbf{q} \cdot \mathbf{r}_m). \quad (2.4)$$

The sum is over all atoms  $M$  at positions  $\mathbf{r}_m$  in the crystal. An ideal crystal is a periodic repetition of a *unit cell* containing one or more molecules. The unit cell is defined by three vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  which has lengths  $a, b, c$ , and angles between the axes  $\alpha, \beta, \gamma$  as shown in Fig. 2.3 (a). These vectors are called the *crystallographic axes* of the unit cell. The lengths and angles are known as the *lattice*



constants or lattice parameters. The position vector  $\mathbf{r}_m$  in Eq. (2.4) is given by the sum of a lattice vector  $\mathbf{r}_{gn}$  and a position vector in the unit cell  $\mathbf{r}_j$  as shown in Fig. 2.3 (b). The structure factor of the crystal with  $n = 1 \dots N$  unit cells and  $j = 1 \dots J$  atoms in the unit cell can be written as

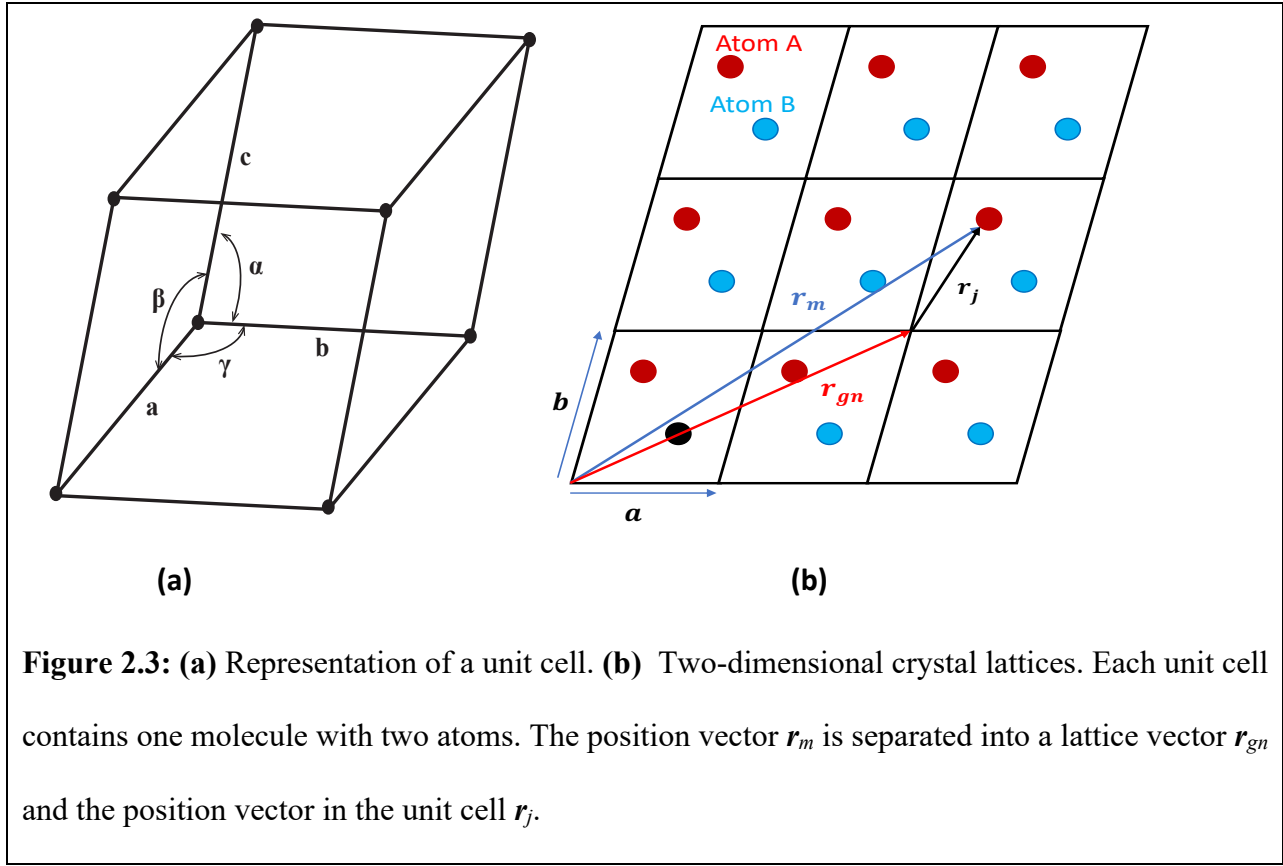
$$\begin{aligned} F^c(\mathbf{q}) &= \sum_{n=1}^N \sum_{j=1}^J f_j \exp(2\pi i \mathbf{q} \cdot (\mathbf{r}_{gn} + \mathbf{r}_j)) \\ &= \sum_{n=1}^N \exp(2\pi i \mathbf{q} \cdot \mathbf{r}_{gn}) \cdot \sum_{j=1}^J f_j \exp(2\pi i \mathbf{q} \cdot \mathbf{r}_j). \end{aligned} \quad (2.5)$$

The first term in Eq. (2.5) is the *lattice factor*,  $G$ , and the second term is the structure factor of the molecule as given in Eq. (2.3).  $\mathbf{r}_{gn}$  is a vector to the origin of each unit cell. It is an integer linear combination of the unit cell vectors written as

$$\mathbf{r}_{gn} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$$

The lattice factor  $G$  can be written as a triple sum of  $u$ ,  $v$ , and  $w$ . For  $U$  unit cells in  $\mathbf{a}$  direction,  $V$  unit cells in  $\mathbf{b}$  direction,  $W$  unit cells in  $\mathbf{c}$  direction such that  $U \cdot V \cdot W = N$ ,  $G$  can be written as

$$\begin{aligned} G &= \sum_{n=1}^N \exp(2\pi i \mathbf{q} \cdot \mathbf{r}_{gn}) \\ &= \sum_{u=0}^{U-1} \exp(2\pi i u \mathbf{q} \cdot \mathbf{a}) \sum_{v=0}^{V-1} \exp(2\pi i v \mathbf{q} \cdot \mathbf{b}) \sum_{w=0}^{W-1} \exp(2\pi i w \mathbf{q} \cdot \mathbf{c}). \end{aligned} \quad (2.6)$$



**Figure 2.3:** (a) Representation of a unit cell. (b) Two-dimensional crystal lattices. Each unit cell contains one molecule with two atoms. The position vector  $r_m$  is separated into a lattice vector  $r_{gn}$  and the position vector in the unit cell  $r_j$ .

### 2.3.4 Laue and Bragg Diffraction Conditions

For large values of  $U$ ,  $V$ , and  $W$ , the sum in Eq. (2.6) is zero for non-integer values of  $\mathbf{q} \cdot \mathbf{a}$ ,  $\mathbf{q} \cdot \mathbf{b}$ , and  $\mathbf{q} \cdot \mathbf{c}$ . But with integer values, the sum is not zero. These are the important conditions in crystallography known as *Laue-conditions* which can be written as

$$\begin{aligned}
 \mathbf{q} \cdot \mathbf{a} &= h \\
 \mathbf{q} \cdot \mathbf{b} &= k \\
 \mathbf{q} \cdot \mathbf{c} &= l
 \end{aligned}
 \tag{2.7}$$

where  $h, k, l$  are integer numbers. The periodicity of the crystal ensures that sets of parallel planes can be drawn passing through the atom centers at regular intervals. These sets of planes are denoted by *Miller indices*  $hkl$ . If one plane is drawn through the unit cell origin, the intercepts of the next plane of the set on the axes are  $a/h, b/k, c/l$ . Suppose  $d$  be the distance from the origin of the reciprocal lattice to the  $hkl$  reciprocal lattice points. Then from Laue-conditions, we can write  $|q| = 1/d$ . The geometrical interpretation is shown in Fig. 2.2. It follows that  $k_{in}$  and  $k_{out}$  make the same angle  $\theta$  with the plane perpendicular to  $q$ .

### 2.3.5 Reciprocal Lattice and Ewald Construction

For any *crystal lattice*, a *reciprocal lattice* exists such that the Laue conditions are automatically fulfilled. It is called the *reciprocal lattice* because many of its properties are reciprocal to those of the real space crystal lattice. For a real space lattice with a unit cell (basis) vectors  $\mathbf{a}, \mathbf{b}$  &  $\mathbf{c}$ , the corresponding reciprocal lattice unit cell (basis) vectors  $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$  are given as

$$\begin{aligned}\mathbf{a}^* &= \frac{1}{V}(\mathbf{b} \times \mathbf{c}), \\ \mathbf{b}^* &= \frac{1}{V}(\mathbf{c} \times \mathbf{a}), \\ \mathbf{c}^* &= \frac{1}{V}(\mathbf{a} \times \mathbf{b}).\end{aligned}\tag{2.8}$$

Here,  $V$  is the volume of the unit cell of the crystal given as  $V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ . The vectors  $\mathbf{a}^*, \mathbf{b}^*$  &  $\mathbf{c}^*$  satisfy the following relationships

$$\begin{aligned}\mathbf{a} \cdot \mathbf{a}^* &= 1; \quad \mathbf{a} \cdot \mathbf{b}^* = 0; \quad \mathbf{a} \cdot \mathbf{c}^* = 0, \\ \mathbf{b} \cdot \mathbf{a}^* &= 0; \quad \mathbf{b} \cdot \mathbf{b}^* = 1; \quad \mathbf{b} \cdot \mathbf{c}^* = 0,\end{aligned}$$

$$\mathbf{c} \cdot \mathbf{a}^* = 0; \quad \mathbf{c} \cdot \mathbf{b}^* = 0; \quad \mathbf{c} \cdot \mathbf{c}^* = 1.$$

The reciprocal lattice is now built up by repeated translations of the reciprocal space basis vectors. A vector  $\mathbf{q}_{hkl}$  drawn from the origin of the reciprocal lattice to any point having coordinates  $hkl$  is perpendicular to the set of planes in the real space crystal lattice with Miller indices  $hkl$ . This vector can be expressed in terms of reciprocal basis vectors as

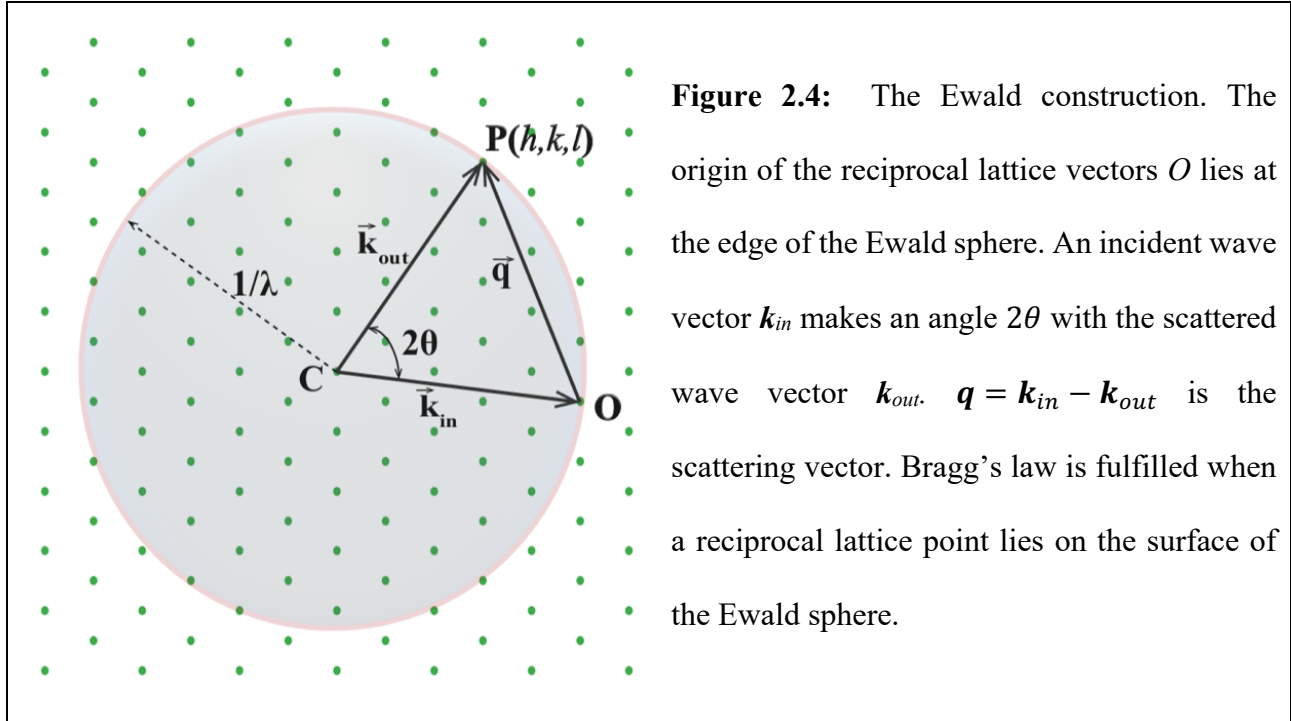
$$\mathbf{q}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*.$$

These are also the solutions of the *Laue-equations* as given in Eq. (2.7). The length of this vector is equal to the reciprocal of the spacing  $d$  of the planes with Miller index  $hkl$

$$q_{hkl} = \frac{1}{d_{hkl}}.$$

Each lattice point is related to a set of planes in the crystal and represents the orientation and spacing of the corresponding set of planes. The usefulness of the reciprocal lattice lies in its connection to the observed diffraction. The diffraction maxima occur only when the Laue equations and elastic scattering conditions are satisfied. The conditions for elastic diffraction can be represented graphically by the *Ewald construction* shown in Fig. 2.4. For elastic diffraction, the vectors  $\hat{\mathbf{k}}_{in}/\lambda$  and  $\hat{\mathbf{k}}_{out}/\lambda$  in the incident and scattered directions have equal magnitudes  $1/\lambda$ .  $\hat{\mathbf{k}}_{in}$  and  $\hat{\mathbf{k}}_{out}$  represent the unit vectors along  $\mathbf{k}_{in}$  and  $\mathbf{k}_{out}$  direction respectively. Since the incident and scattered wave vectors are of fixed length, the scattering vector will cover a sphere in reciprocal space. The sphere will intersect the origin at forward scattering direction. This sphere is called the *Ewald sphere*.

The condition for diffraction from the  $(hkl)$  planes is that the point  $P(h,k,l)$  in the reciprocal lattice touches the surface of the sphere. The direction of the diffracted wave vector  $\mathbf{k}_{out}$  is found by joining  $C$  to  $P$  (Fig. 2.4). When the elastic scattering condition is fulfilled, the vector  $\mathbf{OP}$  is the



scattering vector  $q$ . The magnitude of  $q$  can be written as  $q = (2\sin\theta) / \lambda$ . Combining this with Laue-condition (Eq. 2.7) we can write

$$2d\sin\theta = \lambda. \quad (2.9)$$

Equation (2.9) is a well-known equation in crystallography called *Bragg's law*.

### 2.3.6 Debye-Waller Factor

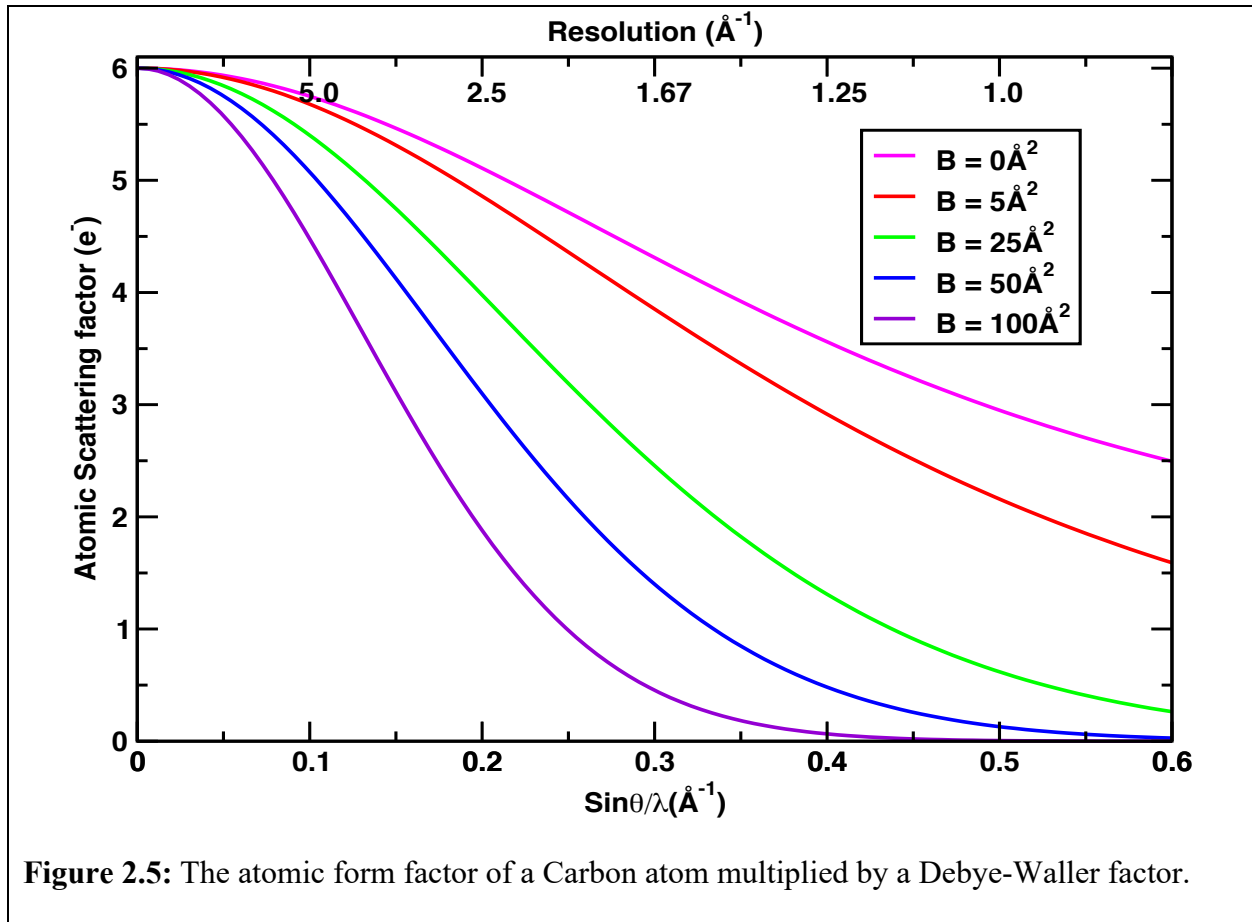
Atoms in a crystal are not located at fixed points in the lattice. These atoms undergo thermal vibration about their mean position. The amplitude of the vibration increases with an increase in temperature. The thermal agitation decreases the intensity of the diffracted beam. This effect can be addressed by introducing a *temperature factor* also called *B-factor*. The B-factor is given by

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

where  $\langle x^2 \rangle$  is the mean square displacement of an atom. As the B-factor increases, the contribution of the atom to the scattering is decreased. To address this effect, the atomic scattering factor (Eq. 2.2) has to be multiplied by the Debye-Waller factor. In a simple case, for isotropic vibration, the Debye-Waller factor can be written as

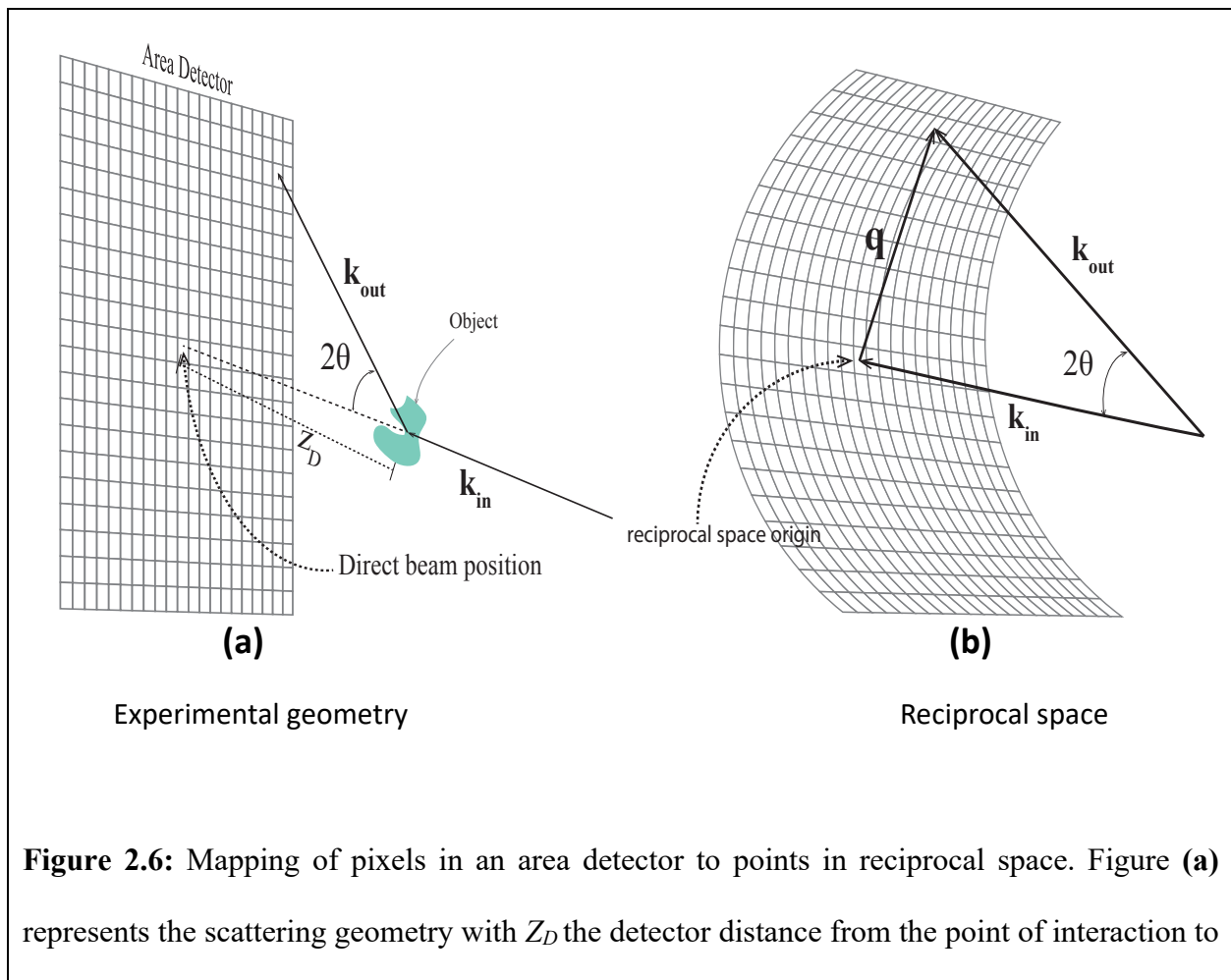
$$T = \exp\left(-B \left(\frac{\sin\theta}{\lambda}\right)^2\right). \quad (2.10)$$

The atomic form factor for carbon is shown for increasing B-factors in Fig. 2.5. The temperature-factor is strongly angular dependent. At high-resolution, for a high B-factor, the atomic form factor approaches zero. Fig. 2.5 shows that atoms with lower B-factors diffract to higher resolutions.



## 2.4 Diffraction Patterns

Only scattering vectors ending on the Ewald sphere contribute to the diffraction pattern of a molecule in a particular orientation. This is true for Bragg reflections as well as for scattering contributions from single particles. The entire 3D diffraction volume has to be recorded to solve the structure in 3D. No matter whether diffraction from crystals or single particles is concerned, a single diffraction pattern records a thin 'slice' through the 3D reciprocal space map of the object as shown in Fig. 2.6. Diffraction patterns are the 2D projections of a 3D object. Full 3D information, therefore, requires many diffraction patterns, each from a different orientation. The geometric relation of a detector pixel to the diffraction space is shown in Fig. 2.6 (b).



the center of the detector. **(b)** Mapping of detector pixels to the diffraction space. The direct beam position always coincides with the origin of the reciprocal space. Figure modified from [60].

### 2.4.1 Mapping the Detector onto the Ewald Sphere

The pixels on the planar detector can be mapped onto the curved Ewald sphere (Fig. 2.6) as described in Eq. (2.11). Consider the incident X-ray beam of wavelength  $\lambda$  traveling along the  $\hat{z}$  (0,0, z) direction. Let a pixel is on the detector whose center is at  $(x, y, z_D)$  in the laboratory frame of reference. Under kinematic approximation, the scattering vector can be written as (see Fig. 2.2 & Fig. 2.6 (b))

$$\mathbf{q} = \mathbf{k}_{out} - \mathbf{k}_{in}$$

The pixel of the detector measures the diffraction intensities of the scatterer. The above equation can be written in the laboratory coordinate system as

$$q_x \hat{x} + q_y \hat{y} + q_z \hat{z} = |\mathbf{k}_{out}| \frac{x\hat{x} + y\hat{y} + z_D\hat{z}}{\sqrt{x^2 + y^2 + z_D^2}} - |\mathbf{k}_{in}| \hat{z}.$$

With this, any pixel  $(x, y, z_D)$  on a planar detector can be mapped onto an Ewald sphere  $(q_x, q_y, q_z)$  as

$$(q_x, q_y, q_z) = \frac{1}{\lambda} \left[ \frac{(x, y, z_D)}{\sqrt{x^2 + y^2 + z_D^2}} - (0, 0, 1) \right]. \quad (2.11)$$

The mapping of the detector pixels onto the Ewald sphere is also shown in Fig. 2.7. From this geometry, the relationship between a pixel  $(x, y)$  on a two-dimensional detector and a scattering vector on the Ewald sphere can be written as

$$q = \frac{2}{\lambda} \sin(\theta)$$



$$q_x = q \left( 1 - \frac{\lambda^2 q^2}{4} \right)^{\frac{1}{2}} \sin(\arctan2(x, y)) \quad (2.12)$$

$$q_y = q \left( 1 - \frac{\lambda^2 q^2}{4} \right)^{\frac{1}{2}} \cos(\arctan2(x, y))$$

$$q_z = -\frac{\lambda}{2} q^2 .$$

#### 2.4.2 Difference Between Single-Particle and Crystal Diffraction Patterns

Without amplification by a periodic lattice, the diffracted intensity from a single molecule is much weaker than that from crystals. At XFELs, the enormous incident intensity compensates for the absence of amplification. Still, a large number of measurements are required for any high-resolution information (discussed in detail in section 2.5). The molecular transforms of single molecules are continuous. With this, diffraction patterns can be sampled finely enough that the unrecorded phase information can be retrieved solely from the set of the measured structure factor amplitudes (discussed in section 2.6). With crystals, strong and discrete diffraction maxima occur only at Bragg angles. The diffracted intensity is sampled in discrete angular intervals only at the reciprocal lattice points. This is sometimes referred to as “Bragg sampling.” Although this facilitates the recording of the crystal’s diffraction pattern, it complicates the process of the reconstruction of the original molecular structure from its diffraction patterns. As intensities are extracted, structure factor amplitudes can only be determined from which electron density cannot be calculated. Additional experiments are necessary to retrieve the phase so that structure factors can be obtained [61]. The advantage is that Bragg peaks can be easily distinguished from slowly varying background in a diffraction pattern. Accurate structure factor amplitudes can be

determined this way. In contrast, if diffraction is continuously varying, the background cannot be readily separated, and a large number of measurements are required to determine accurate amplitudes suitable for solving the structure.

### 2.4.3 Simulation of Diffraction Patterns

For the simulation of single-particle diffraction patterns (Fig. 2.7), we construct a regular two-dimensional Cartesian grid that represents detector pixel positions. The pixel coordinates  $(x,y)$  on the detector are mapped onto the Ewald sphere using Eqs. (2.12). The scattered intensity collected on a detector pixel of solid angle  $d\Omega$  at a scattering vector  $\mathbf{q}$  is given by

$$I(\mathbf{q}) = \phi r_e^2 |\mathbf{F}(\mathbf{q})|^2 d\Omega \quad (2.13)$$

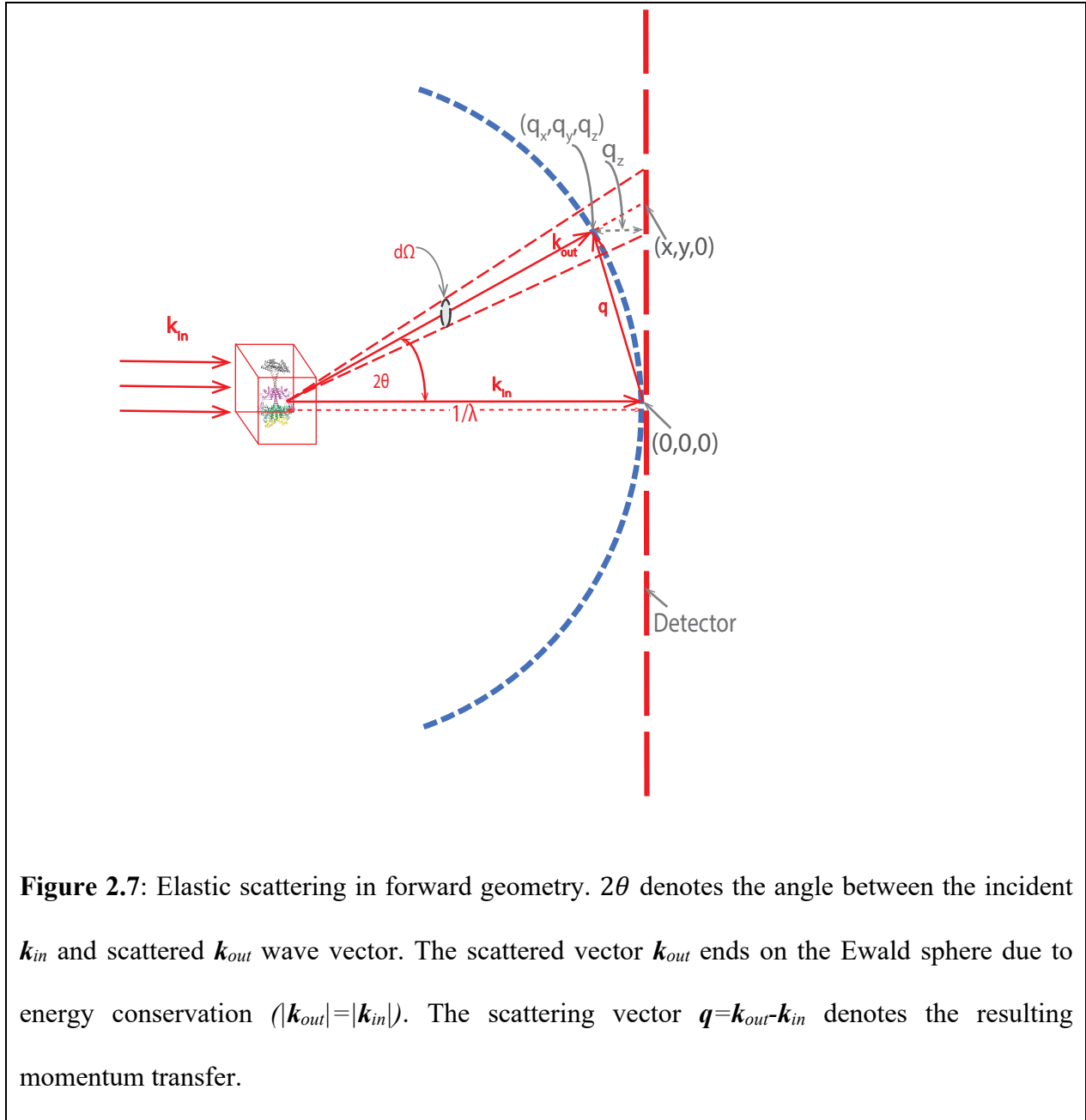
where  $\phi$  is the incident flux ('photons/area'), and ' $r_e$ ' being the classical electron radius.  $\mathbf{F}(\mathbf{q})$  is the structure factor of the molecule given in Eq. (2.3), and  $d\Omega$  the solid angle subtended by the area of the Shannon pixel,

$$d\Omega = \left(\frac{\lambda}{oD}\right)^2 \cos^3(2\theta), \quad (2.14)$$

'o' being the oversampling ratio,  $o \geq 2$  (for details of oversampling see section 2.5.4.3).

With this, the intensity can be written as

$$I(\mathbf{q}) = \Phi r_e^2 |\mathbf{F}(\mathbf{q})|^2 \left(\frac{\lambda}{oD}\right)^2 \cos^3(2\theta).$$



For small scattering angles ( $2\theta$ ), the formula can be approximated by

$$I(\mathbf{q}) = \Phi r_e^2 |\mathbf{F}(\mathbf{q})|^2 \left(\frac{\lambda}{oD}\right)^2.$$

The simulated pattern is obtained by calculating intensity values (Eq. (2.13)) at each pixel of the detector. Randomly oriented diffraction patterns were generated using uniform random rotation quaternions (section 2.4.3).

During an experiment, in addition to single-particle scattering, there will also be scattering from multiple particles. The multiple particles in the beam can be simulated by the coherent or incoherent sum of the scattered waves of each of the particles. For instance, consider two molecules, where one molecule is separated by a distance  $S$  from the other. Let,  $F_1(\mathbf{q})$  and  $F_2(\mathbf{q})$  be the structure factors corresponding to two molecules. For incoherent addition, the intensity of the resultant image is given as

$$I \propto |\mathbf{F}(\mathbf{q})|^2 = |\mathbf{F}_1(\mathbf{q})|^2 + |\mathbf{F}_2(\mathbf{q})|^2.$$

Similarly, for coherent addition, the intensity of the resultant image can be written as

$$I \propto |\mathbf{F}(\mathbf{q})|^2 = |\mathbf{F}_1(\mathbf{q}) + \mathbf{F}_2(\mathbf{q})|^2.$$

After calculating the structure factor of the first molecule  $F_1(\mathbf{q})$ , the structure factor of the second molecule  $F_2(\mathbf{q})$  can be calculated using translation property of the Fourier transform as

$$F_2(\mathbf{q}) = \exp(2\pi i \mathbf{S} \cdot \mathbf{q}) F_1(\mathbf{q}).$$

Here, the second molecule is at a distance  $S$  from the first one. The simulated noise-free single and multiple particles diffraction patterns are shown in Fig. 3.3.

#### 2.4.4 Random Orientations Using Quaternions

A quaternion is a 4-tuple written as  $q_0 + q_1i + q_2j + q_3k$ , where  $q_i$ 's are real numbers and  $i, j, k$  are three imaginary numbers satisfying the following identities:

$$i^2 = j^2 = k^2 = -1,$$

$$ij = k, \quad ji = -k,$$

$$jk = i, \quad kj = -i,$$

$$ki = j, \quad ik = -j.$$

A vector in three-dimensional space can be expressed as a *pure quaternion*, a quaternion with no real part  $Q = 0 + xi + yj + zk$ . A rotation is expressed by a unit quaternion  $Q_R$  i.e. its norm  $|Q_R|$  is equal to 1. For the unit rotation axis  $[u_x, u_y, u_z]$  and angle  $\theta$  the corresponding quaternion is written as  $Q = \cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)u_xi + \sin\left(\frac{\theta}{2}\right)u_yj + \sin\left(\frac{\theta}{2}\right)u_zk$ . Each unit quaternion can be represented as a point in the unit 3-sphere as shown in Fig. 2.8. To generate uniform random oriented diffraction patterns, we use unit quaternion  $\mathbf{Q} = (a + bi + cj + dk)$  as follows [62]:

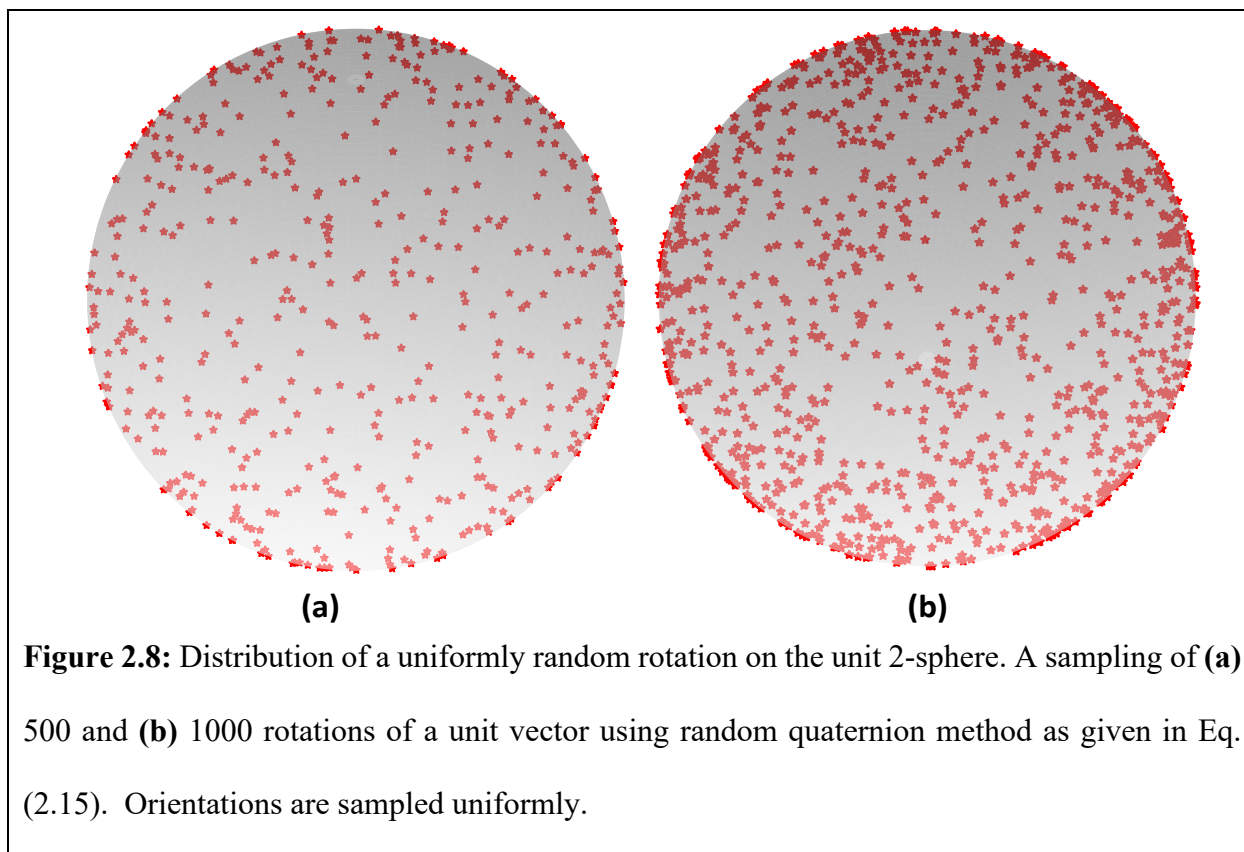
$$\mathbf{Q} = \left( \sqrt{1-u_1} \sin(2\pi u_2), \sqrt{1-u_1} \cos(2\pi u_2), \sqrt{u_1} \sin(2\pi u_3), \sqrt{u_1} \cos(2\pi u_3) \right). \quad (2.15)$$

The three numbers  $u_1, u_2, u_3$  are chosen at random, uniformly distributed in the interval  $[0,1]$ .

The quaternions can be expressed as a rotation matrix [63]:

$$\mathbf{R} = \begin{bmatrix} a^2 + b^2 - c^2 - d^2 & 2bc - 2ad & 2bd + 2ac \\ 2bc + 2ad & a^2 - b^2 + c^2 - d^2 & 2cd - 2ab \\ 2bd - 2ac & 2cd + 2ab & a^2 - b^2 - c^2 + d^2 \end{bmatrix}. \quad (2.16)$$

The rotation matrix is an orthogonal matrix ( $RR^T = R^T R = I$ ;  $\det(R) = 1$ ). It rotates the molecule, represented by the three-dimensional atomic coordinates  $(x_j, y_j, z_j)$  to a new position with coordinates  $(x'_j, y'_j, z'_j)$ , written in matrix form  $\mathbf{X}'_j = \mathbf{R}\mathbf{X}_j$ . These coordinates are then used to calculate the structure factor of the molecule according to Eq. (2.3).



## 2.5 Number of Snapshots

### 2.5.1 Background

Electron density reconstructions from experimental SPI datasets collected at XFELs have achieved resolutions on the order of 5 to 10 nanometers at best [21, 22]. Diffraction up to a resolution of  $5.9\text{\AA}$  was already observed [64] but a reconstruction of a 3D diffraction volume was not attempted due to the small number of diffraction patterns collected. This immediately raises the central question of how many diffraction patterns must be collected for a 3D reconstruction of the electron density at a given resolution. In other words, how many diffraction patterns are required to obtain a diffraction volume at a signal-to-noise ratio (SNR) enough to reach the desired resolution by iterative phasing? Clearly, this depends on the molecule under investigation and

experimental conditions such as wavelength, beam size, and incident X-ray fluence. We formulate a mathematical model to estimate the required number of snapshots.

### 2.5.2 Mathematical Formulation

Let  $D$  be the particle diameter and  $d$  be the aimed resolution. We define a dimensionless quantity  $R = \text{'Number of Resolution Elements'}$  as

$$R = \frac{D}{d}. \quad (2.17)$$

The number of Shannon voxels in the outermost shell covered by a single diffraction pattern for oversampling by a factor of two is

$$nV_{snapshot} = \frac{2\pi(1/d)}{1/2D} = \frac{4\pi D}{d} = 4\pi R. \quad (2.18)$$

Accordingly, the number of Shannon voxels in the resolution shell is

$$nV_{shell} = \frac{4\pi(1/d)^2}{(1/2D)^2} = \frac{16\pi D^2}{d^2} = 16\pi R^2. \quad (2.19)$$

The probability of an outermost shell voxel hit by a randomly oriented diffraction pattern is therefore

$$p = \frac{nV_{snapshot}}{nV_{shell}} = \frac{1}{4R}. \quad (2.20)$$

Let  $\langle n \rangle$  denote the mean number of expected photons per Shannon pixel of a diffraction pattern at the resolution shell. As single photons are counted by the detector, the signal follows Poisson statistics,  $\text{var}(n^2) = \langle n \rangle$ . Accordingly, the SNR is

$$\text{SNR} = \frac{\langle n \rangle}{\sqrt{\text{var}(n^2)}} = \frac{\langle n \rangle}{\sqrt{\langle n \rangle}} = \sqrt{\langle n \rangle}. \quad (2.21)$$

Due to the weak scattering of X-rays from a single molecule, the SNR of a single diffraction snapshot is way too low for any high-resolution information. It is therefore necessary to obtain information from many snapshots by averaging. The number of times  $M$  a voxel must be hit by a diffraction pattern in order to reach the desired SNR is

$$M \times \langle n \rangle = SNR^2. \quad (2.22)$$

As an example, for  $\langle n \rangle = 0.002$  (phytochrome at 10Å resolution, Fig. 3.1 (a) each voxel must be visited at least 500 times to achieve a SNR of 1. The probability  $P$  for a single voxel be visited at least  $M$  times by an ensemble of  $nS$  snapshots is estimated by the following sum of Binomial distributions

$$P(p, M, nS) = 1 - \sum_{k=0}^{M-1} \binom{nS}{k} p^k (1-p)^{nS-k}. \quad (2.23)$$

Under the assumption that individual Shannon voxels are visited independently (justification is given in Appendix 5.1), the joint probability  $\tilde{P}$  to observe all voxels at least  $M$  times is

$$\tilde{P}(p, M, nS, nV_{shell}) = P(p, M, nS)^{\frac{nV_{shell}}{2}}. \quad (2.24)$$

The factor  $\frac{1}{2}$  in front of  $nV_{shell}$  in the equation is due to Friedel's symmetry. Using these relations, we can estimate the total number of diffraction patterns needed to cover the entire diffraction volume at any desired probability  $\tilde{P}$ . For the special case when  $M = 1$ , i.e. at very high signal levels, an estimation of the number of snapshots was proposed previously [22] and is in agreement with our formulation.

Eq. (2.24) cannot be analytically solved for  $nS$ . Instead one can calculate the right-hand side with increasing  $nS$  until the desired probability  $P$  is reached. An implementation of an efficient algorithm in Python is listed in Appendix 5.2. However, an analytical formula for the number of



snapshots can be obtained using the de Moivre-Laplace theorem which approximates the binomial distribution by a Gaussian:

$$\binom{nS}{k} p^k (1-p)^{nS-k} \simeq \frac{1}{\sqrt{2\pi p(1-p)nS}} e^{-\frac{(k-pnS)^2}{2p(1-p)nS}}$$

$$P(p, M, nS) = 1 - \sum_{k=0}^{M-1} \binom{nS}{k} p^k (1-p)^{nS-k} = \sum_{k=M}^{\infty} \binom{nS}{k} p^k (1-p)^{nS-k}$$

$$\simeq \frac{1}{\sqrt{2\pi p(1-p)nS}} \int_M^{\infty} e^{-\frac{(k-pnS)^2}{2p(1-p)nS}} dk = \frac{1}{2} \operatorname{erfc}\left(\frac{M-pnS}{\sqrt{p(1-p)nS}}\right). \quad (2.25)$$

This approximation allows to write  $nS$  explicitly.

$$nS(p, M, P) = \left( \frac{\sqrt{E(P)^2 + 4pM} - E(P)}{2p} \right)^2,$$

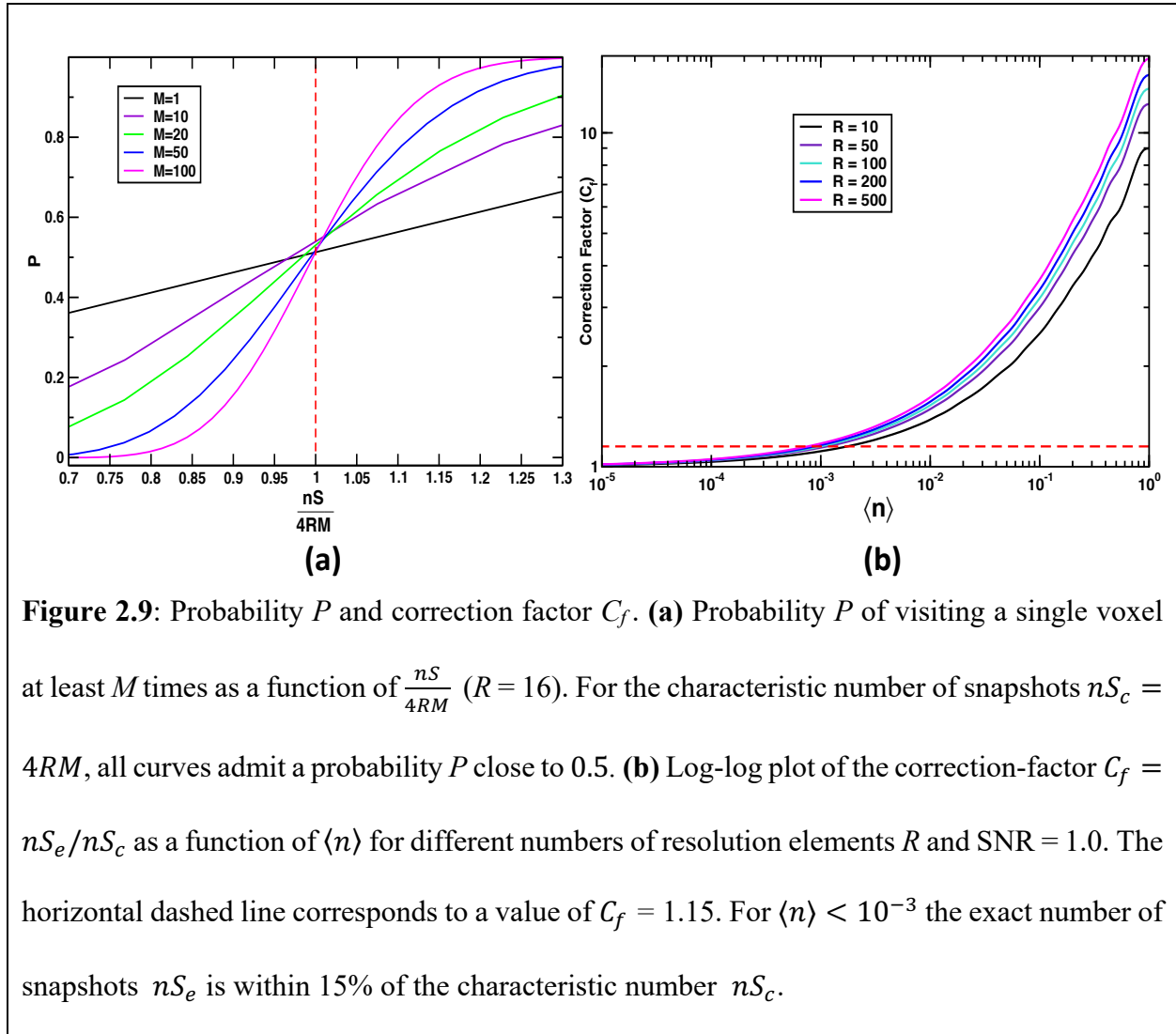
$$\text{where } E(P) = \operatorname{erfc}^{-1}(2P) \sqrt{p(1-p)}.$$

Since  $\operatorname{erfc}(0) = 1$ , the probability  $P$  becomes 0.5 for  $nS_c = \frac{M}{p}$ . We call  $nS_c$  the characteristic number of snapshots, the number required for a single voxel being visited at least  $M$  times with probability 50%. Together with Eq. (2.20) and (2.22) this yields

$$nS_c = 4RM = \frac{4R \operatorname{SNR}^2}{\langle n \rangle}. \quad (2.26)$$

To verify the approximation, the exact probability  $P$  was calculated according to Eq. (2.23) as a function of  $nS/4RM$  for different values of  $M$ . The result is depicted in Fig. 2.9 (a). All curves admit a value of  $P = 0.5$  for  $nS/4RM = 1$ , in close agreement with the approximation (3.9). The ratio of the exact number of snapshots  $nS_e$ , derived from Eq. (2.24) to the characteristic number of snapshots from Eq. (2.26) is expressed as a correction factor  $C_f = nS_e/nS_c$ . A plot of  $C_f$  as a function of the mean number of photons  $\langle n \rangle$  for a SNR value of 1.0 is shown in Fig. 2.9 (b). In this plot  $nS_e$  approaches  $nS_c$  as the number of photons is lowered. The exact number  $nS_e$  can easily

be estimated from the characteristic number  $nS_c$  by multiplication with the proper correction factor taken from the graph, without the need to solve Eq. (2.24).



For all voxels being jointly visited at least  $M$  times, the number of snapshots will be larger of course. With a joint probability  $\tilde{P} = 0.5$  in Eq. (2.24) a single-voxel probability  $P = (1/2)^{2/nV}$  must be reached instead, a value much closer to one.

According to Eq. (2.26) the most important parameter for estimating the number of diffraction patterns is the mean number of expected photons per Shannon pixel  $\langle n \rangle$  at the desired resolution  $d$ . Different methods to calculate  $\langle n \rangle$  are explained in the next section.

### 2.5.3 Estimating the Average Number of Photons

The average number of photons  $\langle n \rangle$  in the outermost resolution shell can be calculated from the structure factor  $F(\mathbf{q})$  as described in Eq. (2.13). However, these calculations require an atomic model, which is usually not available. Fortunately, there are useful approximations which only require the molecular weight and the approximate size of the molecule. Representing all non-hydrogen atoms by carbon atoms, the number of atoms  $N_c$  can be estimated from the molecular weight  $MW$ ,

$$N_c = MW/12\text{Da}.$$

At low-resolution one can approximate the molecule by a sphere with diameter  $D$  [65]. The squared structure factor of a sphere filled with  $N_c$  carbon atoms of atomic number  $Z_c = 6$  is

$$|F(\mathbf{q})|^2 = N_c^2 Z_c^2 \left( 3 \frac{\sin(\pi q D) - \pi q D \cos(\pi q D)}{(\pi q D)^3} \right)^2.$$

Accordingly, the average number of photons becomes

$$\langle n \rangle = \phi r_e^2 N_c^2 Z_c^2 \int \left( 3 \frac{\sin(\pi q D) - \pi q D \cos(\pi q D)}{(\pi q D)^3} \right)^2 d\Omega.$$

With the number of resolution elements  $R = D/d$  and  $d = 1/q$  this can also be written as

$$\langle n \rangle = \phi r_e^2 N_c^2 Z_c^2 \int \left( 3 \frac{\sin(\pi R) - \pi R \cos(\pi R)}{(\pi R)^3} \right)^2 d\Omega. \quad (2.27)$$

Thus, for the low-resolution approximation, the geometric part of the structure factor depends only on the number of resolution elements.

For the high-resolution regime, a different approximation is quite useful. With  $N_c$  carbon atoms the squared structure factor is

$$|\mathbf{F}(\mathbf{q})|^2 = |f_c(q) \sum_{j=1}^{N_c} \exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j)|^2 = |f_c(q)|^2 \left| \sum_{j=1}^{N_c} \exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j) \right|^2,$$

where  $f_c(q)$  denotes the atomic form factor of carbon. Taking the average over a shell with radius  $q$  yields

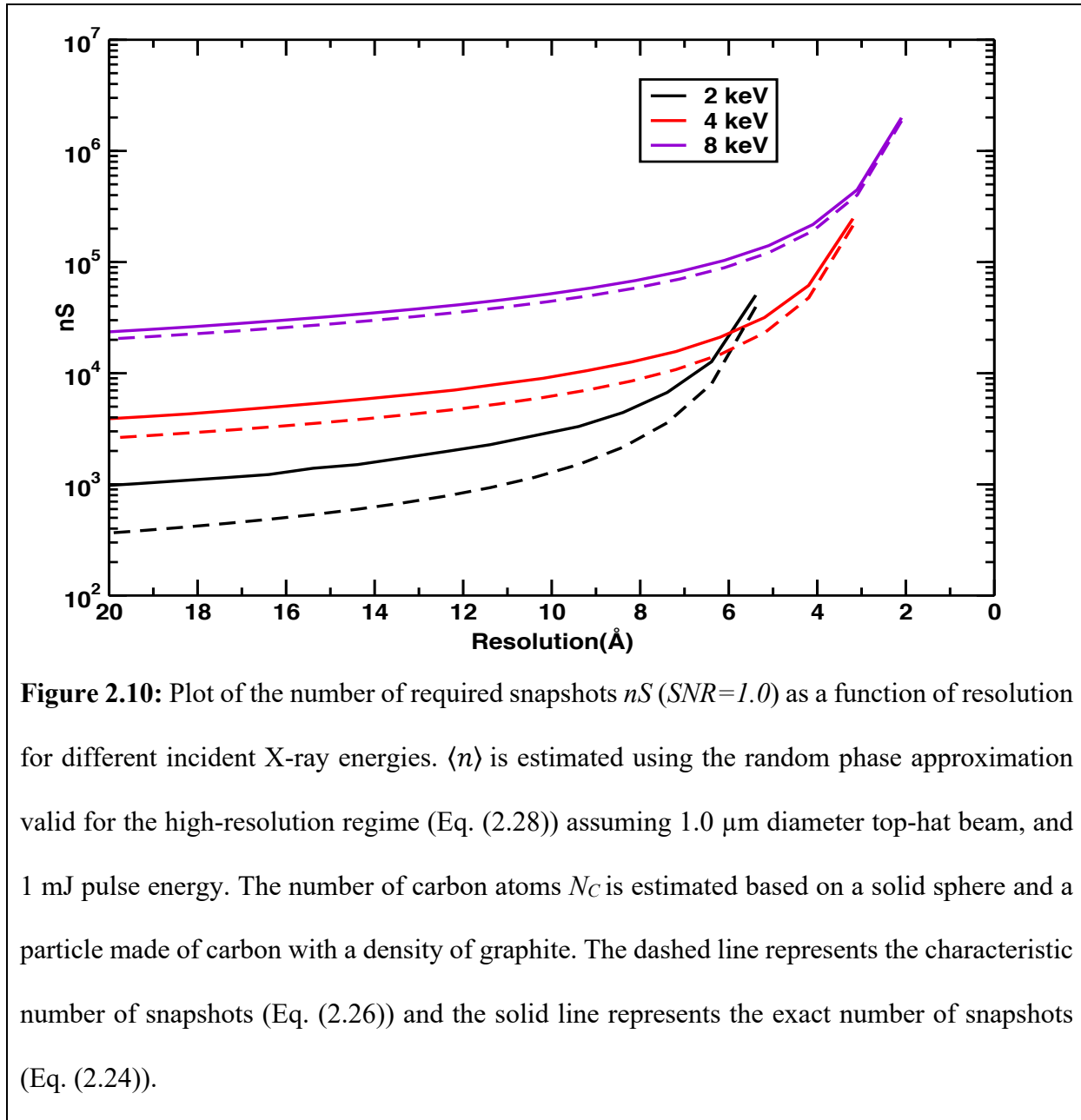
$$\langle |\mathbf{F}(\mathbf{q})|^2 \rangle = |f_c(q)|^2 \langle \left| \sum_{j=1}^{N_c} \exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j) \right|^2 \rangle.$$

If the complex number  $\exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j)$  can be approximated by a random phasor (discussed in Appendix 5.3), the quantity  $\langle \left| \sum_{j=1}^{N_c} \exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j) \right|^2 \rangle$  becomes equal to  $N_c$  [66], so that

$$\langle n \rangle = \phi r_e^2 N_c |f_c(q)|^2 d\Omega. \quad (2.28)$$

Eq. (2.28) assumes that the positions of atoms are completely uncorrelated which is approximately true at resolutions approaching the atomic scale. For the high-resolution approximation  $\langle n \rangle$  is directly proportional to  $N_c$  and  $|f_c(q)|^2$ .

Assuming a solid sphere and a particle made of carbon with a density of graphite, then  $N_c = \frac{4}{3} \pi \left(\frac{D}{2}\right)^3 \rho$ , with  $\rho$  being the number density of atoms. As  $d\Omega$  is proportional with  $D^{-2}$  (Eq. (2.14)),  $\langle n \rangle$  is linearly proportional to the size of the object. Applying the high-resolution approximation for the value of  $\langle n \rangle$ , Fig. 2.10 shows a plot of the number of snapshots required at different X-ray energies. As  $\langle n \rangle \propto D$ , from Eq. (2.26) we can say the number of snapshots is independent of the particle size (assuming the density is constant) for high-resolution approximation.



## 2.6 Single-particle Reconstruction

XFELs produce large datasets of 2D diffraction patterns of biomolecules, containing up to millions of snapshots. These entire datasets not only contain diffraction patterns of single particles, but also snapshots of multiple particles, molecular aggregates, blank shots (no particle intercepted),

buffer impurities, and contaminated materials as discussed in [67]. Single-particle diffraction patterns must be separated from those originating from multiple particle diffraction and other impurities. The X-ray pulses produced from XFELs destroy the sample, making it impossible to get more than one diffraction pattern per particle. Accordingly, every single molecule is injected into the X-ray interaction region in an unknown orientation. The orientation of each particle that contributes to a diffraction pattern must be recovered. Intensities must be extracted from the diffraction pattern and must be merged and converted to amplitudes. Finally, the phase is determined for each structure factor amplitude using iterative phasing from which ultimately electron density map is calculated.

### 2.6.1 Orientation Determination

The particles are injected in random orientations into the X-ray beam. The orientations of the molecules relative to each other have to be determined from the snapshots. Orientation recovery is one of the most crucial steps in single-particle structure determination. Multiple algorithmic methods of orientation recovery have been developed to assign orientations to single-particle X-ray diffraction patterns [68–71]. For this dissertation, the *Dragonfly* package [72] was used for orientation determination. *Dragonfly* uses the *Expansion Maximization Compression* (EMC) algorithm [70].

#### 2.6.1.1 Expansion Maximization and Compression (EMC)

EMC uses a Bayesian approach for orientation determination. This is an iterative algorithm, and each iteration consists of three steps: *Expansion* (E), *Maximization* (M), and

*Compression* (C). Initially, it can be started with any prior knowledge about the particle or could be seeded from a guess model comprising a random set of numbers. In the first step (*Expansion* step) of each iteration, the current 3D volume is expanded into a set of diffraction patterns. The expansion is based on each potential measurement angle or orientations. The experimental (measured) patterns are compared separately to the expanded model calculated in the previous step for all possible orientations. The log-likelihood probability in each orientation is calculated. The patterns are assigned orientations based on where they fit best. This gives a new expanded maximum-likelihood model according to the log-likelihood probability (*Maximization* step). In the last step of the iteration, a new expanded model, using the newly assigned orientations, is compressed back into a 3D volume (*Compression* step). After this, a new iteration is started until orientations converge.

### 2.6.2 3D Merging

The diffraction patterns are merged into a 3D volume after orientation determination. A so-called “Cone-Gridding” algorithm is used for this purpose. This algorithm has previously been used to analyze real SPI data collected at an XFEL [73]. To estimate the intensity,  $I(q)$ , at a point  $q$ , the algorithm identifies pixels in all the diffraction patterns that correspond to the scattering vector  $q$ . This happens only if the Ewald sphere associated with a diffraction pattern intersects  $q$ . Geometrically, the directions of the incident beam of diffraction patterns which contribute to a point  $q$  form a cone, hence the name “Cone-Gridding.” For a finite number of diffraction patterns, the Ewald spheres rarely pass through any of the grid points. This is addressed by relaxing the elastic scattering condition to allow an angular deviation up to a Shannon angle. The gridpoint  $q$  is projected onto the 2D diffraction pattern and the scattering amplitude is interpolated from the

neighboring grid points. At each grid point of the diffraction volume, the intensity is estimated by the weighted sum:

$$I(q_{i,j,k}) = \sum_{d=1}^D w(q_{i,j,k})_d I_d(P(q_{i,j,k})) / \sum_{d=1}^D w(q_{i,j,k})_m.$$

where  $D$  is the number of diffraction patterns contributing to the grid point  $(i,j,k)$ ,  $w(q_{i,j,k})_d$  represents the Gaussian weights and  $I_d(P(q_{i,j,k}))$  denotes the interpolated intensity retrieved from diffraction pattern  $d$  by projecting the scattering vector into the diffraction plane. The Gaussian weights are determined by taking the deviation from the Ewald condition

$$w(q_{i,j,k})_d = \exp\left[-\frac{\Delta\theta_d^2}{2\sigma^2}\right].$$

Here  $\sigma$  represents the Shannon angle and  $\Delta\theta$  is the deviation from the elastic scattering condition (how much distance from the surface of the Ewald sphere). For all the pixels with  $\Delta\theta_d > 2\sigma$  the weight is set equal to zero.

### 2.6.3 Phase Retrieval

In SPI, the electron density distribution of a biomolecule is reconstructed from the merged 3D diffraction volume via iterative phase retrieval methods.

#### 2.6.3.1 The Phase Problem

The phases of the structure factor amplitudes must be known to reconstruct the electron density. The electromagnetic waves X-rays have oscillation periods on the order of  $10^{-20}$  to  $10^{-16}$  seconds. X-ray detectors only sense the intensity in a pixel and cannot resolve the extremely rapid



oscillations. This leads to the phase problem in both SPI and crystallography. Phases are retrieved using iterative calculations with specific boundary conditions. The boundary conditions are based on the properties of the real and reciprocal spaces. Solutions to the phase problem are discussed in Section (2.5.4.4).

### 2.6.3.2 Importance of Phases

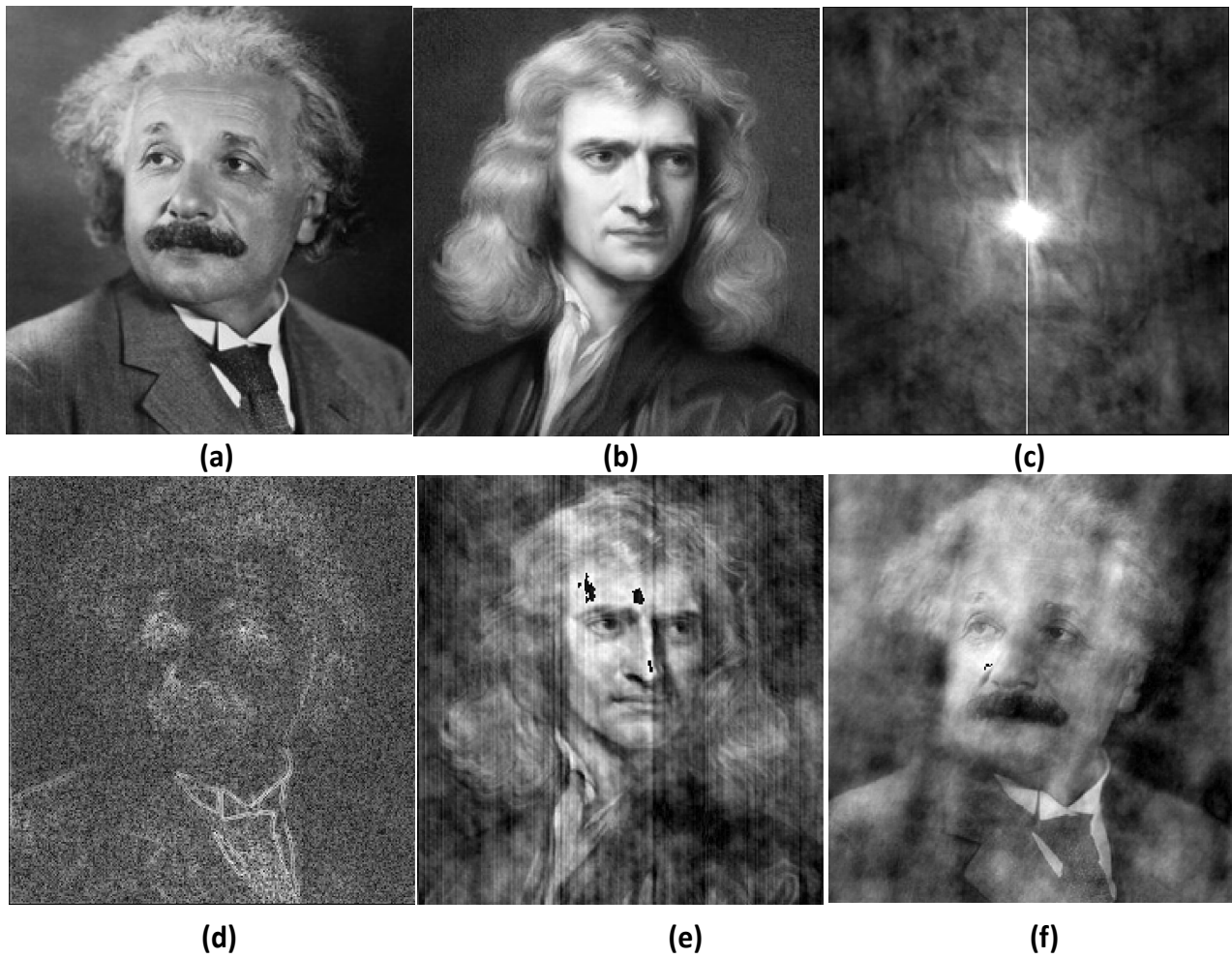
The importance of phases in reconstructing the correct image is illustrated in Fig 2.11. As phase information is unknown, one could try to reconstruct an image assuming that the phases have an arbitrary value, say  $0^\circ$ , using amplitudes (square root of the measured intensities) with inverse FT as  $FT^{-1}\{\sqrt{I} \exp(i \cdot 0)\}$ . This results in an uninterpretable image (see Fig. 2.11 (c)). On the other hand, if the phases were known and amplitudes unknown, the resulting image obtained with inverse FT as  $FT^{-1}\{1 \cdot \exp(i \cdot \phi)\}$  is quite similar to the original image (see Fig. 2.11 (d)). For reconstruction with the amplitude information of one image and the phase information of the other, the image corresponding to the phase data is prominent but somewhat degraded as in Fig. (2.11 (e) & 2.11 (f)). This suggests that the phases carry more structural information than the amplitudes.

### 2.6.4.3 Oversampling

Since X-ray diffraction intensities of a single molecule are continuous (Fig. 3.3) iterative phasing becomes possible [74, 75]. The phase information of continuous diffraction patterns can be recovered by sampling the intensities at angular intervals at least with the ‘Shannon spacing’. For an object with diameter  $D$ , the Shannon sampling interval is given as  $\Delta q_s = 1/(2D)$ .

Measuring intensities more finely than this do not increase the information content in the diffraction pattern. Sampling with the Shannon frequency is called (2-fold) *oversampling*. If the diffraction pattern is measured to a maximum resolution  $q_{max} = 1/d$ , from range  $-q_{max}$  to  $+q_{max}$ , then there are  $N_s = \frac{2q_{max}}{\Delta q_s} = \frac{4D}{d}$  Shannon intervals.

Oversampling the diffraction pattern in reciprocal space has a corresponding relationship in real space. There are two regions in real space, one where the object is allowed to have density and another which is empty (see Fig. 2.13 (d)). How large is the empty region relative to the region that contains density depends on the sampling frequency: the higher the oversampling rate, the larger is the empty region.



**Figure 2.11:** Illustration of the importance of phase in far-field diffraction. **(a)** and **(b)** are the images of Einstein and Newton respectively. **(c)** Reconstruction using Fourier amplitudes obtained from **(a)** and setting phase to a constant value, say zero. **(d)** Reconstruction using known phase information from Einstein **(a)** but unknown amplitude (setting the amplitude to a constant value of 1). For two images **(a)** and **(b)**, their Fourier phases are now interchanged while keeping the amplitude: **(e)** Fourier synthesis using the amplitudes from Einstein **(a)** and phases from Newton **(b)**. **(f)** Fourier synthesis using the amplitudes from Newton **(b)** and phases from Einstein **(a)**. The images of Einstein and Newton are taken from the public domain and can be found through a web search.

#### 2.6.4.4 Iterative Phasing Algorithms

The phase retrieval algorithms iterate back and forth between real space and reciprocal space. Constraints in both real space and reciprocal space are applied. The constraint in the reciprocal space is given by the experimental data, the Fourier amplitudes. This constraint is referred to as the *Fourier-space constraint (modulus constraint)*. The second constraint is given by the oversampling. The region of the real space that allows density is called the object's support and the constraint imposed by it is called the *real-space constraint (support constraint)*. The iterative algorithm begins with the application of random phases to the experimental amplitudes. The algorithm implements the following steps.

1. Assign a random phase set as an initial guess to every voxel of the diffraction volume.
2. Combine this random phase set with the measured Fourier magnitude and apply an inverse Fourier transform.
3. Estimate support (i.e. a boundary slightly larger than envelope or the diameter of the molecule) based on the oversampling rate. Set all the voxels outside the support to zero.
4. Apply Fourier transform (FT) to the updated volume.
5. Replace the amplitudes of the resulting computed FT with the measured amplitudes (the magnitude of the amplitude is the *Fourier-space constraint*) but retain the phases.
6. Repeat from step 2 until the computed phases converge and computed density becomes chemically meaningful. In this way, a solution is obtained.

Depending on how the supports change in every iteration, there are several phasing algorithms.

### 2.6.4.5 The Error-Reduction and Hybrid Input-Output Methods

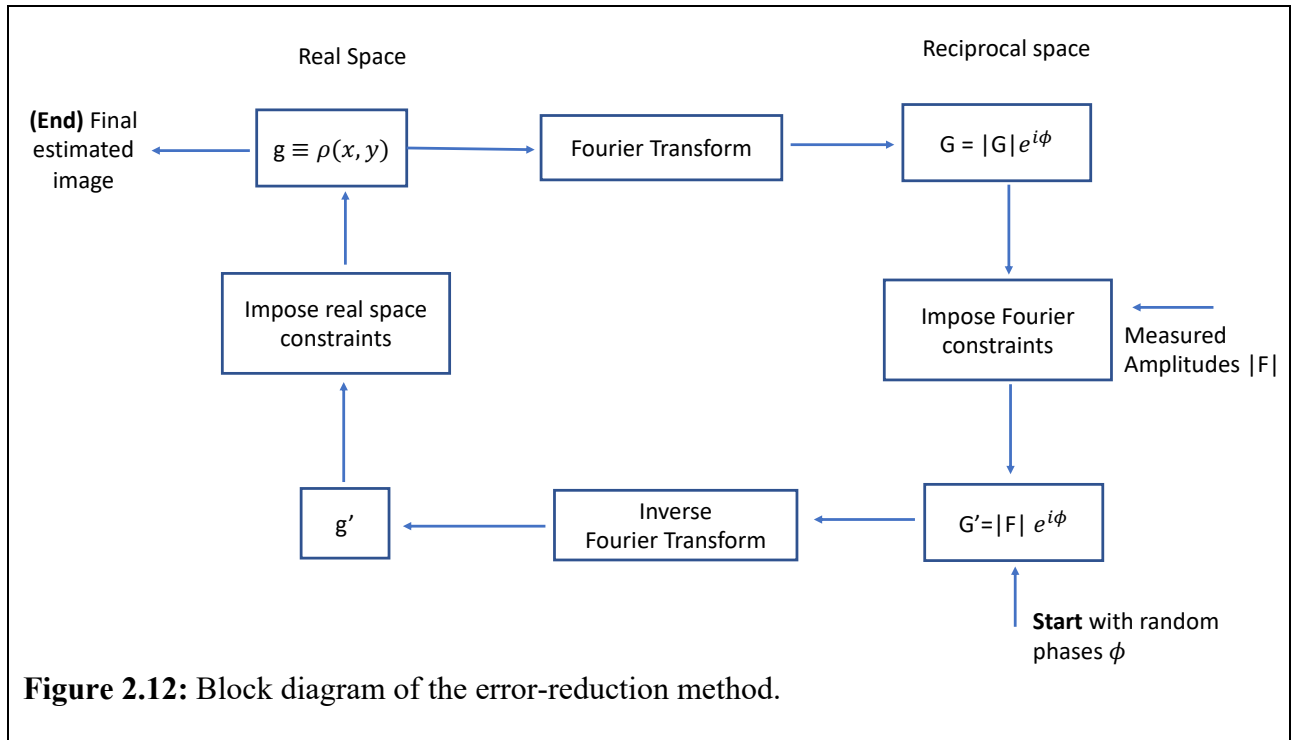
The first iterative algorithm was pioneered by Gerchberg and Saxton [76] which was later improved by Fienup, and called error-reduction (ER) method [77]. This algorithm is shown graphically in Fig. 2.12. The ER algorithm starts with an outline of the space where the object is defined i.e. the support. It is assumed that the real space outside the support will be set to zero for all the iterative transformations. The prior knowledge of the support  $S$  is derived from the autocorrelation function (Patterson function) obtained by the inverse FT of the measured diffraction intensity [78]. The constraint for the ER algorithm has the form

$$g_{k+1}(x, y) = \begin{cases} g'_k(x, y), & (x, y) \in S \\ 0, & (x, y) \notin S \end{cases}$$

The notations used in the above constraint are explained in Fig. 2.12. To overcome the stagnation problems encountered by the ER method [79], Fienup devised some modifications to the method and called it Hybrid Input-Output Method (HIO) [80]. This method employs a feedback parameter  $\beta$  to push the pixel values outside the support to zero without rigidly constraining them. The constraint for the HIO algorithm has the form

$$g_{k+1}(x, y) = \begin{cases} g'_k(x, y), & (x, y) \in S \\ g_k(x, y) - \beta g'_k(x, y), & (x, y) \notin S \end{cases} \quad (2.29)$$

For  $\beta=1$ , HIO and ER algorithm coincide. This algorithm is called a *hybrid algorithm* as a new iteration depends on the value of the previous iteration.

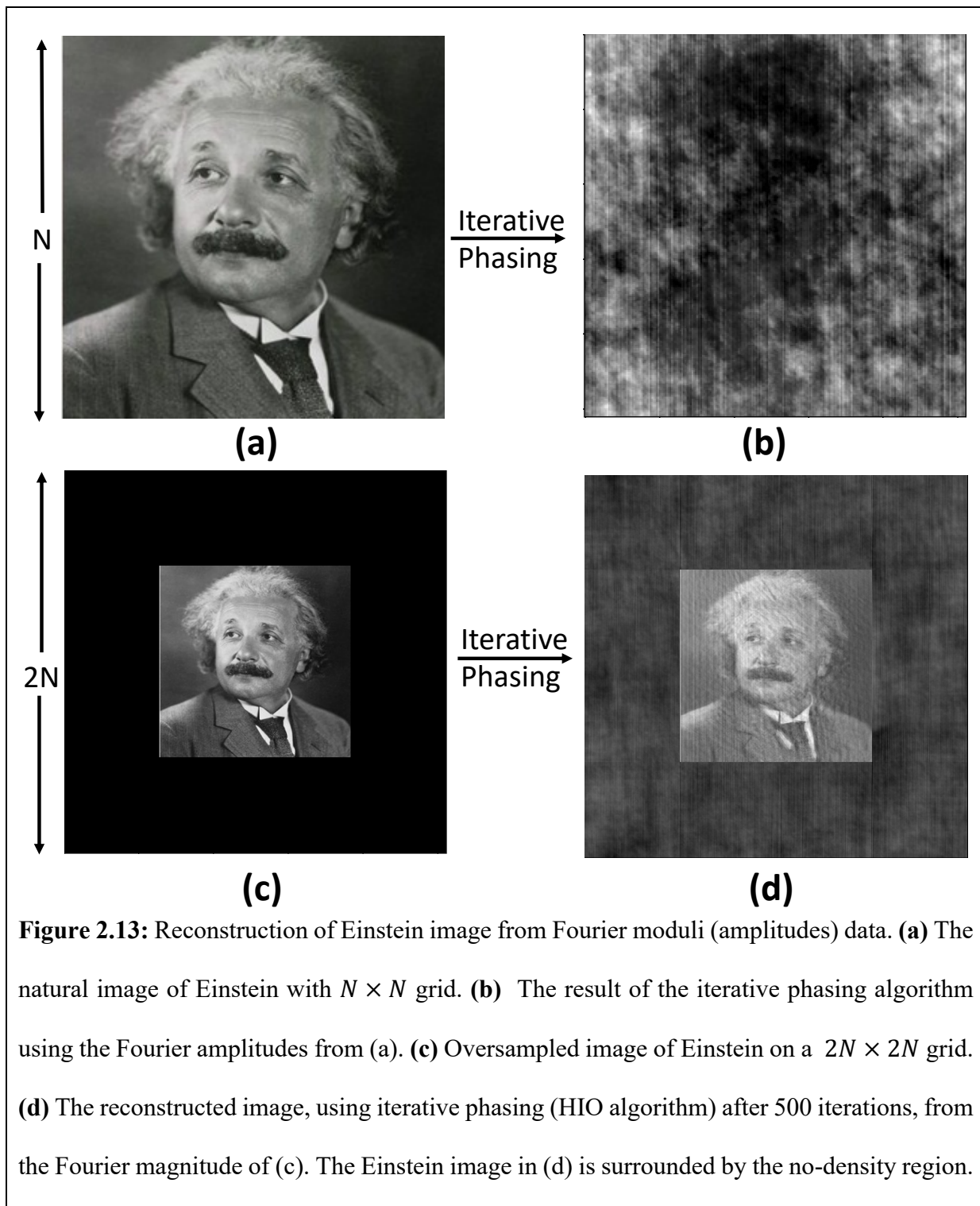


#### 2.6.4.6 Different Phase Retrieval Algorithms

Many phase retrieval algorithms proposed to date are based on the HIO algorithm [80]. A comprehensive list of most of the iterative algorithms and comparison were compiled by Marchesini [81]. The support constraints (the volume or size of the 3D support) in ER and HIO algorithms were held constant for the entire iterations in one run of the algorithm. The optimization of the support constraint based on the hit and trial method was very inefficient. In 2003, Marchesini et al. [82] developed a method to dynamically update the support during the iterations. This provided an efficient approach to optimize the support constraint. This algorithm is known as the *shrink-wrap algorithm* in which the initial support is gradually updated in regular steps. The original idea is to find a new support mask by convolving the reconstructed image (the absolute value of the reconstructed wavefield) with a Gaussian of a certain width  $\sigma$ . The updated support is obtained by applying a threshold that can be chosen by the user. The width  $\sigma$  can be reduced by a

certain value in every iteration. The threshold and the Gaussian width are factors that may be systematically varied during the reconstruction process.

A few algorithms are based on slightly different ideas. For example, the *charge flipping algorithm* doesn't use support but instead treats low-density regions in real space different from high-density regions [83]. The sign of the charge density below a threshold is flipped while above the threshold remains unchanged. This leads to an effective correction for large false-negative electron density. This algorithm seems to work for high-resolution data only. The robustness of these iterative algorithms can be improved by applying additional object-specific constraints. As electron densities are real and positive, commonly used constraints are the *reality constraint* and the *positivity constraint*. This can be enforced by simply setting an imaginary electron density component to zero after applying the real-space constraint and forcing the value inside the support to be positive or enforcing Friedel's symmetry in reciprocal space. The reconstruction of a 2D image using iterative phasing algorithm is shown in Fig. 2.13.



**Figure 2.13:** Reconstruction of Einstein image from Fourier moduli (amplitudes) data. **(a)** The natural image of Einstein with  $N \times N$  grid. **(b)** The result of the iterative phasing algorithm using the Fourier amplitudes from (a). **(c)** Oversampled image of Einstein on a  $2N \times 2N$  grid. **(d)** The reconstructed image, using iterative phasing (HIO algorithm) after 500 iterations, from the Fourier magnitude of (c). The Einstein image in (d) is surrounded by the no-density region.



#### 2.6.4.7 Missing data

Most diffraction patterns from single-particle experiments are incomplete because of various experimental difficulties. Even in the best case, due to the presence of a beamstop to block the direct beam, there are some missing pixels in the center. For a given structure, increasing the amount of missing data usually means that the number of iteration cycles needed for a solution also increases. All iterative algorithms handle the missing data in the same way. When applying reciprocal-space constraints, pixels with missing experimental amplitudes retain their recovered (calculated) amplitude, just like all pixels retain their phases. Thus, the missing amplitudes are recovered in the same way as the missing phases.

So far, the focus was on the necessary details of the simulation of single-molecule diffraction patterns and the reconstruction of its electron density. Biological processes are highly dynamic. To understand proteins in action, a fourth dimension, time, must be added to the diffraction patterns. To date, there are significant experimental challenges [84, 85] to conduct a time-resolved SPI experiment. However, the determination of both structure and dynamics is possible using time-resolved crystallography [86–88]. The changes in the atomic coordinates of the transient states can be recorded ‘in real-time’ [41, 89]. For this, proteins need to be functionally active in their crystalline state. The strategies to trap the transient states are referred as ‘kinetic crystallography’ [90]. The time-resolved X-ray crystallography is one of the several ‘kinetic crystallography’ techniques that trap the intermediates in a reaction. The next few sections will detail all the necessary formalisms required to understand the time-resolved pump-probe serial femtosecond crystallography experiments at modern XFELs.

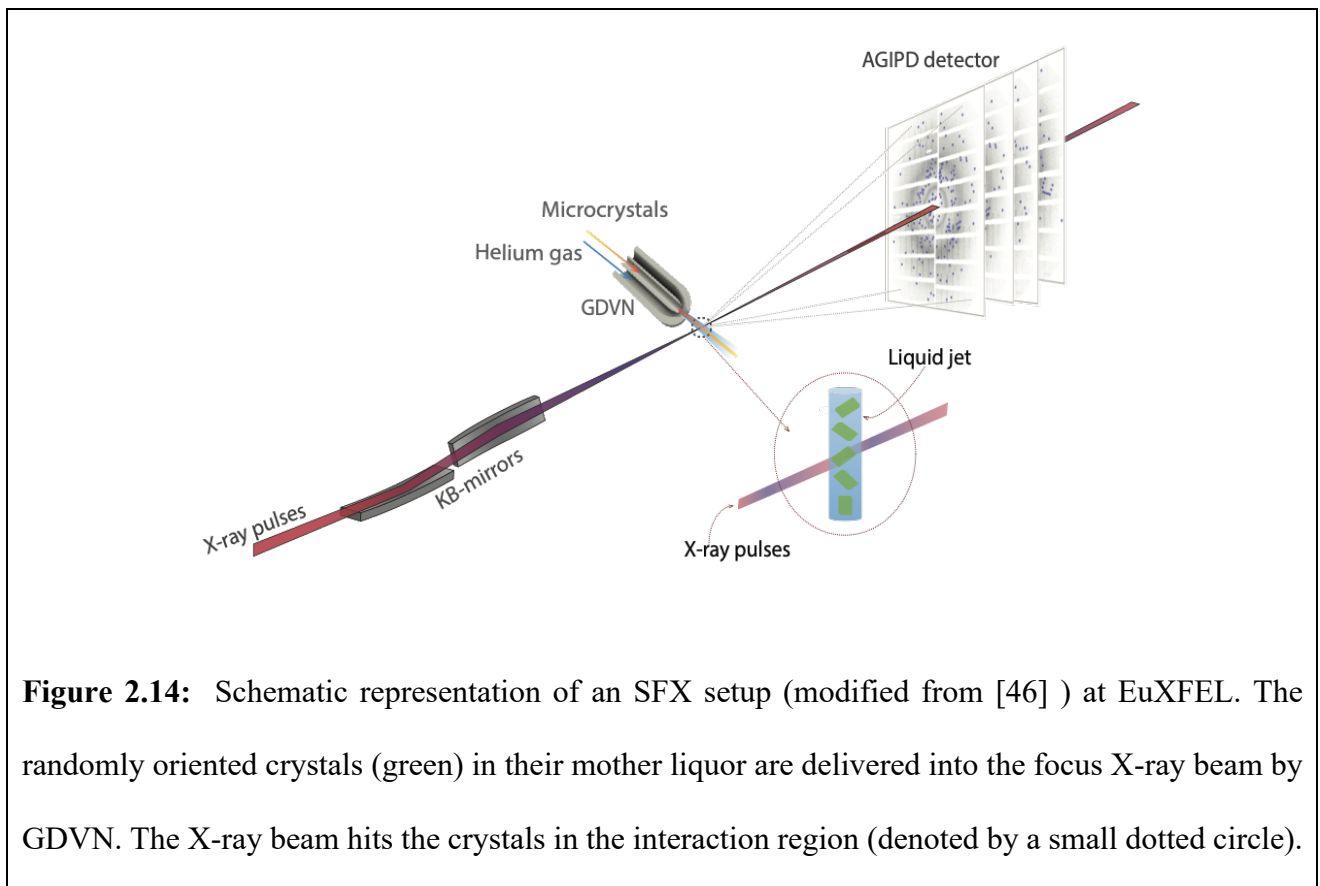
## 2.7 Serial Femtosecond Crystallography

The incident X-ray intensity at XFELs is so large, that microcrystals can be probed. Furthermore, diffraction experiments are conducted at room temperatures that maintain the dynamics and function of the biomolecule. Each microcrystal exposed to an XFEL pulse is destroyed. Because of this, a dataset must be collected from thousands of randomly oriented crystals, each exposed only once. This data collection approach is termed serial femtosecond crystallography (SFX). The method is called '*serial*' because the crystals are injected serially into the X-ray interaction region. This way each snapshot is from a fresh crystal. It is called '*femtosecond*' because the crystals interact with a single X-ray pulse that lasts only about 10-40 femtoseconds.

### 2.7.1 Sample Injection

The typical SFX experiment is shown in Fig. 2.14. The crystals are suspended in their mother liquor or in a viscous carrier. They are introduced into the X-ray interaction region one-by-one in a serial fashion in a jet. The sample must be delivered and replenished at faster rates than XFEL pulses to avoid the impact from the previous pulse. The delivery method should be compatible with the experimental environment like vacuum, helium, air. The crystals need to be treated delicately during the injection process to avoid mechanical and chemical stress. The quality of the recorded pattern depends on the background level. The background signal is mainly due to the diffraction signal from the mother liquor. Optimally, the size of the liquid jet should match the size of the crystals.

Several devices were developed for injecting microcrystals, such as the gas dynamic virtual nozzle (GDVN) [91], the liquid cubic phase (LCP) injector (high viscosity injectors) [92], and the mixing injector [93–95]. The commonly used injector system for the SFX experiment is the GDVN which produces a fast-flowing, stable  $\mu\text{m}$ -sized liquid jet [41, 44, 46]. The major limitation of the GDVN is its relatively high sample consumption. The LCP injector consumes only a small amount of sample but produces a high scattering background. At the Spring-8 Angstrom Compact free electron LASer (SACLA) mineral oil-based grease matrices are used as a carrier of protein microcrystals for SFX experiments [96–98]. The sample delivery methods are evolving to minimize sample consumption, reduce the frequency of clogging or freezing, and minimize the background scattering signal.



## 2.7.2 X-ray Detectors

2D detectors or area detectors are the core of an X-ray diffraction experiment. An area detector contains a 2D array of X-ray sensitive elements (pixels). The larger the area of the detector, the greater the number of pixels on the detector, and the larger is the solid angle that can be covered at a particular sample-to-detector distance.

2D detectors can be classified into two broad categories: *photon-counting detectors* and *integrating detectors* [99]. In a photon-counting detector, each absorbed X-ray photon is converted to an electronic pulse. The processing electronics can distinguish the signal generated by an X-ray photon from that produced by other photons. It can also discriminate against the electronic background. The photon-counting detectors cannot handle large instantaneous fluxes because the electronic pulses from each must be distinguishable which becomes impossible at high counting rates caused by high X-ray intensities. Instead, integrating detectors are used which integrate the total signal, including noise, during a user-selected, short period. These types of detectors are used at XFEL sources.

XFELs demand that the detector readout frequency is compatible with the X-ray pulse frequency. Besides, their dynamic range must be sufficient, to detect single photons for example to support single-molecule imaging and intense Bragg spots for crystallographic experiments. The first detector developed for LCLS was a 2.3-megapixel Cornell-SLAC Pixel Array Detector (CSPAD) [100]. Every single pixel in CSPAD can be set to either a high or a low-gain mode. The CSPAD has a dynamic range of  $\sim 2700$  photons at 8 keV in the low gain mode [100]. The newly developed Adaptive Gain Integrating Pixel Detector (AGIPD) for the EuXFEL automatically switches the

gain for each pixel, resulting in three distinct gain levels (high/medium/low) which depend on the incoming X-ray intensity. Single photons can be recorded in the high gain mode. The low gain provides a dynamic range of  $> 10^4$  photons at 12 keV [101]. Detector pixels are arranged on tiles. Several tiles form quadrants. The quadrants are arranged so that the strong direct beam passes through the hole in the center. The unused beam, passing through the center hole, can be refocused downstream and used for another experiment [102] in parallel.

### 2.7.3 Processing of XFEL Data

During an SFX experiment, a single X-ray pulse produces a single diffraction pattern on the detector. Due to the quasi-monochromatic nature of the XFEL radiation, many partial reflection intensities are collected from only a small part of reciprocal space. To reconstruct the integrated intensities that cover the entire reciprocal space, tens of thousands of diffraction patterns must be collected and analyzed properly. Due to the nature of the sample delivery process, not every X-ray pulse hits a crystal, therefore, many blank images are produced. The data must be sorted and filtered so that only diffraction patterns which have a high likelihood of being usable for autoindexing and intensity integration are picked.

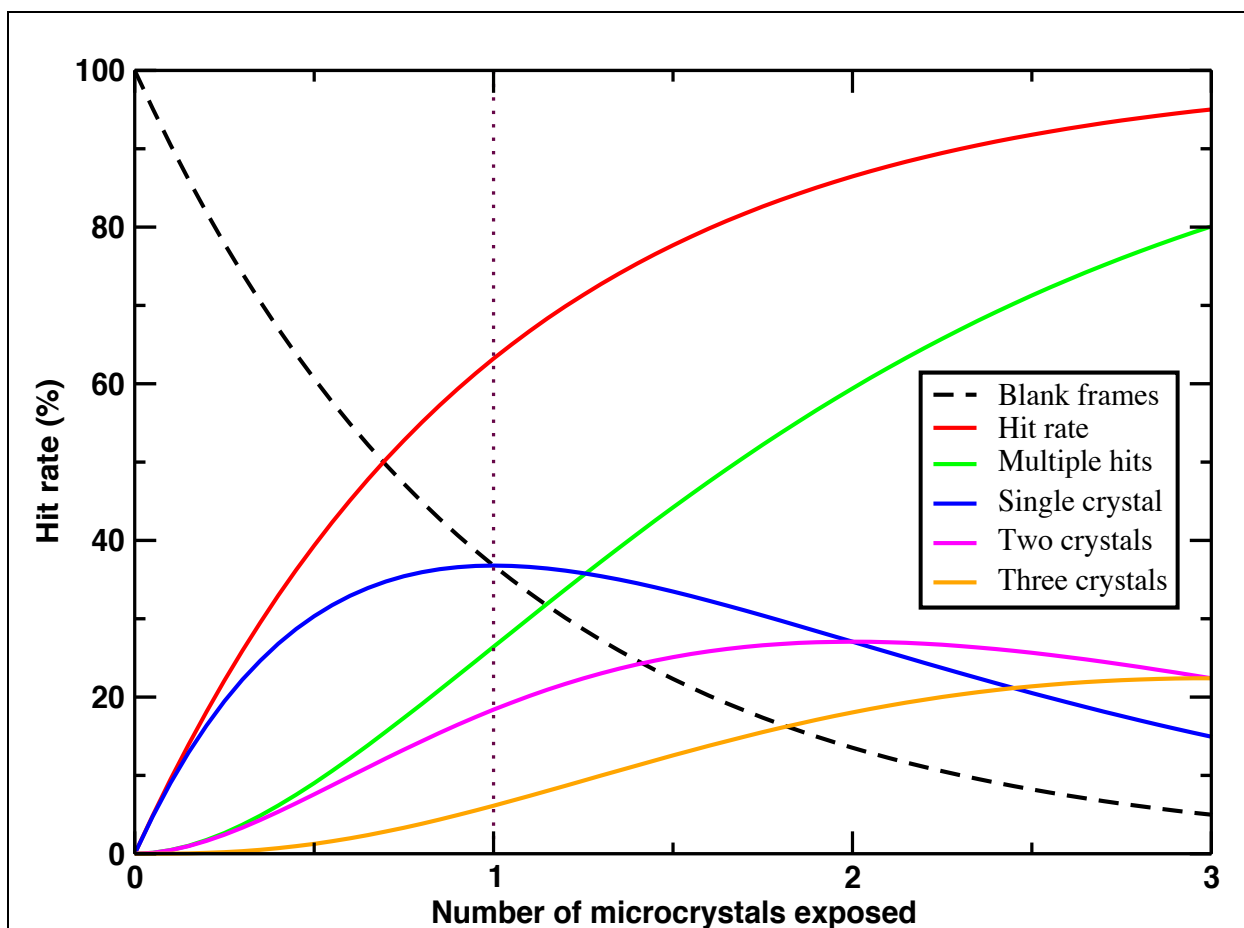
#### 2.7.3.1 Pre-processing

Detector images contain either Bragg reflections (*hits*) or blanks without Bragg reflections. The first step of the analysis is to extract the hits. By detecting and counting Bragg peaks on the detector a hit can be distinguished from a blank. *Cheetah*, an open-source program, [103] is commonly used for preprocessing. Cheetah also performs background subtraction and detector

geometry corrections. It also stores the images in a condensed form in the HDF5 (Hierarchical Data Format, version 5) file format.

### 2.7.3.2 Hit rate

Not every X-ray pulse diffracts from a microcrystal. In fact, over 99% of the microcrystals are never exposed. A microcrystal interacts with an X-ray pulse by chance. The hit rate is the ratio of diffraction patterns with hits to all diffraction patterns and is an important parameter for monitoring the success of an SFX experiment. The hit rate can be controlled by altering the concentration of the crystals (crystal density) or the thickness of the liquid jet (amongst others). A high microcrystal density will increase the likelihood of hits but also increases the chance of clogging the nozzle. Moreover, the probability to expose multiple crystals with the same X-ray pulse increases. The arrival of microcrystals in the X-ray interaction region is a stochastic event and determined by Poisson statistics [104]. The relationship between the hit rate and the crystal density is given by a Poisson distribution. The ideal hit rate in SFX is 63% [104]. The theoretical maximum for getting ‘single crystal hits’ (only one crystal hit by the X-ray pulse) is 37% as shown in Fig. 2.15. The remaining 26% of the hits are multiple hits. In most SFX experiments the actual hit rate is much lower on the order of 1% to 10% with liquid jets and up to 25% with LCP jets. The Poisson statistics for the hit rate is equally applicable for single particles used in an SPI experiment [105].

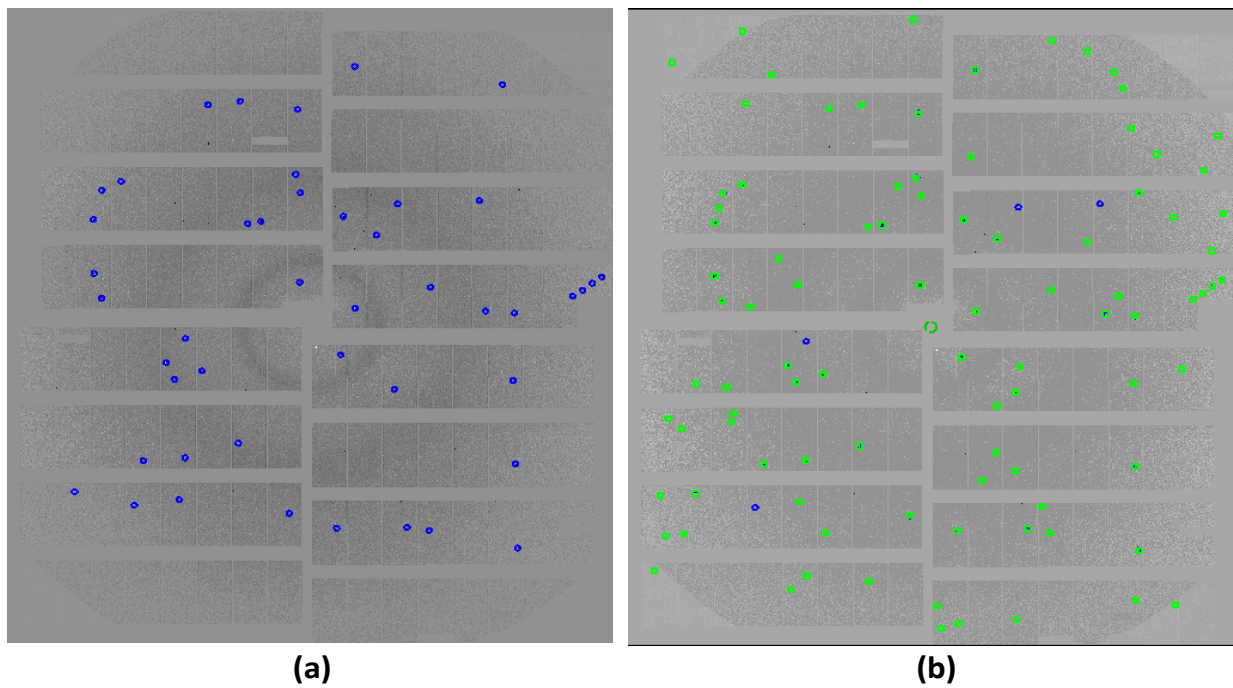


**Figure 2.15:** Theoretical estimate of the hit rate using Poisson statistics [104, 105]. The hit rate depends on the crystal density. With increasing crystal density more than one crystal is exposed by the same X-ray pulse. When the hit rate approaches 100%, multiple hits dominate the diffraction patterns. Diffraction patterns with multiple hits are difficult if not impossible to analyze because they contain diffraction from multiple, randomly oriented crystal lattices.

### 2.7.3.3 Indexing

The “hits” as identified by Cheetah are now indexed and integrated by *indexamajig*, a program of the CrystFEL software suite [106]. Indexing and integration in CrystFEL is performed by making use of external programs such as MOSFLM [107] , DirAX [108], XDS [109], and XGANDALF [110]. Indexing of diffraction pattern means assigning Miller indices to the Bragg peaks. In addition, the indexing algorithm can determine the cell parameters *ab initio* for each diffraction pattern. Once the crystal orientation is found, it is used to predict the location of the Bragg spots on the image. An image is considered indexed if a minimum percentage of found peaks lay close to the predicted peaks. The peak-finding and the indexing results for a diffraction pattern of photoactive yellow protein (PYP) [46] are shown in Fig. 2.16. The success rate of the indexing process is measured by the ‘indexing rate ’which is the ratio of the indexed patterns to the total number of hits. This largely depends on the number of hits identified by the preprocessing software.





**Figure 2.16** : Representative of a peak-finding and indexing results from PYP crystals [46] recorded on the AGIPD detector. **(a)** A diffraction pattern of a PYP crystal. The small blue circles represent the peaks as identified by the cheetah software. **(b)** The green circles represent the predicted reflections, as identified by the CrystFEL software [106], overlaid with the peaks identified in (a).

#### 2.7.3.4 Merging of Intensities

Each diffraction pattern from the SFX experiment is a thin slice through the reciprocal lattice and only contains partial reflections. The integrated intensity cannot be determined from a single snapshot. Only the merged data can be used for structure determination. The partial intensities are merged via Monte Carlo integration [111]. A sufficiently large number of random measurements around each Bragg spot will eventually provide a good estimate of the scattered intensity. Merging can be done using *process\_hkl* in CrystFEL that takes the symmetry of the

crystal into account. The merged intensities can be visualized by plotting flat central sections through reciprocal space using *render\_hkl* built into CrystFEL as shown in Fig. 2.17. The quality of the final reflections can be improved by simply scaling the intensities to account for variations in the pulse intensity and crystal size. The weaker intensities from smaller crystals and the stronger intensities from larger crystals are appropriately scaled to a common scale. More sophisticated post-refinement methods, such as scaling, and the determination of the partiality of the observed Bragg peaks are also implemented in CrystFEL with program *partialator*.

During merging, it is important that the indexing choices are compatible. For several indexing possibilities, the autoindexing programs only make a random choice. To address indexing ambiguity [112] in crystals *ambigator* module [113] in CrystFEL is used. For example, PYP crystallizes in hexagonal space group  $P6_3$ . It has a twofold degenerate indexing ambiguity because  $a^*$  and  $b^*$  have exactly the same length. An indexing ambiguity may also arise if a unit cell has a diagonal of similar lengths to one of the cell axes. The indexing ambiguity for PYP is  $(h\ k\ l)$  and  $(k\ h\ l)$ . This means crystals with indexing ambiguity could produce two diffraction patterns with identical Bragg peak positions but with different intensities. By not resolving this ambiguity the final dataset will artificially be of higher symmetry as shown in Fig. 2.17. After solving the indexing ambiguity, scaling, and merging the partial intensities the program CCP4 [114] is used for refinement and reconstruction of the electron-densities.

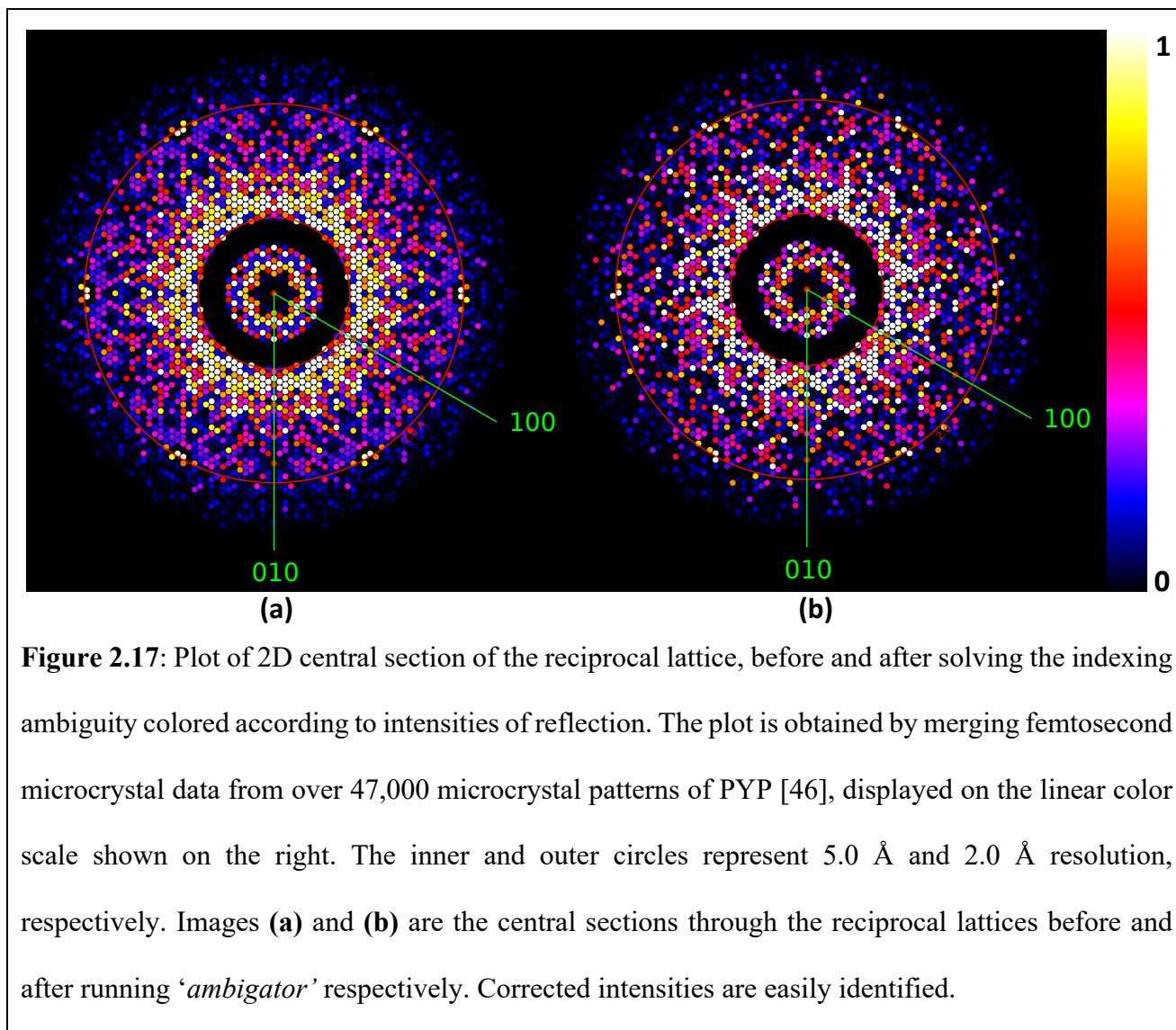
### 2.7.3.5 Evaluating the Data Quality

The quality of the data is measured by splitting the collected dataset in halves, merging each one separately, then comparing the two half-datasets using correlation coefficients and an R-factor. The  $R_{split}$  is computed as follows:

$$R_{split} = \frac{1}{\sqrt{2}} \frac{\sum |I_{even} - I_{odd}|}{\frac{1}{2} \sum (I_{even} + I_{odd})}. \quad (2.32)$$

Here  $I_{even}$  represents the intensity of a reflection produced by merging even-numbered patterns,  $I_{odd}$  represents the intensity of the equivalent reflection from the odd-numbered patterns and the sum is over all reflection indices. The ratio is divided by  $\sqrt{2}$  because the R-split overestimates the error relative to conventional R-factor computed from the full number of observations [56]. Another figure of merit is  $CC_{1/2}$ , which is the Pearson correlation coefficient between the intensities in each half dataset. The  $CC^*$  estimates the correlation coefficient between the final merged dataset and a hypothetical true dataset.  $CC^*$  is derived from  $CC_{1/2}$  using the formula [115]

$$CC^* = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}. \quad (2.33)$$



## 2.7.4 Experiments at XFELs

### 2.7.4.1 List of Experiments at XFELs

I participated in a number of experiments at various XFELs. They are shown in table 1.

Table 2.1: Experiments at the worldwide XFELs

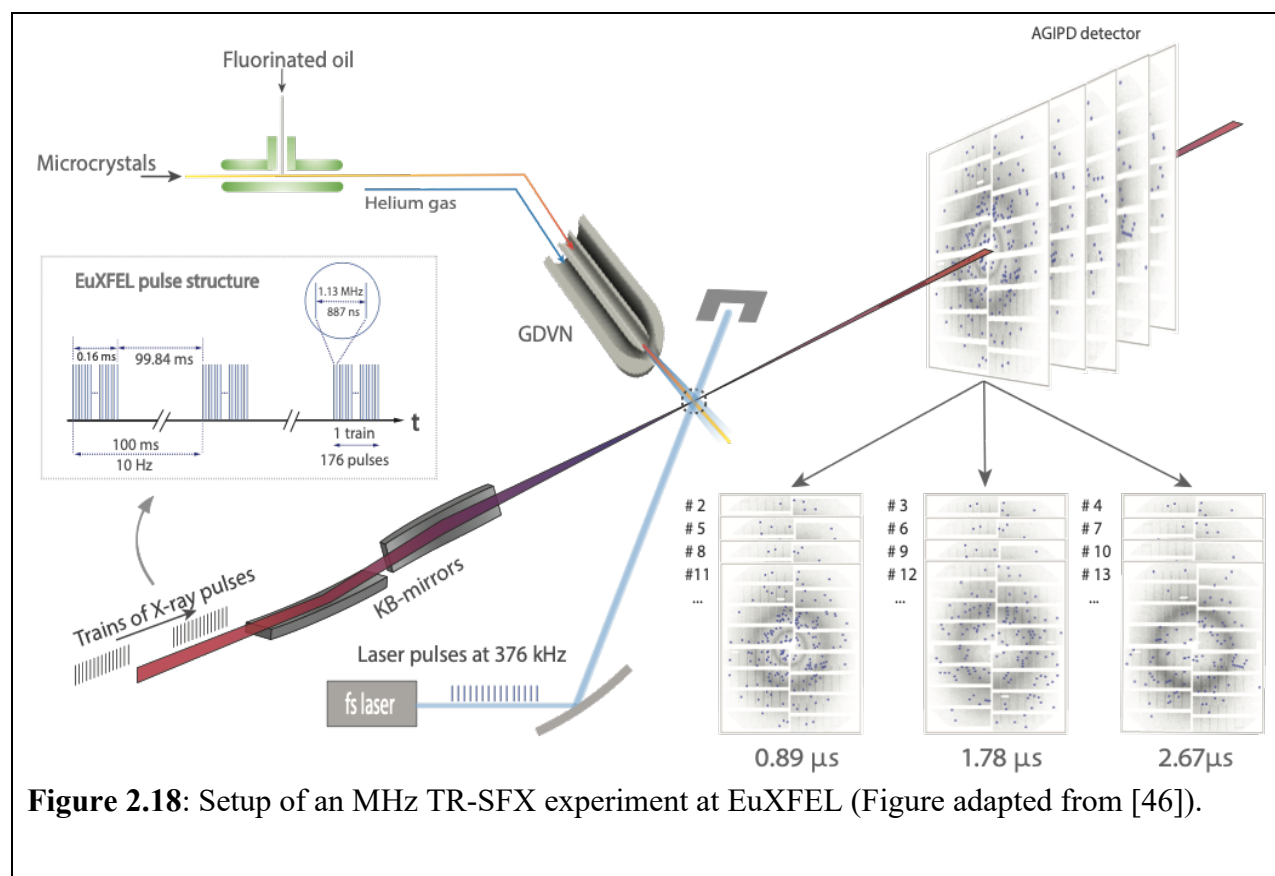
Experiment Station	Date	Principal Investigator	Description	Role*
LCLS	Dec, 2015	Prof. Dr. Marius Schmidt	Structure-Based Enzymology at the LCLS.	DC, DA
LCLS	January, 2016	Prof. Dr. Sebastian Westenhoff	Phytochrome study by picosecond time-resolved SFX.	DC, DA
LCLS	September, 2016	Prof. Dr. Marius Schmidt	Structure-Based Enzymology at LCLS.	DC, DA
LCLS	November, 2017	Prof. Dr. Marius Schmidt	Z/E Isomerization in Bacteriophytochromes.	DC, DA
EuXFEL	March, 2019	Prof. Dr. Marius Schmidt	Time-Resolved Serial Crystallography with MHz Data Collection Rates.	DC, DA, SP
SACLA	June, 2019	Prof. Dr. Marius Schmidt	Time-Resolved Macromolecular Structure Determination of Phytochromes.	DC, DA
EuXFEL	March, 2020	Prof. Dr. Marius Schmidt	Time-Resolved Enzymology to Capture a Full Catalytic Series.	DA, SP

\*DC – Data Collection, DA – Data Analysis, SP – Sample Preparation

#### 2.7.4.2 Time-Resolved SFX at EuXFEL

In photoactive molecules, a reaction can be initiated by a brief laser flash, called a ‘pump’ pulse. After an adjustable time-delay ( $\Delta t$ ), the crystals are illuminated by an X-ray pulse, called a ‘probe’ pulse, which generates a diffraction pattern. This is called a ‘pump-probe’ experiment. PYP has been previously studied by TRX experiments at both synchrotrons and XFELs [41, 42, 44, 116]. Here, PYP has been used to establish TR-SFX at the EuXFEL. For this TR-SFX experiment, PYP microcrystals need to be grown, and a suitable laser setup is necessary. The laser

setup must be synchronized with the X-ray pulses. The experimental design using the MHz pulse structure of the EuXFEL is shown in Fig. 2.18.

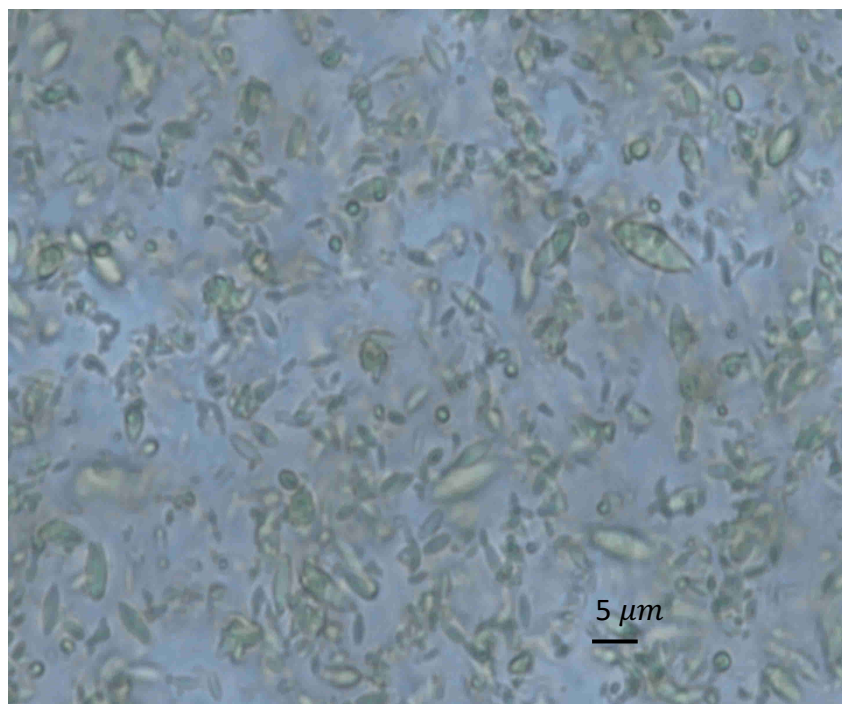


**Figure 2.18:** Setup of an MHz TR-SFX experiment at EuXFEL (Figure adapted from [46]).

### 2.7.4.3 Sample Preparation

PYP was overexpressed and purified as reported [41]. PYP microcrystals were grown with Na-malonate (pH 7.0) as a precipitant. PYP solution was concentrated to 100 mg/mL and 4.0 M Na-malonate pH 7.0 was added once to a final concentration of 3.3 M. To obtain microcrystals, the mixture was vigorously stirred for 8 hours and matured overnight. With this method crystals with sizes between 5 and 10  $\mu\text{m}$  in diameter were obtained. The resulting microcrystals slurry were centrifuged at 8,000  $g$  for 10  $\text{min}$ . After centrifugation, two distinct layers formed a dense cloudy yellow bottom layer and a clear top layer. The top layer was removed, and the remaining

mixture was mixed with 2.8 M Na-malonate, pH 7. This was done to avoid further growth of the crystal. The crystals before centrifugation are shown in Fig. 2.19. Crystals were then filtered subsequently through 20  $\mu\text{m}$  and 10  $\mu\text{m}$  stainless steel filters. The crystals are then injected into the serial crystallographic instrument using GDVN to collect the diffraction patterns.



**Figure 2.19:** PYP microcrystals. Crystalline suspension in Neubauer counting chamber.

#### 2.7.4.4 Instrumentation

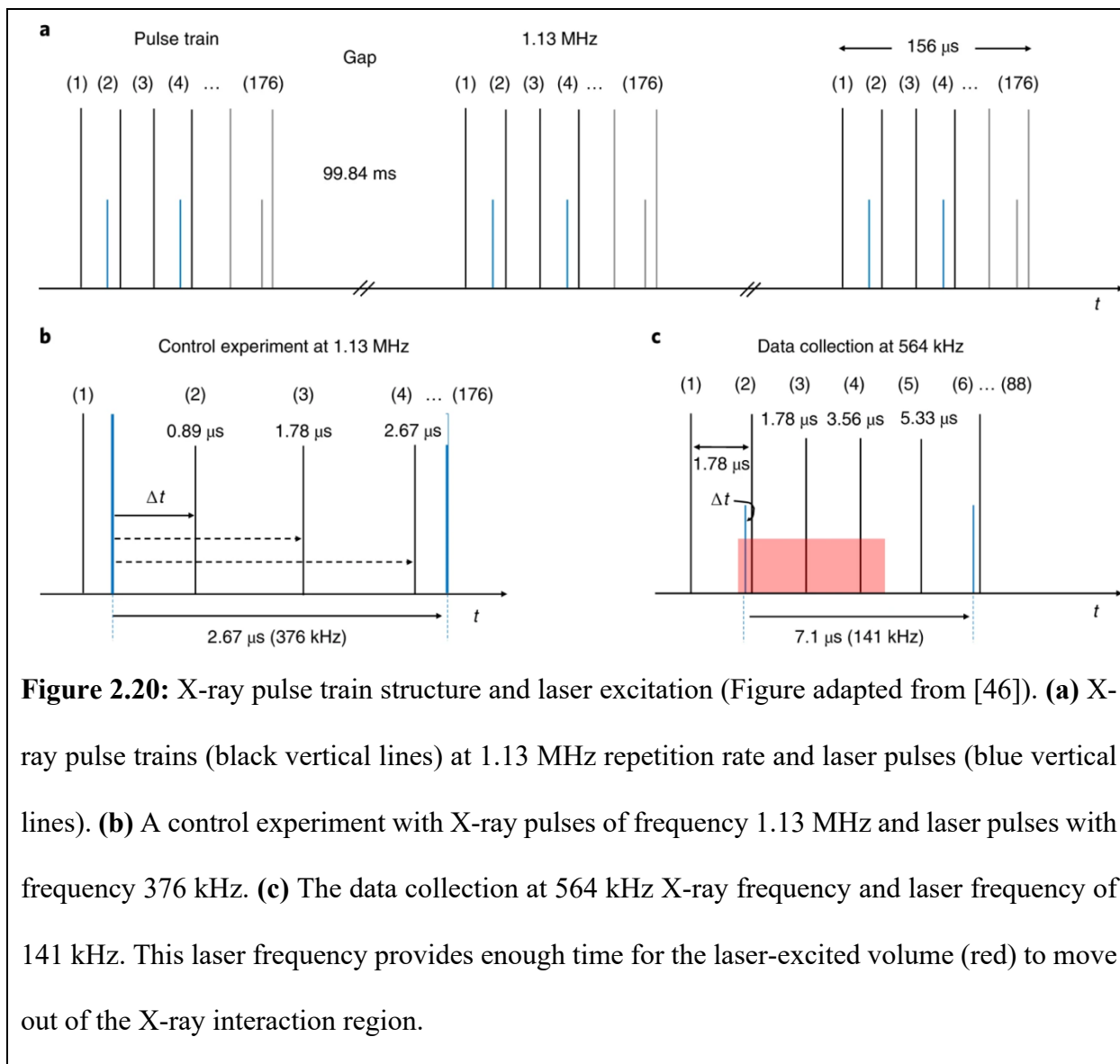
A TR-SFX experiment was conducted on PYP at the newly operational, high repetition rate XFEL, the EuXFEL over 5 twelve-hour shifts. The EuXFEL uses superconducting technology to generate unprecedentedly high X-ray pulse rates. X-ray pulses arrive in pulse-trains that contain up to 2700 pulses (final design specification). Pulses within the train repeat with a rate of up to 4.5 MHz. There are 10 trains per second. Each pulse train is 600  $\mu\text{s}$  long with 99.4 ms gaps between the trains. In the first SFX experiment at EuXFEL with the MHz rate, there were only 15 pulses per train, yet datasets were successfully collected on several proteins [117]. The number of

available pulses increased steadily. 178 pulses/train were available for the PYP TR-SFX experiment.

The experiment aimed to cover the previously unexplored picosecond time range of the photocycle (Fig. 1.5 (b)) and establish TR-SFX at MHz repetition rates XFEL. The experiment was carried out at 9.3 keV X-ray energy in a pump-probe manner. The X-ray pulse energy corresponds to 700  $\mu\text{J}$ . The KB-mirror system focuses the X-ray beam to a 2-3  $\mu\text{m}$  focal spot as shown in Fig. 2.18. The reaction was initiated using laser pulses of 240  $\text{fs}$  at a wavelength of 420  $\text{nm}$  with a flux density of 1.6  $\text{mJ}/\text{mm}^2$  in a 42  $\mu\text{m}$  (full-width half-maximum, FWHM) focal spot. A GDVN was used to flow the sample to the X-ray interaction region as shown in Fig. 2.18. Two types of X-ray and laser pulse sequences were used for the control experiment and for the data collection. For the control experiment, an X-ray repetition rate of 1.13 MHz and optical laser pulses rate of 375 kHz repetition was used. This resulted in a scheme with the pump-probe sequence and 2 interleaved X-ray pulses in the dark without laser excitation (Light-dark-dark scheme). This was repeated as depicted in Fig. 2.20. The two interleaved X-ray pulses correspond to time delays of 1.78  $\mu\text{s}$  and 2.67  $\mu\text{s}$  respectively. This sequence was used to determine when and whether the laser-excited jet volume had passed the X-ray interaction region. For the  $\text{ps}$  data collection, repetition rates of X-ray pulses and the optical laser pulses were reduced to 564 kHz and 141 kHz respectively. The laser excitation occurs before every fourth X-ray pulse, as shown in Fig. 2.20. Diffraction patterns were collected using an AGIPD [101] of 1 megapixel located 117.7-118.6  $\text{mm}$  downstream of the sample interaction region as shown in Fig. 2.18. The AGIPD's readout rate matches the X-ray pulse rate in the trains. AGIPD reads 352 frames per train or 3520 frames per second [101]. The data were pre-processed using Cheetah [103] and indexed using CrystFEL [106] as explained in



section 2.6.3. The dark and light data files were sorted using the cell ID information stored in the CrystFEL output file (stream file) as shown in Fig 2.26. These files then can be used by CCP4 programs to generate difference electron density maps.



## 3. Results

### 3.1 Simulation Results of a Phytochrome Molecule

#### 3.1.1 Simulation Strategy

The simulations were done for the full-length *Idiomarina sp.* phytochrome molecule (Fig. 1.3 (a)) whose structure was recently determined [34]. This structure has an approximate diameter  $D$  of 164 Å and consists of about 11,000 atoms. Diffraction patterns were simulated according to the formalism described in section 2.4.2 and implemented in Python.

The phytochrome molecule was illuminated by a simulated X-ray free-electron laser with a wavelength of 5.0 Å corresponding to a photon energy of 2.48 keV. The resolution at the edge of the detector was 10 Å. The phases needed to recover the electron density can be retrieved by sampling the continuous diffraction pattern at sufficiently small intervals in reciprocal space. To retain the phase information, these intervals must be smaller than  $1/(2D)$  [14, 118], i.e. oversampled at least twice with respect to the molecular diameter (detail is explained in section 2.5.4.3). This determines the size of the Shannon pixels in the diffraction pattern. For phytochrome,  $73 \times 73$  detector pixels are required to reach a resolution of 10 Å. The simulated signal for each detector pixel was converted to the expected number of photons for an incident X-ray fluence of  $10^{20}$  photons/cm<sup>2</sup> achievable at an XFEL. For an X-ray focal spot of  $1\mu m$ , this corresponds to a pulse energy of  $320\mu J$ . The measured photon counts follow Poisson statistics [66]. Accordingly, diffraction patterns were simulated by adding Poisson noise ('shot noise') to the calculated diffraction signal. To evaluate whether the near-atomic resolution could be

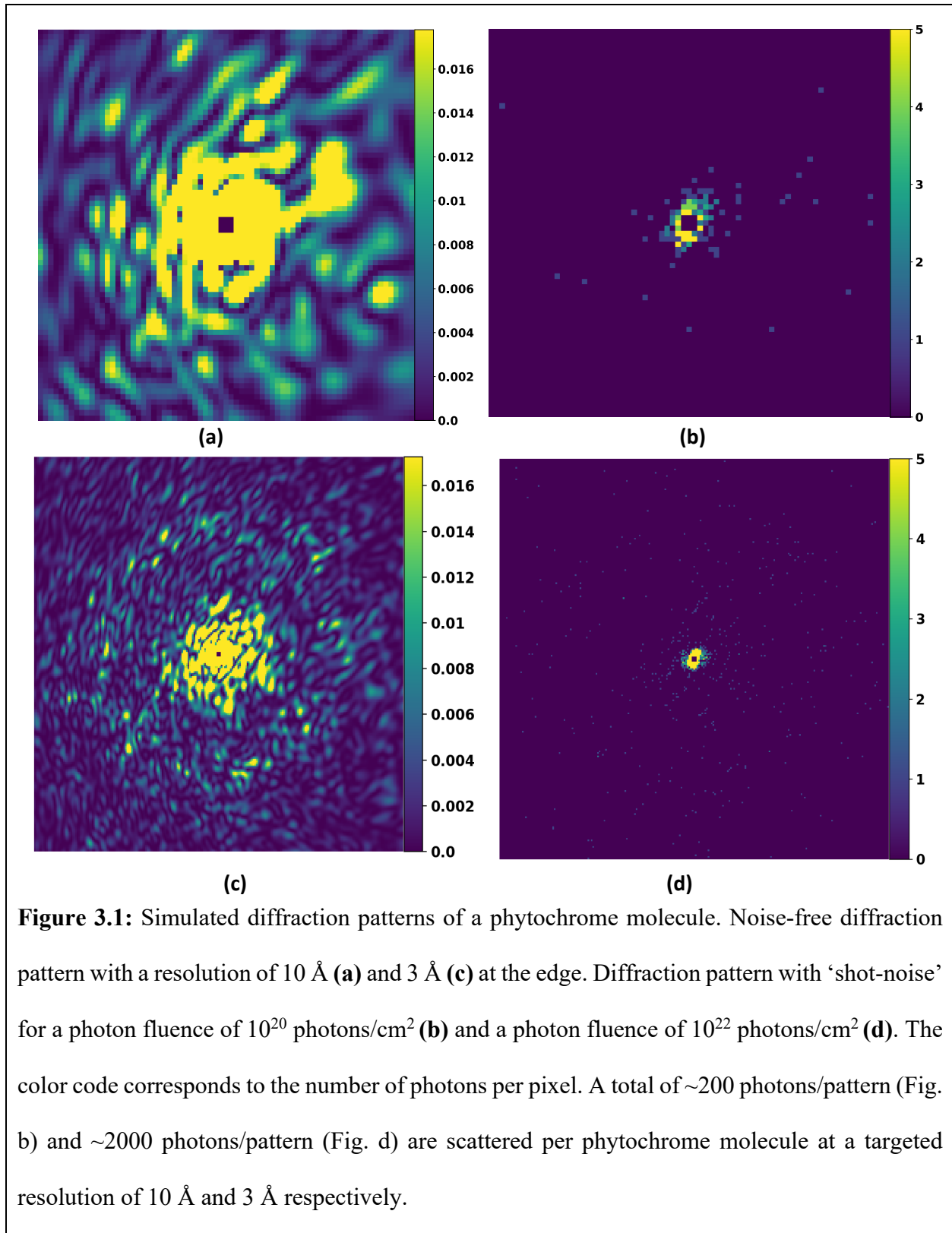
realistically reached by modern X-ray sources, the simulation was also done for the target resolution of 3 Å. For this, higher photon energy of 8.27 keV is used instead, which corresponds to a wavelength of 1.5 Å. For a focal spot of 1  $\mu\text{m}$ , there are only 20 photons per pattern so a smaller X-ray focal spot with 0.1  $\mu\text{m}$  diameter is chosen which yields a photon fluence of  $1 \times 10^{22}$  photons/cm<sup>2</sup>. Randomly oriented diffraction patterns were generated using uniform random rotation quaternions as described in section (2.4.3). The required number of diffraction snapshots, using our formalism (Eq. (3.8)), at different experimental conditions to reach a target resolution for a SNR 1.0 is tabulated in Table 3.1.

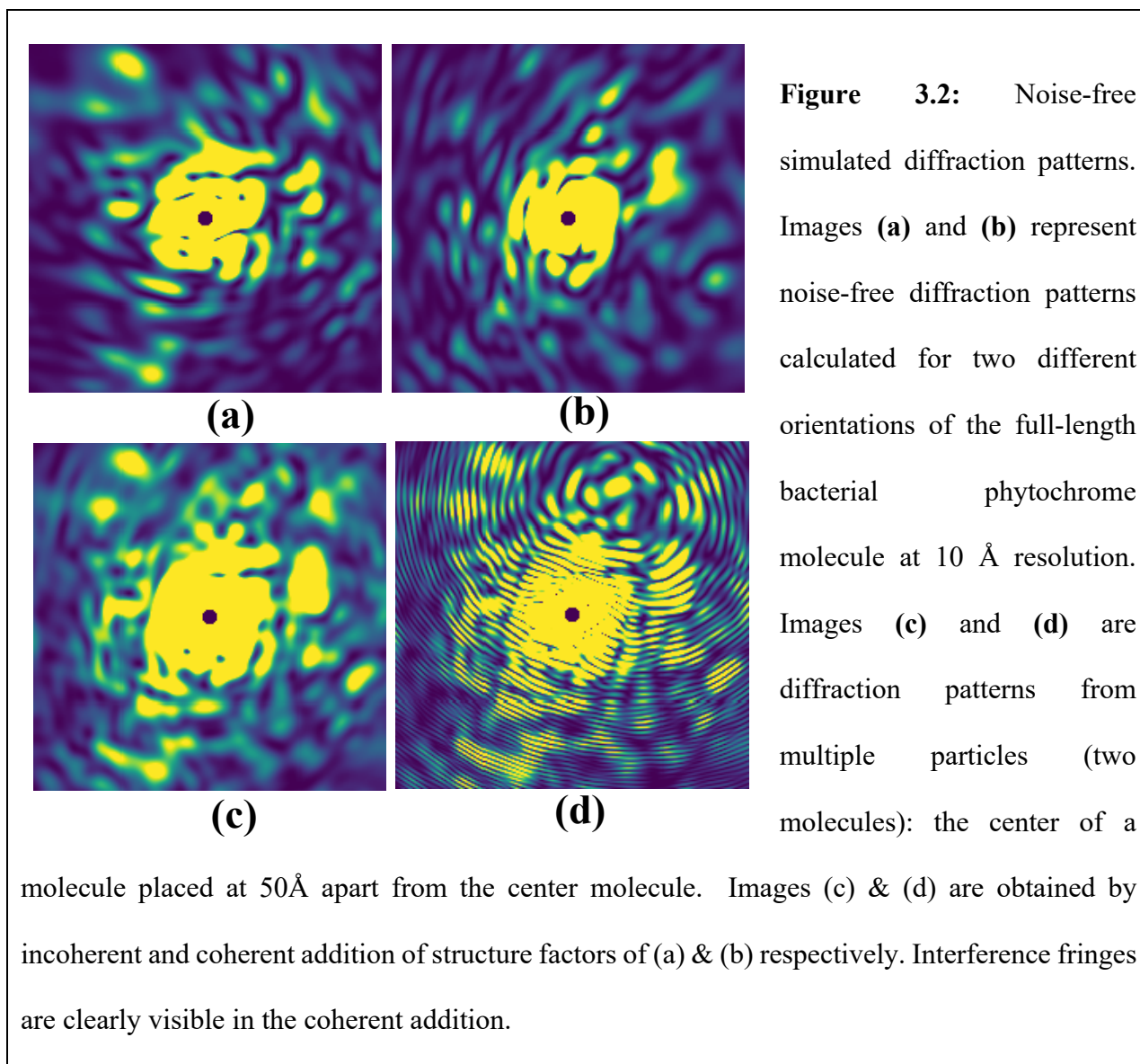
### 3.1.2 Simulated Diffraction Patterns

Simulated single-molecule diffraction patterns of full-length phytochrome are shown in Fig. 3.1. As discussed in section 2.7.3.2 there is also a probability for multiparticle exposures. A noise-free multiple particles diffraction snapshot is shown in Fig. 3.2.

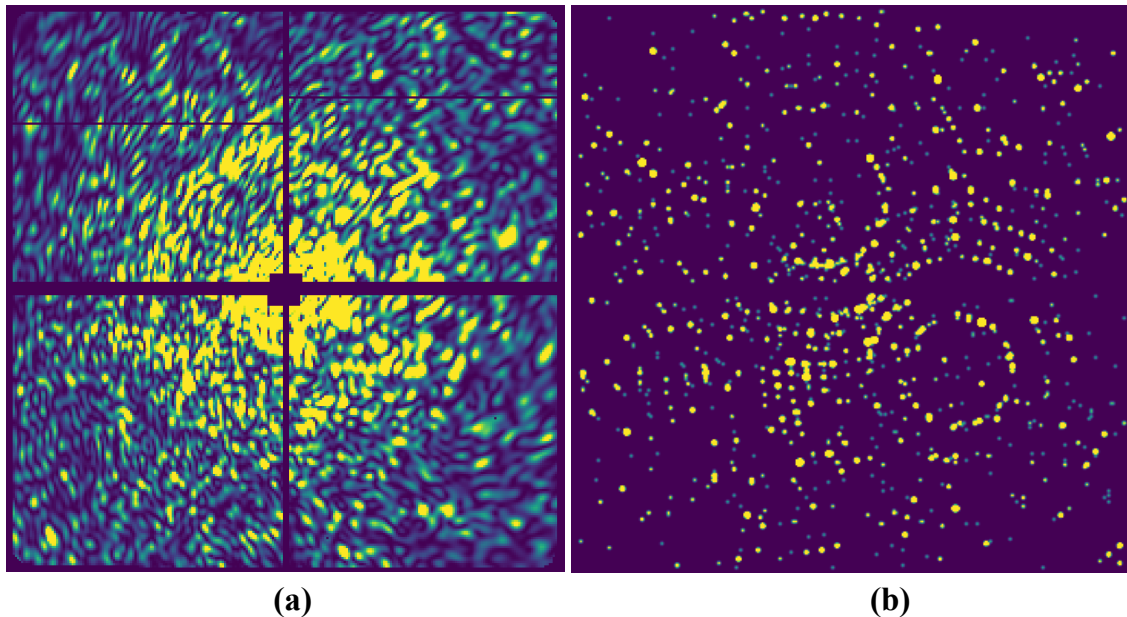
As it is interesting to see the difference to a diffraction pattern from a small crystal, a simulated crystal diffraction pattern is shown in Fig. (3.3). For small crystal, the intensity can be calculated as the product of the lattice factor for a limited number of unit cells, and the squared structure factors of the unit cell as outlined in references [106, 111].

$$I(\mathbf{q}) = \Phi r_e^2 \frac{\text{Sin}^2(\pi N_a \mathbf{q} \cdot \mathbf{a})}{\text{Sin}^2(\pi \mathbf{q} \cdot \mathbf{a})} \frac{\text{Sin}^2(\pi N_b \mathbf{q} \cdot \mathbf{b})}{\text{Sin}^2(\pi \mathbf{q} \cdot \mathbf{b})} \frac{\text{Sin}^2(\pi N_c \mathbf{q} \cdot \mathbf{c})}{\text{Sin}^2(\pi \mathbf{q} \cdot \mathbf{c})} |\mathbf{F}(\mathbf{q})|^2 d\Omega. \quad (3.1)$$





The crystal is modeled as a parallelepiped with a  $N_a$ ,  $N_b$  and  $N_c$  number of unit cells along the  $a$ ,  $b$ , and  $c$  directions respectively. The maximum intensity scales with the square of the number of periods  $N_a^2 N_b^2 N_c^2$  in the three directions  $a$ ,  $b$  and  $c$ . Intrinsic crystal disorder was modeled by a Debye-Waller (Eq. (2.10)).



**Figure 3.3:** Noise-free diffraction patterns of a single-molecule (a) and a crystal (b) at 3 Å resolution). The intensity is continuously distributed on the single-molecule diffraction pattern whereas it is concentrated in Bragg peaks for a crystal. The pulse fluence was chosen to be  $10^{12}$  photons focused to a  $1\mu\text{m}$  spot, with a photon of wavelength 2.5 Å. For a simulation of the crystal diffraction pattern, the number of unit cells along *a*, *b*, and *c* directions is 50. Every atom was assigned the same B-factor of 40 Å<sup>2</sup>. The simulated patterns are shown on a simulated pnCCD detector [119].

**Table 3.1:** Estimated number of snapshots required to reach SNR = 1 at different resolutions for the full-length phytochrome molecule at various experimental conditions.

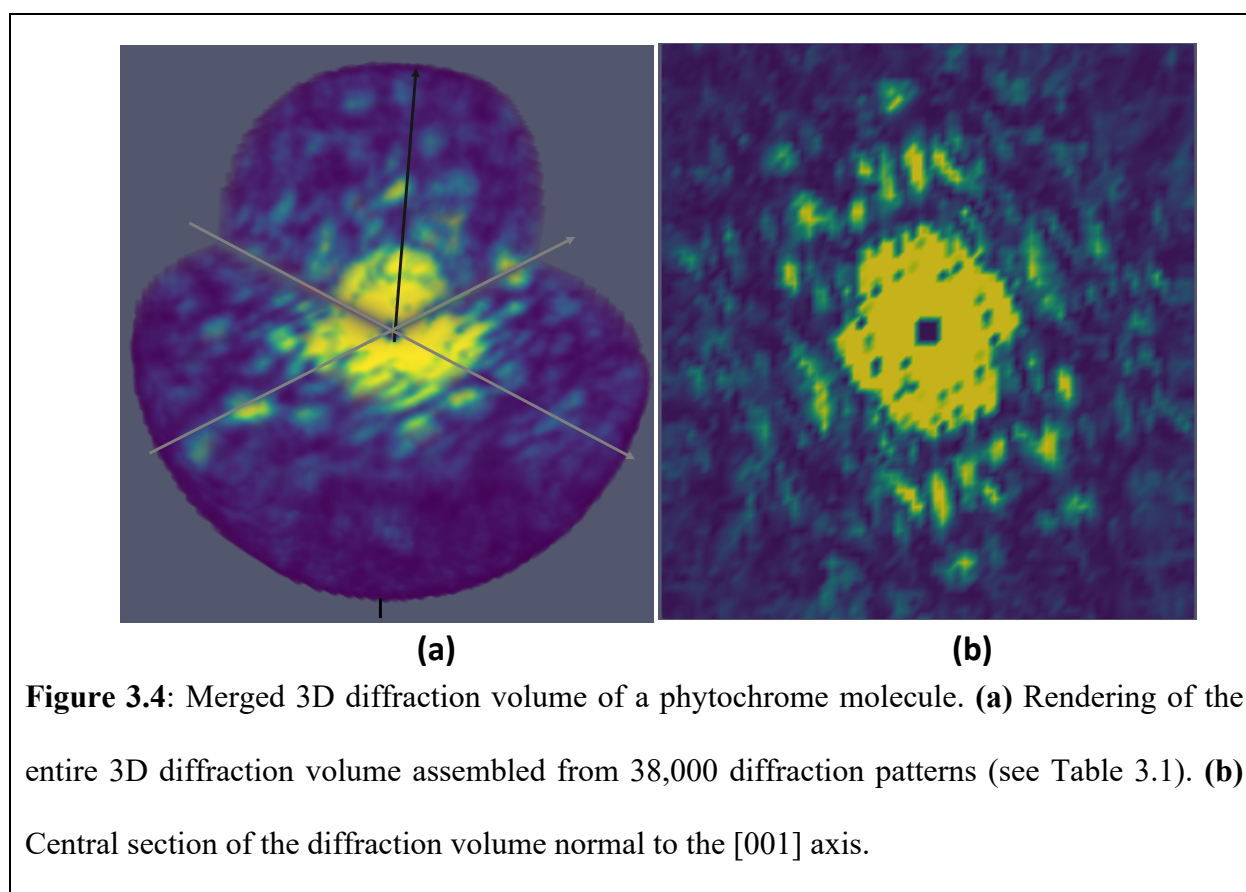
Resolution	Soft X-rays Energy = 2.48 keV; $\lambda = 5 \text{ \AA}$ Beam size = $1.0\mu\text{m} \times 1.0\mu\text{m}$ Fluence = $10^{20}\text{ph}/\text{cm}^2$	Hard X-rays Energy = 8.0 keV; $\lambda = 1.5 \text{ \AA}$ Beam size = $0.1\mu\text{m} \times 0.1\mu\text{m}$ Fluence = $10^{22}\text{ph}/\text{cm}^2$	Hard X-rays (Larger Beam) Energy = 6.0 keV; $\lambda = 2.07 \text{ \AA}$ Beam size = $0.5\mu\text{m} \times 0.5\mu\text{m}$ Fluence = $4.0 \times 10^{20}\text{ph}/\text{cm}^2$

	$\langle n \rangle$	# Snapshots	$\langle n \rangle$	# Snapshots	$\langle n \rangle$	# Snapshots
30 Å	$1.8 \times 10^{-2}$	1,774	$1.0 \times 10^{-1}$	488	$8.4 \times 10^{-3}$	3,394
25 Å	$6.7 \times 10^{-2}$	5000	$4.8 \times 10^{-2}$	1,007	$3.8 \times 10^{-3}$	8,302
10 Å	$2.0 \times 10^{-3}$	38,244	$2.6 \times 10^{-2}$	4,323	$2.2 \times 10^{-3}$	35,063
5 Å	*	*	$6.6 \times 10^{-3}$	26,978	$5.0 \times 10^{-4}$	284,978
3 Å	*	*	$4.3 \times 10^{-3}$	66,075	$3.5 \times 10^{-4}$	672,010

$\langle n \rangle$  is the mean number of photons per Shannon pixel at the desired resolution  $d$ .

\*Not accessible due to wavelength or unpractically high scattering angle.

### 3.1.3 Merged 3D Diffraction Volume



The ensemble of all simulated snapshots is merged into a diffraction volume using the ‘cone-gridding’ algorithm (section 2.6.3). The 3D diffraction volume of the molecule was obtained by orienting the noisy diffraction patterns relative to each other. The merged 3D volume is shown in Fig. 3.4. In the simulation, the central three voxels of the diffraction volume were set to zero, which considers the experimentally inaccessible central area of the detectors used at the XFELs.

### 3.1.4 Phasing and its Validation

The electron density from the diffraction volume was recovered using a combination of the hybrid-input-output HIO [80] and the shrink-wrap algorithm [82] as explained in section 2.6.4.4. The HIO algorithm was applied for the first fifty iterations with the feedback parameter  $\beta = 0.9$ . After this, the shrink-wrap algorithm was used with an adaptive support constraint determined anew for each iteration cycle. For the shrink-wrap process, the electron density was convoluted with a Gaussian of width  $\sigma$ . The initial width  $\sigma$  was set to 6 voxels and reduced by 5% after each iteration until a minimum of 1 voxel was reached. Voxels that contain electron densities larger than 14% of the maximum were assigned to the new support constraint. This algorithm converged after a few hundred iterations. The reconstructed electron density is shown in Fig. 3.5.

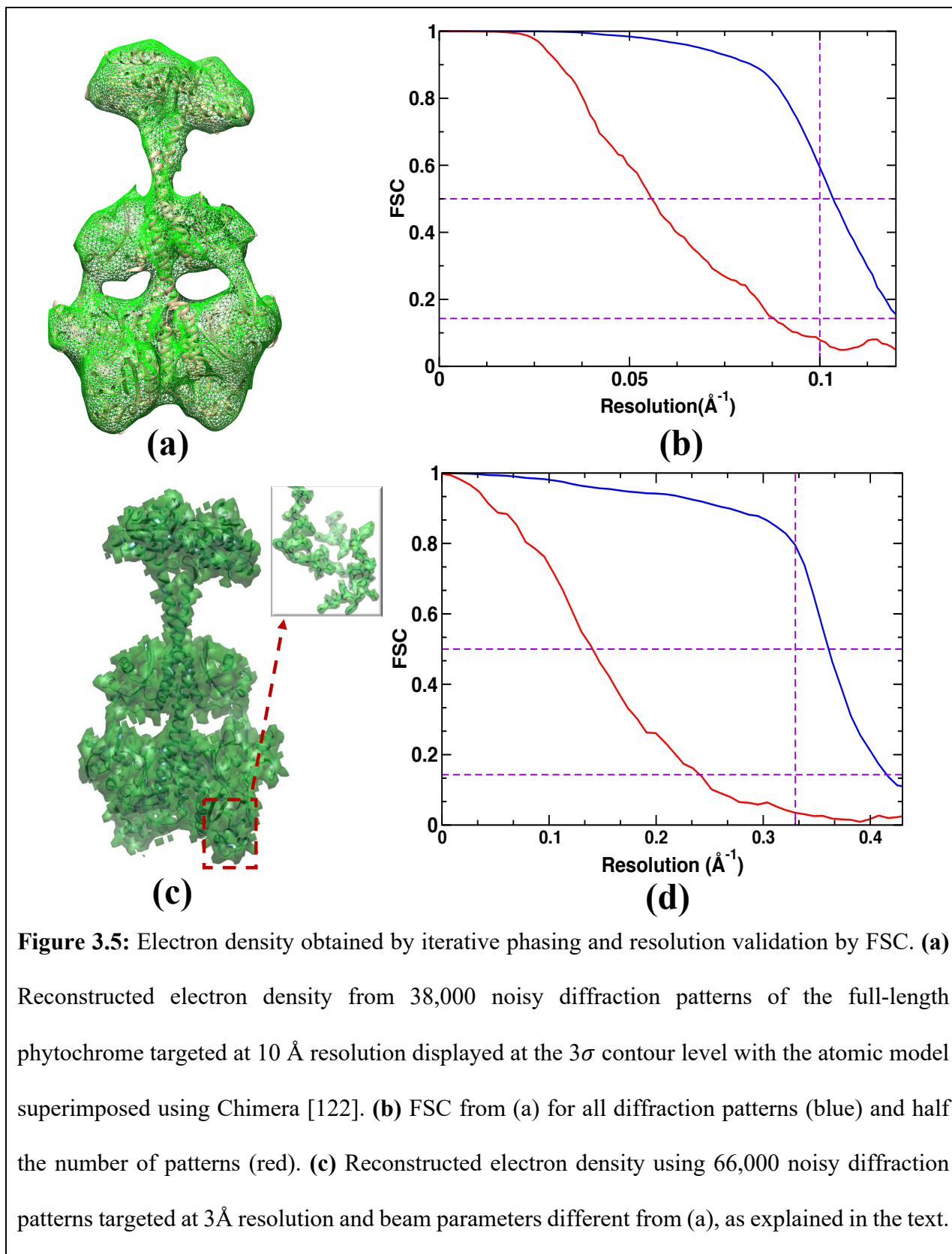
The resolution and reproducibility of the reconstructed electron density were accessed using the Fourier Shell Correlation (FSC). For this, the diffraction patterns are split randomly into two disjoint sets ‘1’ and ‘2’ which were processed independently, resulting in two electron densities maps. The FSC is calculated from the Fourier transform of the two maps by

$$FSC(\mathbf{q}) = \frac{\sum_{\mathbf{q}} F_1(\mathbf{q}) \cdot F_2^*(\mathbf{q})}{\sqrt{\sum_{\mathbf{q}} |F_1(\mathbf{q})|^2} \sqrt{\sum_{\mathbf{q}} |F_2(\mathbf{q})|^2}}, \quad (3.13)$$



where  $F_1(\mathbf{q})$  and  $F_2(\mathbf{q})$  are the Fourier transforms of maps '1' and '2', respectively,  $q$  is the magnitude of the scattering vector and  $*$  denotes the complex conjugate. The FSC measures the degree of correlation between the two reconstructions as a function of scattering angle (spatial frequency in Fourier space). The FSC curve starts with a value of 1, indicating excellent agreement at low spatial frequencies and drops gradually to small values indicating poor agreement as shown in Fig. 3.5. The resolution limit is defined by the value where the FSC drops below a certain threshold. Conventional thresholds used by the cryo-EM community are 0.143 and 0.5 [120, 121]. These thresholds are represented by horizontal dashed lines in Fig. 3.5. The FSC drops below 0.5 at around 10 Å (Fig. 3.5 (b), blue line). This demonstrates that the number of diffraction patterns estimated by Eq. (2.24), with  $\tilde{P} = 0.5$  and  $\text{SNR} = 1$ , is sufficient to reach the targeted resolution.

To test whether a smaller number of snapshots could be sufficient to obtain the electron density at the same resolution, the reconstruction workflow is repeated with half the number of patterns (19,000), which corresponds to a SNR of 0.67. The FSC (Fig. 3.5b, red line) reveals that instead of 10 Å only about 20 Å is reached in this case. Similarly, the reconstructed electron density from 66,000 patterns at 3 Å is shown in Fig 3.5c. Details of the structure can be identified from the inset of Fig 3.5c. With half the number of patterns (33,000 resulting again in a  $\text{SNR} = 0.65$  in the highest resolution shell), the resolution reached is only about 8 Å (Fig. 3.5d, red line).



**Figure 3.5:** Electron density obtained by iterative phasing and resolution validation by FSC. **(a)** Reconstructed electron density from 38,000 noisy diffraction patterns of the full-length phytochrome targeted at 10 Å resolution displayed at the  $3\sigma$  contour level with the atomic model superimposed using Chimera [122]. **(b)** FSC from (a) for all diffraction patterns (blue) and half the number of patterns (red). **(c)** Reconstructed electron density using 66,000 noisy diffraction patterns targeted at 3 Å resolution and beam parameters different from (a), as explained in the text.

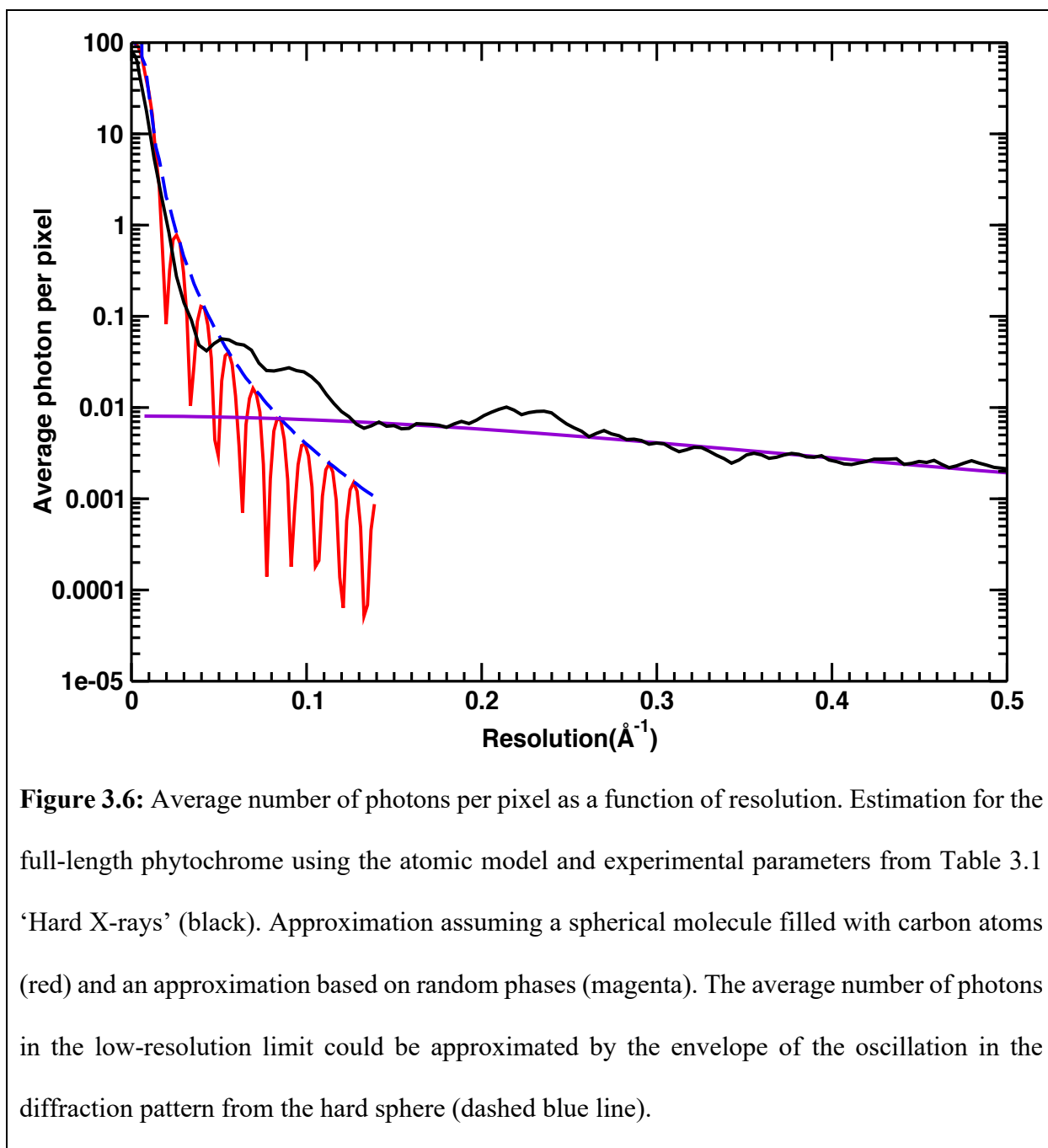
(d) FSC from (c) for all diffraction patterns (blue) and half the number of patterns (red). The horizontal dashed lines show the established threshold for FSC (0.143 and 0.5). The vertical dashed line corresponds to the target resolution.

### 3.1.5 Low- and High-Resolution Approximation

To verify these results, for the low-resolution the spherical approximation is used (Eq. (2.27)) and for high-resolution random phase approximation is used (Eq. (2.28)). The comparison of the value of  $\langle n \rangle$  based on approximation and the quantity obtained from the simulation is shown in Fig. 3.6. For this, all non-hydrogen atoms in the phytochrome molecule are represented by Carbon atoms. The number of carbon atoms  $N_c$  is estimated from the molecular weight  $MW$  as

$$N_c = MW/12 \text{ Da.}$$

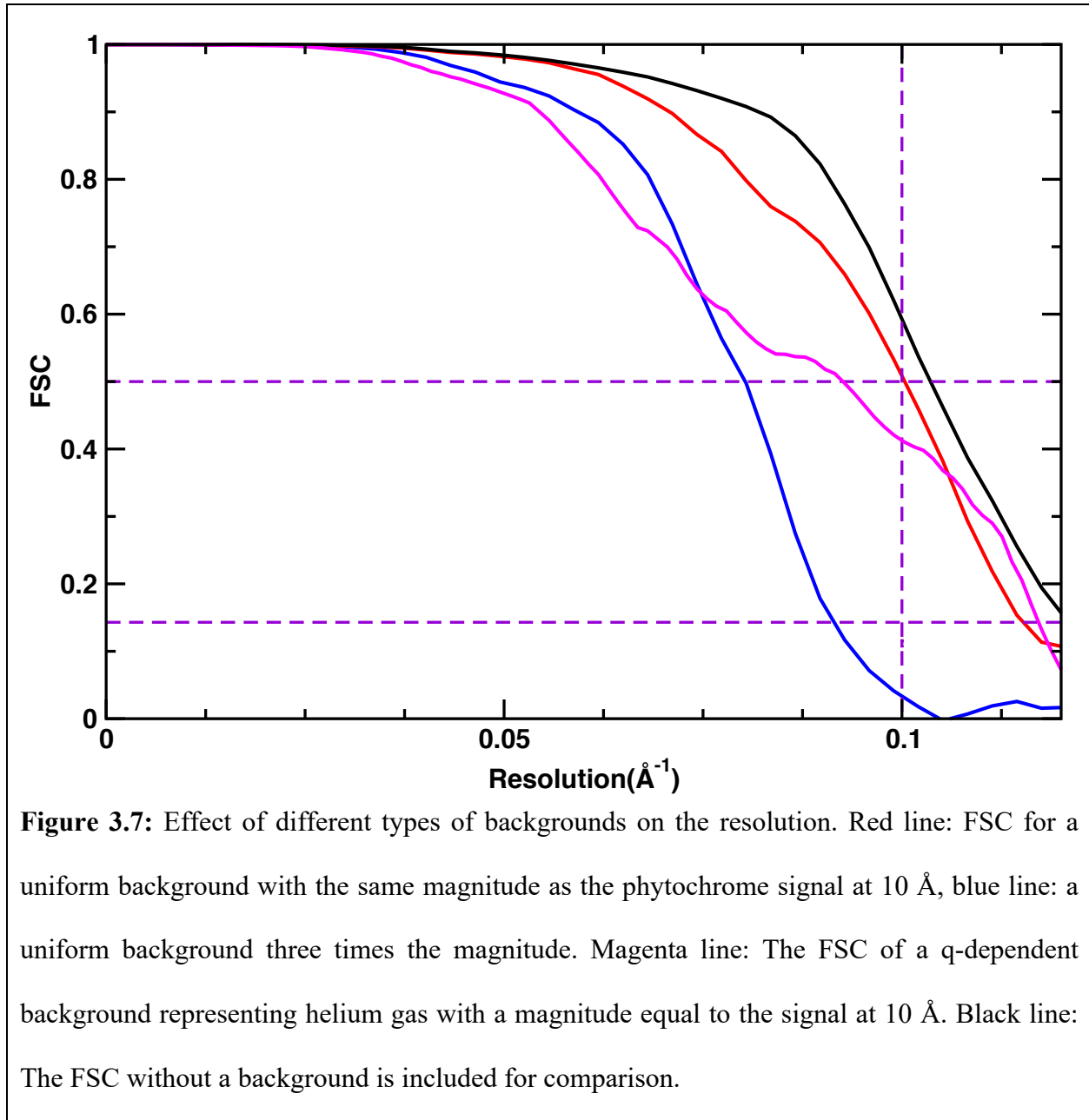
As shown in Fig. 3.6., the low-resolution approximation is reasonable up to a resolution of  $\sim 20 \text{ \AA}$ , whereas the high-resolution approximation is in excellent agreement for sub-nanometer resolution and below.



### 3.1.6 Effect of Background

The background essentially can be classified as an instrument background (including the detector noise), scattering background from the carrier gas (usually Helium gas) and the scattering

from a hydration layer. The first type (instrument noise) is usually small and may be ignored. The second and third types typically give a relevant  $q$ -dependent signal.

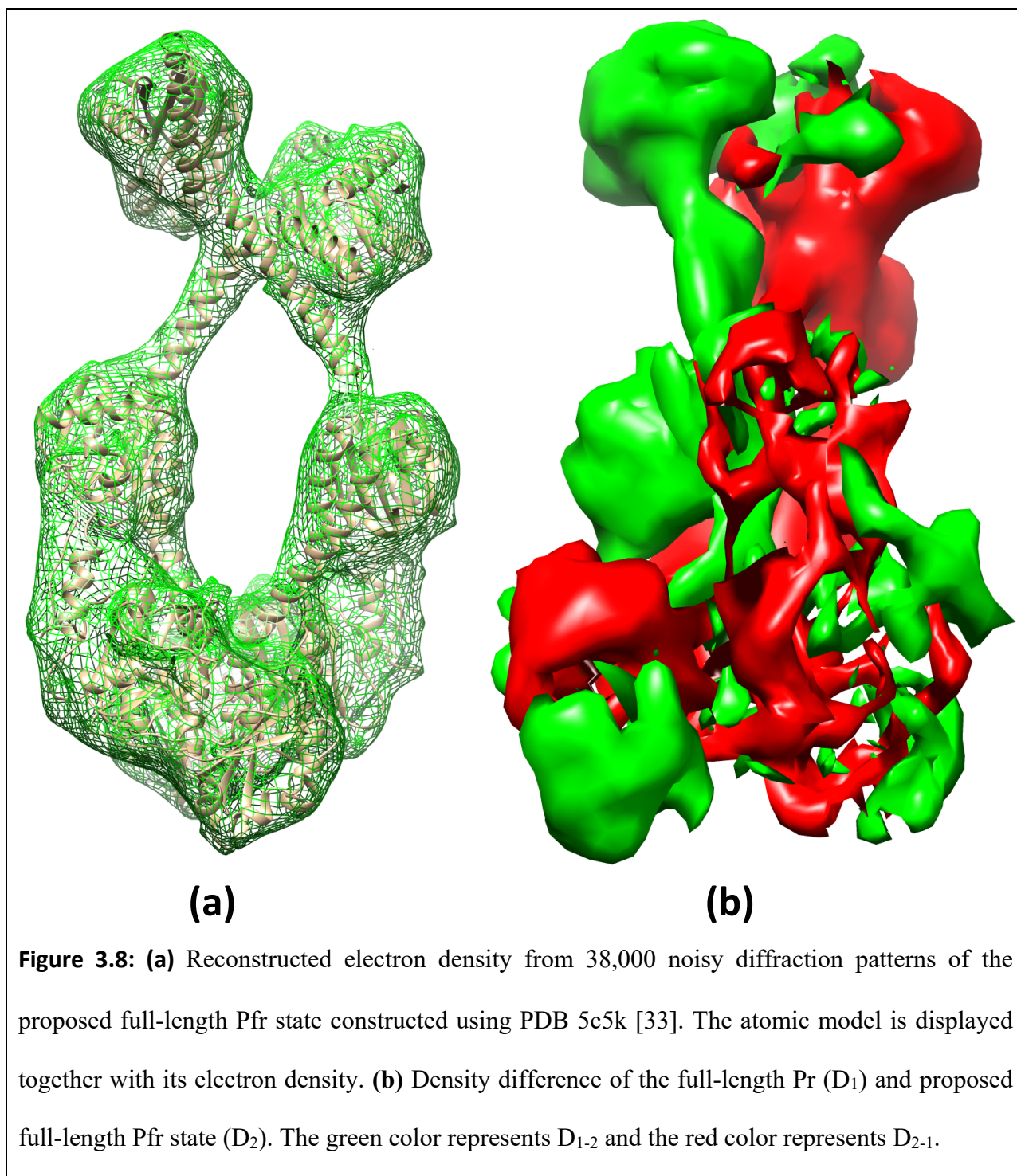


A uniform background is considered with the same and three times the magnitude as the phytochrome signal at 10Å, respectively. Additionally, a q-dependent background of helium gas is assumed, using the atomic scattering factor of *He*. The magnitude of the helium background is set equal to the signal of the phytochrome at 10 Å. The diffraction patterns, including the background, are converted to photon counts by the addition of Poisson noise as described above. The merging and reconstruction processes are repeated to obtain 3D electron density maps. The FSC with a background is shown in Fig. (3.7) and compared to the FSC without any background. Addition of a uniform background (magenta line) has a small effect, and the target resolution of 10 Å can still be reached with 38,000 diffraction patterns. For a background three times the molecular signal, however, the resolution is reduced to 12 Å (blue line). This shows a background that is comparable to the molecular signal can be tolerated as it does not result in substantially more diffraction patterns to reach the target resolution.

### 3.1.7 Modeled Pfr Structure

As the full-length Pfr structure of phytochrome is unknown, a crude structure is assumed. The structure of the photosensory core module (PAS-GAF-PHY) of *D. radiodurans* in the Pfr form (PDB 5c5k [33]) is used as a template. The enzymatic domain of the *Idiomarina* spec. phytochrome (PDB 5llw) is attached to the PHY domain in *D. radiodurans*. The model is shown in Fig. 3.8(a). Noisy diffraction patterns of the proposed Pfr structure were simulated with parameters as explained in section 3.1.1 at a target resolution of 10 Å. Procedures as detailed in sections 3.1.3 and 3.1.4 are used to reconstruct an electron-density shown in Fig 3.8 (a). The electron density of the modeled Pfr state (Fig. 3.8 (a)) is compared to the electron density of the Pr state (Fig. 3.5 (a)). The difference between electron densities is calculated and displayed in Fig.

3.8 (b). This shows that the large conformational changes between the Pr and Pfr states in this enzyme could be examined by using the SPI method.



## 3.2 Simulation Results of a Thiol Decorated Gold Nanoparticles.

### 3.2.1 Simulation Strategy

The diffraction patterns of thiol monolayer protected gold nanoparticle (Fig 1.4) were simulated for a photon energy of 7 keV corresponding to wavelength 1.8 Å with a pulse energy of 2 mJ at the sample and a top-hat beam profile of 200 nm diameter. The simulated diffraction pattern is shown in Fig. 3.9. The resolution at the edge of the detector is 2.0 Å. A total of ~1100 photons/pattern are scattered per gold-thiol molecule out of which ~50 photons are contributed from thiol atoms only. According to our formalism, we need 19,000 snapshots to achieve a target resolution of 2.0 Å at SNR of 1.0. Accordingly, we simulated 19,000 diffraction patterns. The randomly oriented diffraction patterns were generated using uniform random rotation quaternions as described in section (2.4.3).

### 3.2.2 Orientation Determination and Iterative Phasing

The orientation determination of the diffraction snapshots was performed using *Dragonfly* software [72] as discussed in section 2.6.2. Data was provided to *Dragonfly* in photon counts. The photon counts were first converted to the sparse *.emc* format using the script *h5toemc.py* available in the *Dragonfly* software. Orientation determination was performed starting with quaternion sampling (*num\_div*) of 9 and was increased in steps up to a value of 11. The number of rotations is given by  $10 \times (5 \times num\_div^3 + num\_div)$  [70]. Initially the number of rotations was 36,540 and was later increased to 66,660. 160 iterations of the EMC algorithm were performed starting from a random (uniform) model. The oriented diffraction patterns were then subsequently merged

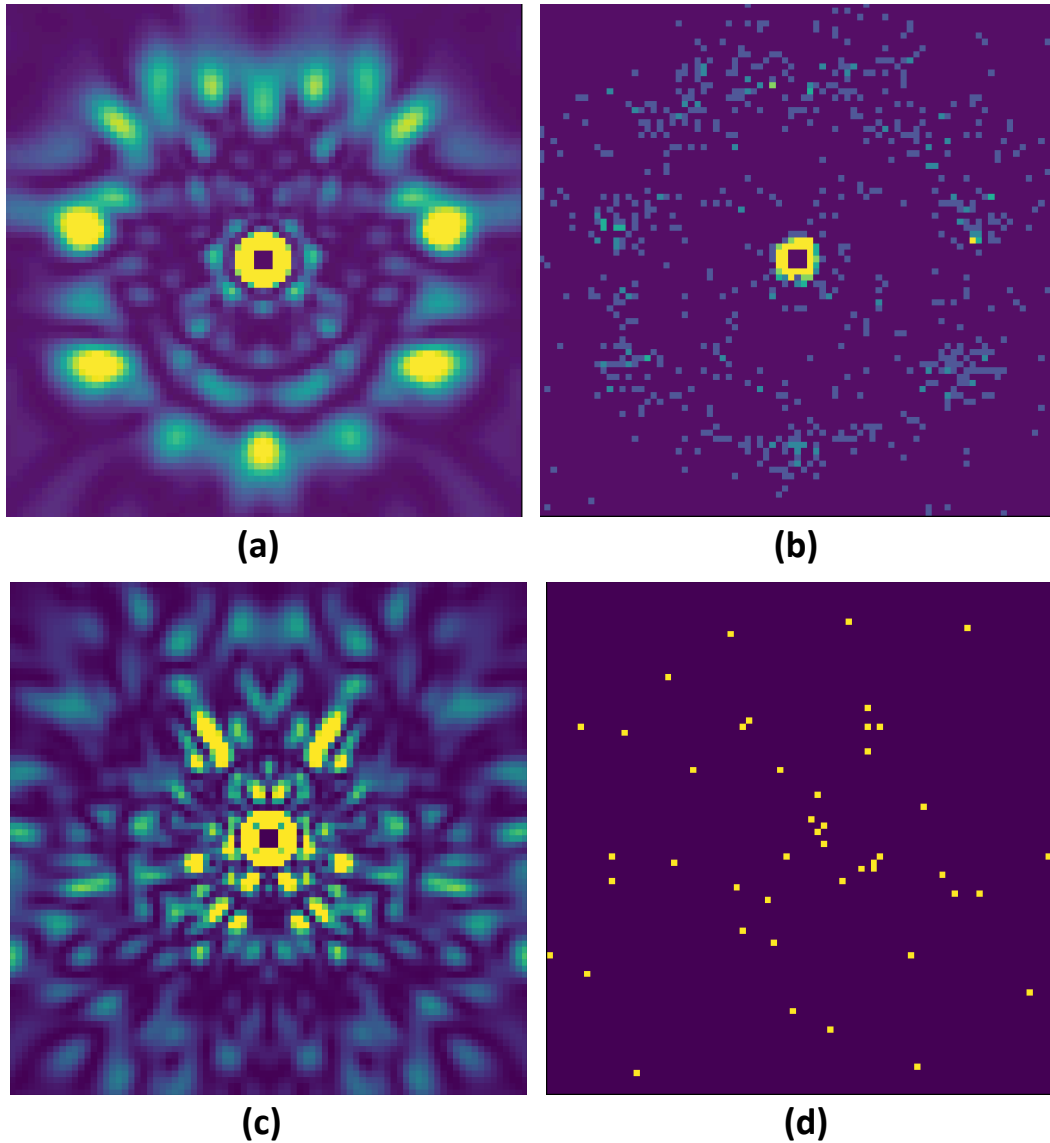


into a 3D diffraction volume as shown in Fig. 3.10. The 3D diffraction volume was phased to generate the real space electron density (Fig. 3.11 (a)) using a combination of HIO and shrink-wrap algorithm. The datasets were split into two halves and independently reconstructed to calculate the ‘gold-standard’ FSC as shown in Fig 3.11 (b). The targeted resolution (vertical dashed line) is within the thresholds used by the cryo-EM community. This demonstrates that the number of diffraction patterns estimated by Eq. (2.24), with  $\tilde{P} = 0.5$  and  $\text{SNR} = 1$ , is sufficient to reach the targeted resolution.

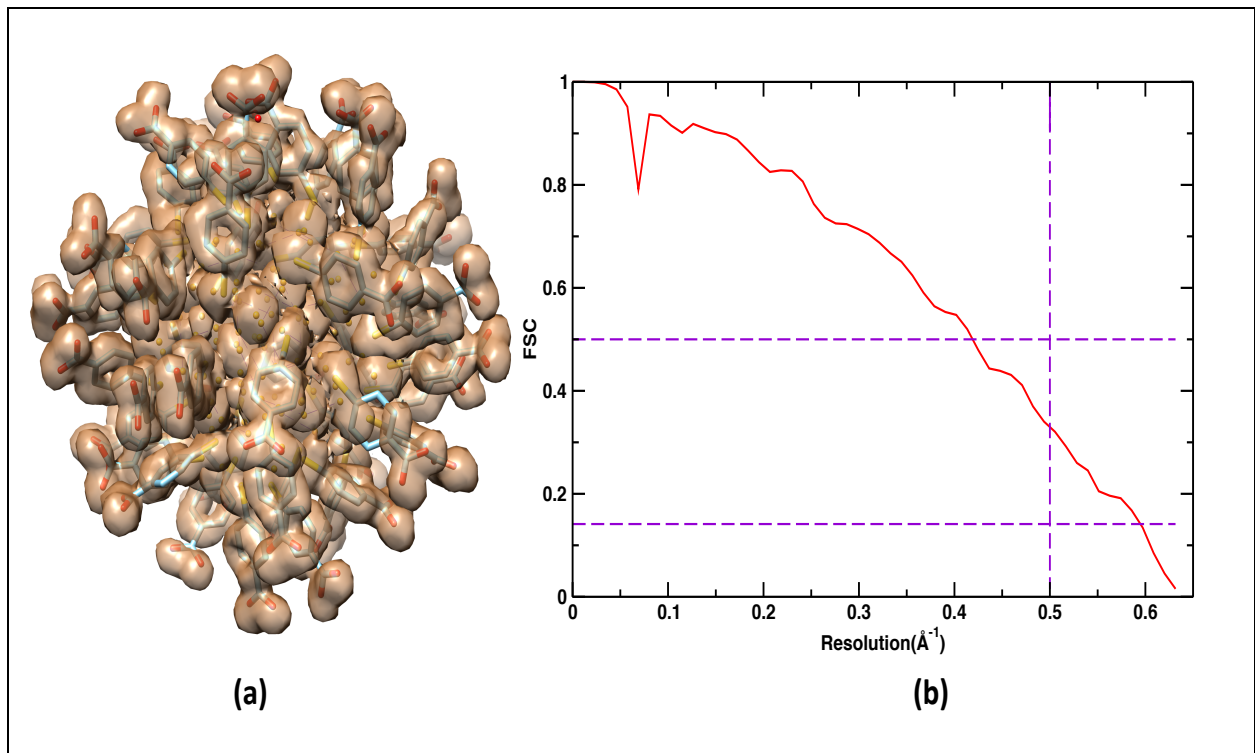
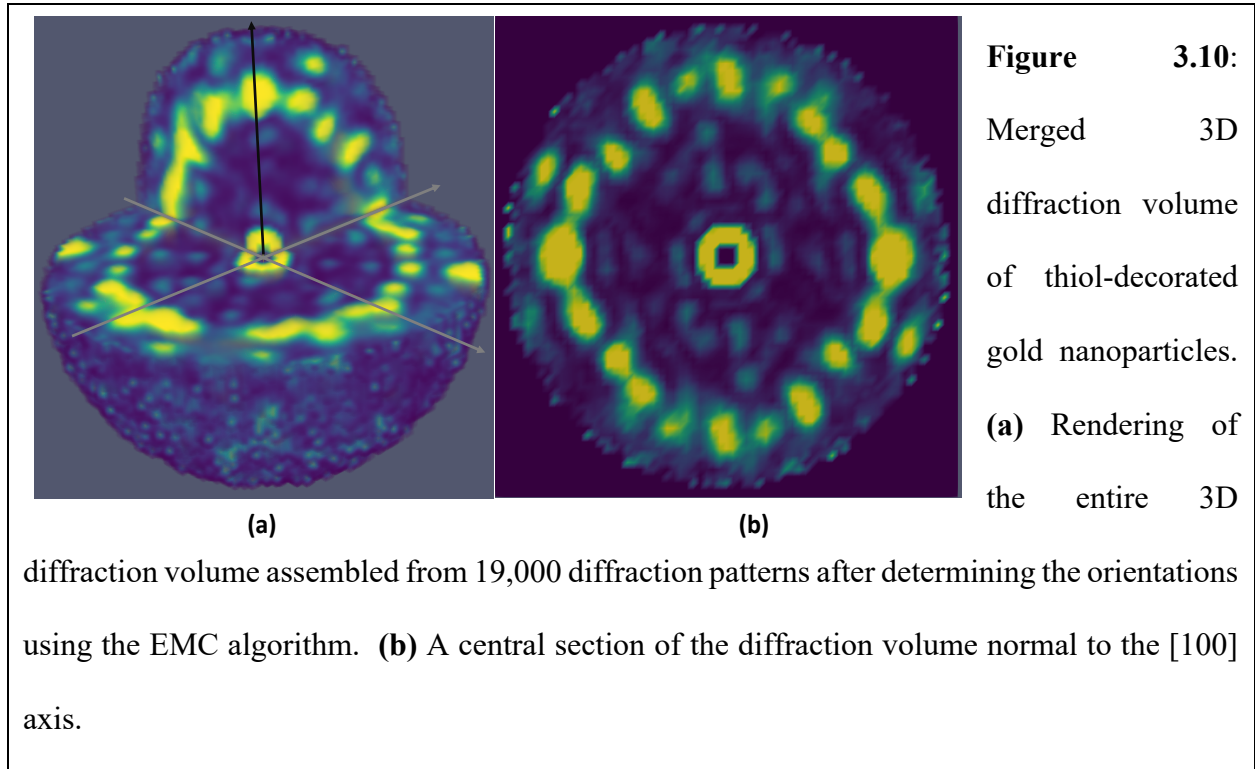
The iterative phase algorithm iterates back and forth between the real and Fourier space. The gold nanoparticles scatter much stronger than the thiol molecules. Because of this, ‘ringing’ or ‘overshoot’ artifacts appear along the edges in real space. To avoid these artifacts the transitions in Fourier space are made smooth using a low pass filter. Butter-worth filter [123] is applied to the merged 3D volume with a cut off at the targeted resolution. Butter-worth filter has the following profile

$$G(q) = \frac{1}{1 + \left(\frac{q}{q_0}\right)^{2k}}, \quad (3.14)$$

where  $k$  is a parameter that determines the order of the filter and  $q_0$  is the cut-off frequency.



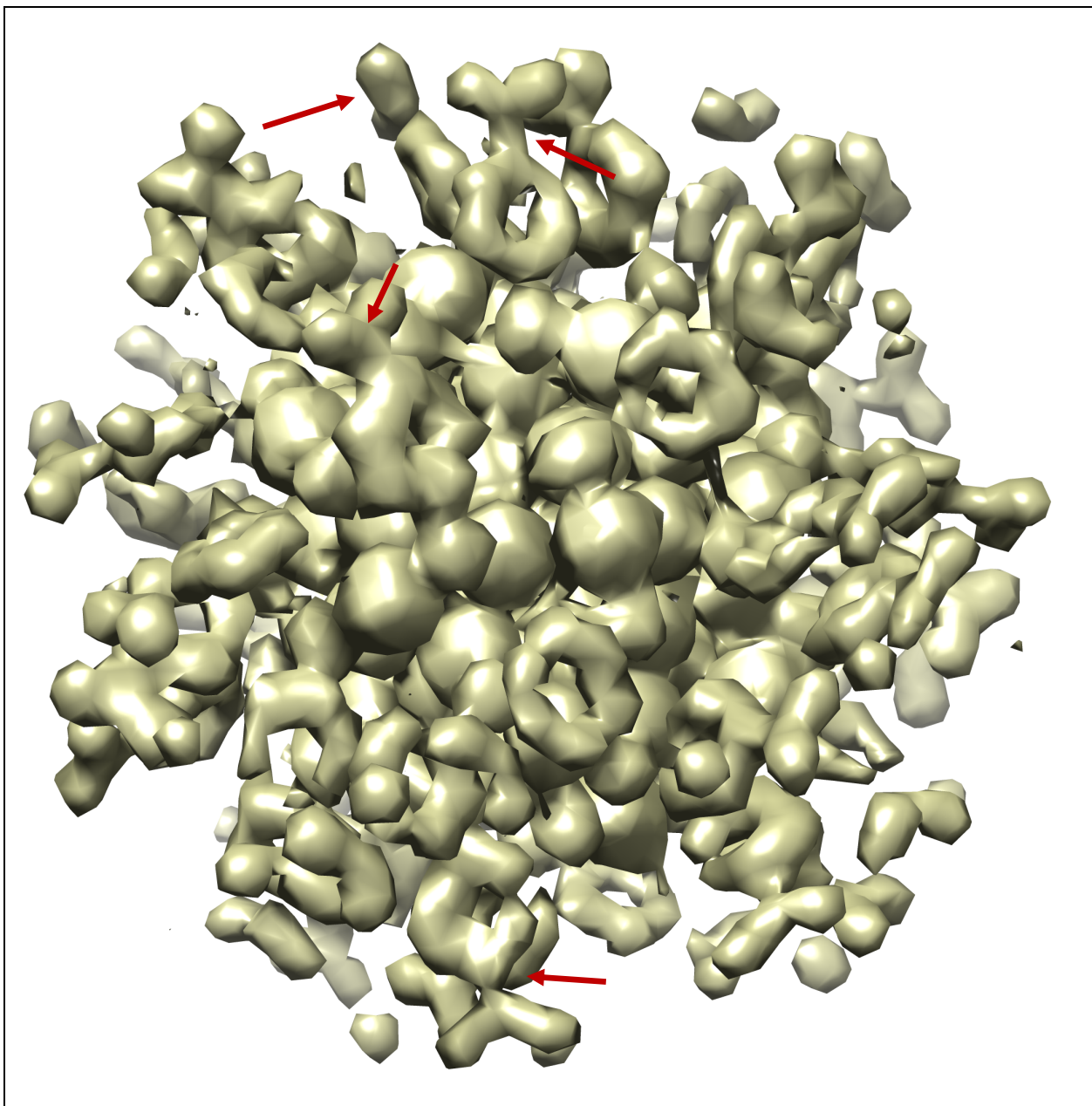
**Figure 3.9:** Simulated diffraction pattern from a thiol-decorated gold nanoparticle. **(a)** Noise-free diffraction pattern with a resolution of  $2.0 \text{ \AA}$  at the edge of the detector. **(b)** Diffraction pattern with the addition of Poisson noise. **(c)** Noise-free pattern of molecules consisting of the thiol part only. **(d)** Diffraction pattern with the addition of Poisson noise. The intensity is converted into photon counts for an X-ray energy of  $7.0 \text{ keV}$ , a  $0.2 \text{ }\mu\text{m}$  diameter top-hat beam, and a  $2 \text{ mJ}$  pulse energy. The total number of photons in the pattern (b) is 1080 of which thiol contributes only 48 (d).



**Figure 3.11:** Electron density and validation of resolution using FSC. (a) Reconstructed electron density of thiol decorated gold nanoparticles obtained by iterative phasing. The atomic model is displayed together with its electron density. (b) FSC as a function of resolution. The horizontal dashed lines show the established threshold for FSC (0.143 and 0.5). The vertical dashed line corresponds to the target resolution of 2.0 Å.

### 3.2.3 Difference Fourier Technique

When a partial model of the protein is initially known, it is often possible to determine the rest of the structure by using the difference Fourier technique [124]. For this technique to work, the partial model that explain the major part of the structure must produce good enough phases that, when combined with the measured amplitude, electron density of the minor part is obtained. The minor part of the structure consists of the thiol molecules which are rigidly attached on the surface of the gold cluster. Using this approach, a fewer number of diffraction patterns would be sufficient to reconstruct the electron density. Figure 3.12 shows the reconstructed electron density using the known phases from the gold cluster and the amplitudes of the entire thiol-gold molecule. The electron density that determines the position of the thiols can be easily identified (as marked with a red arrow in Fig 3.12).



**Figure 3.12:** Reconstructed electron density of a thiol decorated gold nanoparticle using the phases of the gold cluster only. The red arrow shows the position of a few well-resolved thiol molecules.

### 3.3 Experimental Results

#### 3.3.1 Time-Resolved experiment at EuXFEL

A summary of the results from data reduction is given in Table 3.2. The results of the control experiment (X-ray repetition rate of 1.13 MHz and optical laser repetition rate of 375 kHz) are shown in Fig. 3.13. With an optical laser spot size of  $42 \mu\text{m}$  (FWHM) and the sample jet speed of  $30 \text{ m/s}$ , the excited volume of the sample crystals should leave the X-ray interaction region within  $2 \mu\text{s}$ . The sample probed at  $2.67 \mu\text{s}$  should be free of laser influence. But some contaminations were still observed (Fig. 3.13 (b) & (c)). This shows that the excited volume has not left the X-ray interaction region. In addition, the hit rate abruptly decreased from 2% at the first X-ray pulse in the train to 1% (see Fig. 3.14 (a)) in all subsequent pulses. This shows the jet velocities do not reliably replace the sample at the X-ray interaction point. These problems can be solved either by decreasing the laser spot size or by increasing the jet speed. However, these parameters were already at their maximum practical values. Because of this, the X-ray and optical laser repetition rates were reduced to 564 kHz and 141 kHz respectively and an extra X-ray pulse was added between each optical laser pulses. With these parameters, only the first of four pulses contribute to the ultrafast time delays. A time delay of  $10 \text{ ps}$  was set. Consequently, the second, third, and fourth pulses have time delays of  $1.78 \mu\text{s}$ ,  $3.56 \mu\text{s}$ , and  $5.33 \mu\text{s}$  respectively. The difference electron density (DED) maps calculated at  $1.78 \mu\text{s}$  differs completely from the  $10 \text{ ps}$  DED map. A DED map at  $5.33 \mu\text{s}$  is free from contamination (Fig. 3.13 (g)). This means at a shorter time after  $3.56 \mu\text{s}$  the laser-excited jet volume left the X-ray interaction region. Additionally, the hit rate across the entire pulse train remains essentially constant about 2% (Fig. 3.14 (b)). This shows at 564 kHz the sample is sufficiently refreshed before the next X-ray pulse

arrives. Finally, two more time points, 30 *ps*, and 80 *ps* were collected. DED maps of the chromophore binding pocket for time delays 10 *ps*, 30 *ps*, and 80 *ps* are shown in Fig. 3.15. All time delays resulted in excellent DED maps that contain chemically meaningful positive and negative DED features (red and blue features in Fig. 3.15). From the side view of the pCA chromophore in Fig. 3.15 (d-f), we see the chromophore is in a twisted *cis* configuration throughout. These results cover the region previously unknown in the photocycle. This also demonstrates the way how the TR-SFX datasets can be collected at the EuXFEL.

**Table 3.2:** Statistics of the control data collected with 1.13 MHz X-ray pulses. Parentheses: highest resolution shell (1.71-1.74) Å. The table adapted from [46].

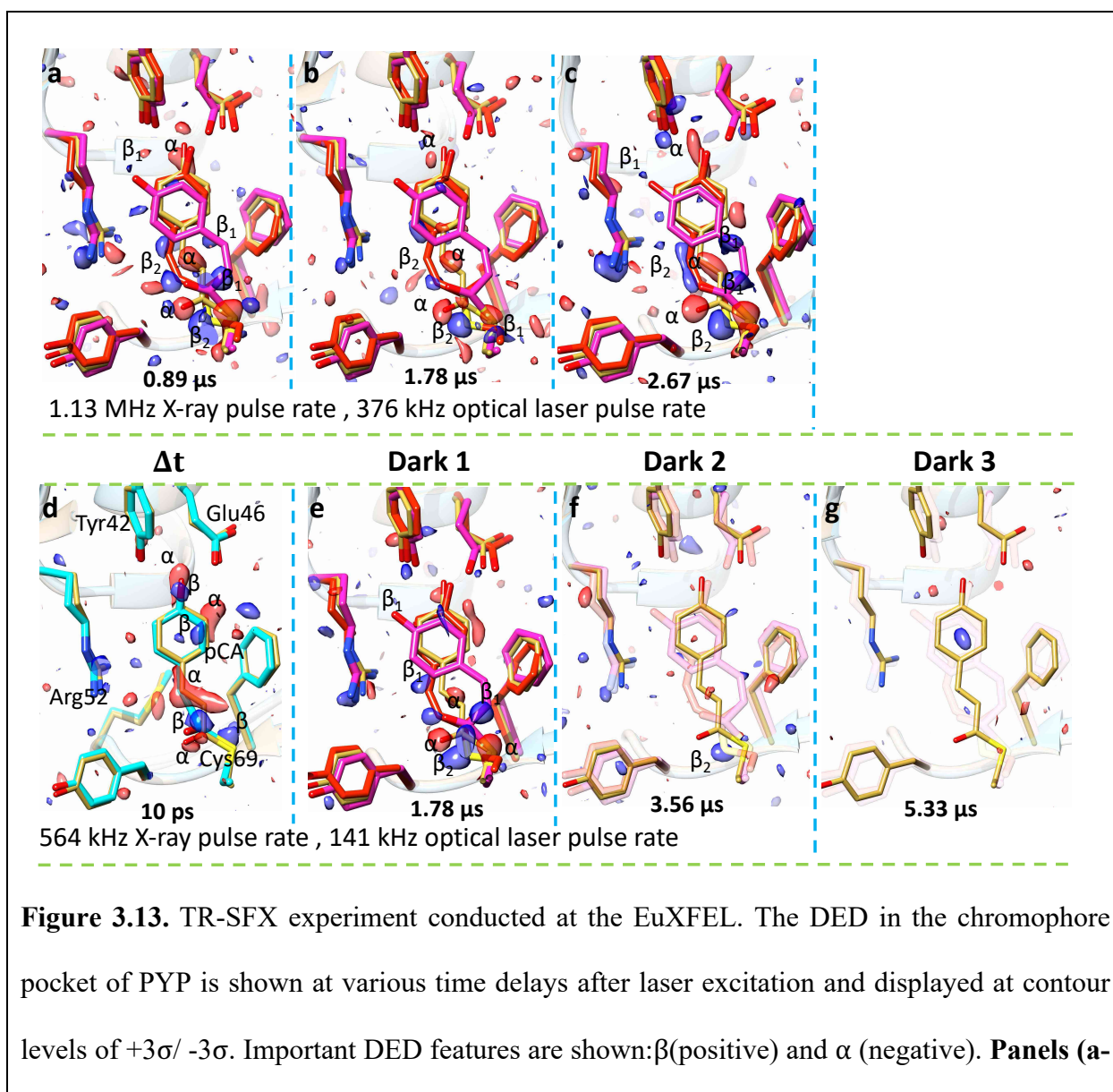
<b>Control experiments</b>				
Repetition rates X-ray/laser	1.13 MHz/376 kHz			
Temperature	285 K			
<b>Data collection</b>				
Space group	P <sub>6</sub> <sub>3</sub> (173)			
Cell dimensions				
<i>a, b, c</i> (Å)	66.9, 66.9, 40.8			
$\alpha, \beta, \gamma$ (°)	90, 90, 120			
Resolution (Å)	1.71 (1.71 - 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	13642	13734	14183	5725
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	11502	11510	11497	11504
$R_{split}$ (%)	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)

**Table 3.3:** Statistics of the data collected with 564 kHz X-ray pulses. Parentheses: highest resolution shell (1.6-1.63) Å. Table adapted from [46].

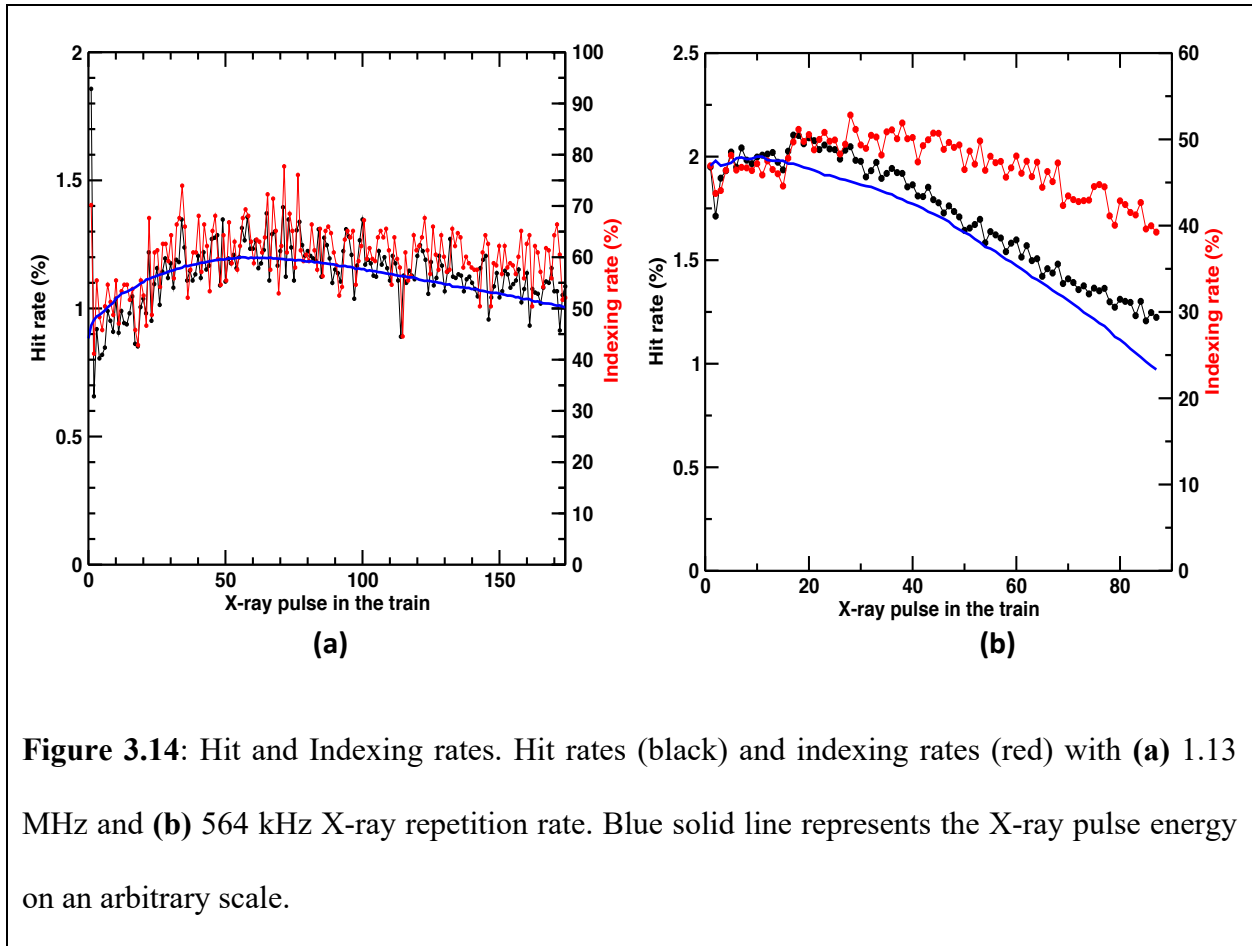
Repetition rates X-ray/Laser	564 kHz/141 kHz			
Resolution	1.6 (1.6 - 1.63) Å			
Temperature	285 K			
Space group	P6 <sub>3</sub> (173)			
Unit cell	a = 66.9 Å b = 66.9 Å c = 40.8 Å α=90° β=90° γ=120°			
time delays	<b>10 ps</b>	<b>dark1 (1.78 μs)</b>	<b>dark2 (3.56 μs)</b>	<b>dark3 (5.33 μs)</b>
Hits	93130	91184	92365	88373
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	14124	14142	14141	14157
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)
time delays	<b>30 ps</b>	<b>dark1 (1.78 μs)</b>	<b>dark2 (3.56 μs)</b>	<b>dark3 (5.33 μs)</b>
Hits	81066	79580	80027	77608
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	13727	13722	13726	13725
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)
time delays	<b>80 ps</b>	<b>dark1 (1.78 μs)</b>	<b>dark2 (3.56 μs)</b>	<b>dark3 (5.33 μs)</b>
Hits	30860	29264	28332	28235
Hit/indexing rate [%]	1.1/60.4	1.0/60.1	1.0/62.2	1.0/60.3

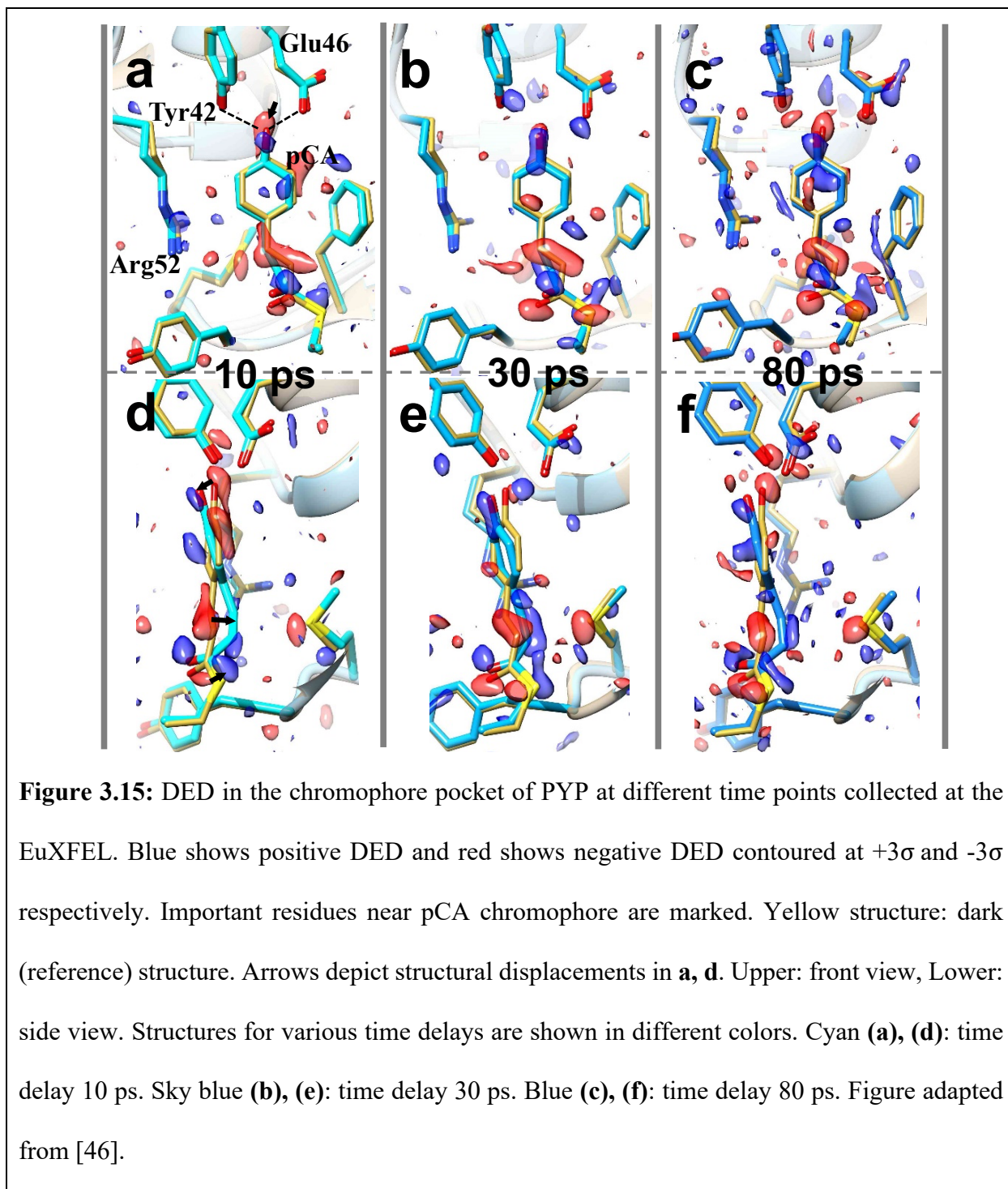


Reflections observed	2,148,177	2,020,504	2,052,291	1,969,452
Unique reflections	13711	13700	13705	13703
*Redundancy	156.7 (15.1)	147.5 (14.1)	149.7 (15.1)	143.7 (14.7)
Completeness (%)	98.0 (97.3)	97.9 (97.1)	98.0 (97.3)	97.9 (97.5)
R-split	8.88 (49.9)	9.58 (49.8)	9.62 (49.5)	9.73 (50.3)
CC1/2	99.1 (68.4)	98.9 (70.8)	98.9 (70.5)	98.8 (67.0)



**c)** Results with 1.13 MHz X-ray repetition rate and 364 kHz optical pulse rate with a pump-probe sequence and two intermittent X-ray pulses without laser activation in between. **(a)** Pump-probe delay of 0.89  $\mu\text{s}$  **(b)** 1.78  $\mu\text{s}$  after the laser pulse **(c)** 2.67  $\mu\text{s}$  after the laser pulse. **Panels (d-g):** Results with 564 kHz X-ray and 141 kHz optical laser pulse rates with a pump-probe sequence and three intermittent X-ray pulses without laser activation in between. **(d)** Time delay of 10 ps. **(e)** 1.78  $\mu\text{s}$  after the laser pulse **(f)** 3.56  $\mu\text{s}$  after the laser pulse **(g)** 5.33  $\mu\text{s}$  after the laser pulse. The DED signal persists until 3.56  $\mu\text{s}$  ( $\beta_2$ , panel f) and completely vanishes at 5.33  $\mu\text{s}$  (panel g). Figure adapted from [46, 125].





## 4. Discussions

### 4.1 Future of SPI

FELs capable of operating at the short wavelength and producing femtosecond ultrabright pulses open the possibilities for the determination of the structures of single biological macromolecules, subcellular organelles at high resolution. Thus far, the resolution obtained from 3D reconstructions of experimental SPI datasets at XFELs [21, 22] has not substantially exceeded that from single-shot imaging [64]. This shows the need to obtain larger datasets to extend the resolution. We have established a sound mathematical formalism to determine the number of snapshots required to answer a specific biological question at any resolution. This formalism, in the future, could be accommodated for a close estimate of the number of patterns for a given resolution incorporating experimental conditions like sample heterogeneity, background, or other nuisances such as X-ray streaks and variations in the detector response [84, 85, 126]. Quantitative results for the number of snapshots are derived as a function of the desired resolution, the molecular size, the expected average number of photons per Shannon pixel, and the SNR in the resolution shell (Eq. 2.26).

There have been significant challenges in collecting data of sufficient quality (low background and high signal) and quantity to enable orientation determination and merging into a 3D diffraction volume for reconstructing electron-density near-atomic resolution. Challenges include a weak signal from the sample compared to the instrumental background, contamination of the aerosolized molecules by nonvolatile contaminants in the solution, heterogeneity of the samples. These challenges result in the collection of a limited number of usable data frames. Other systematics like fluctuations of the beam position and intensity, scattering from the apertures, problems with

sample delivery hinder the collection of diffraction patterns with a reasonable SNR value. The intensity of the XFEL pulses could be increased to boost the SNR value which ultimately increases the resolution. SNR value could also be improved by averaging from many diffraction patterns. To systematically tackle these limitations and push the methodology further an international scientific collaboration is built named the ‘SPI initiative’ [84]. The ‘SPI initiative’ has received several experimental beam times at LCLS. The collected data have been published [64, 127, 128]. These data sets were released to the CXIDB [129] and are available for use to assist in the development of analysis methods. SPI is facing experimental obstacles rather than theoretical boundaries to make the ‘resolution revolution’ possible. But with high repetition-rate, high fluence XFELs, specialized detectors, improvement in the sample delivery technology, and noise-robust data analysis algorithms, the challenges of the SPI technique may be overcome in the near future. It may become possible to collect millions of usable diffraction patterns during an experiment with a high repetition rate of EuXFEL. The experimental datasets of necessary quality and quantity to obtain sub-nanometer resolution (see Table 3.1 for our estimate) from a non-crystalline biological sample can be collected using modern sources. Given improvements in X-ray source and sample delivery, high-resolution structures of a biomolecule at ambient temperature using SPI is on the horizon.

#### **4.2 Reference-enhanced SPI**

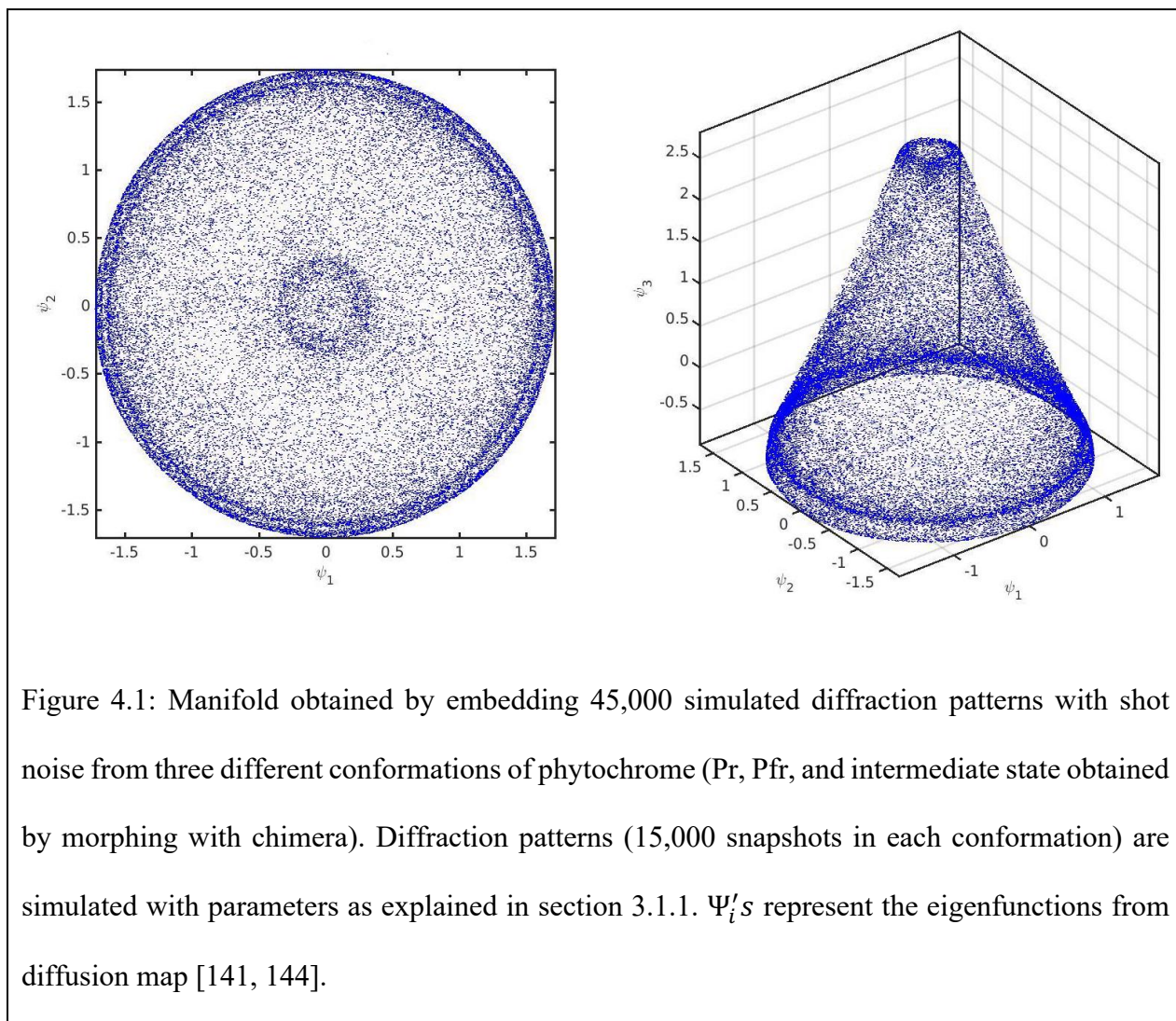
SPI is one of the demanding methods for determining the structure of uncrystallized single biomolecule. However, most of the SPI experiments at XFELs [21, 22, 130] focused on the method development using larger particles. The feasibility of transitioning to smaller particles could be possible with a reference-enhanced diffractive imaging method. This may allow imaging of

weakly scattering objects at XFELs to high resolution [131, 132]. The general principle is to gain SNR value including scattering from a strongly scattering reference. Recently two methodologies, which are slightly modified than the conventional SPI, have been proposed for obtaining diffraction patterns with low background and high signal [133]. This is based on a holographic principle where the strongly scattering holographic references are used in close proximity to the sample of interest. The first proposed method is one where a strong reference scatterer like a gold nanoparticle is chemically attached to the target object. The second system uses a 2D crystal reference in a scanning fixed-target sample geometry. Gold nanoparticles conjugated to biological samples were also studied to improve resolution using cryo-electron microscopy [134–137]. This offers a powerful way to identify and precisely localize specific macromolecular components as gold nanoparticles provide strong contrast in noisy protein images. Here, we showed that it is possible to resolve weakly scattering thiol molecules from the strongly scattering gold-thiol system using conventional SPI technique. Adding highly scattering gold particles hide the signals measured from thiol molecules. However, it appears adding the reference signal is advantageous: when the signal from the object is weaker, or more generally when the signal is less than the threshold required for reliable orientation determination and phase retrieval. Also, the reference could make hit detection easier and improves the hit rate.

### **4.3 Mapping Conformational Spectrum**

Despite the discovery of phytochromes in photosynthetic cyanobacteria in the 1990s [138–140], only one crystal structure of a full-length, intact phytochrome (with diguanylyl cyclase as an enzymatic) is obtained for Pr state, till today. The intact structure with histidine-kinase enzymatic domain for any states is still unknown. The full-length structures of phytochrome, both in their Pr

and Pfr state, as well as the intermediate structures, are necessary for a better understanding of the signaling mechanisms of bacteriophytochromes. As discussed in previous chapters, crystallography provides structural snapshots that are frozen in space and time. Proteins are locked in one conformation due to crystal contacts, which restricts potential structural changes. Our calculations show the structure of an intact phytochrome molecule with SPI is within reach at modern XFELs. The advantage of SPI lies in separating conformational states and thus to image the entire ensemble of structures of a protein. The task of separating different states requires the examination of large datasets, sophisticated programs, and significant computing power. With an increase of computational resources, the algorithm based on diffusion map manifold embedding [141–143] made it possible to study the conformational heterogeneity of the system [21]. Figure 4.1 shows the diffusion map manifolds formed by 45,000 simulated diffraction snapshots of full-length phytochrome molecule in three different conformations (Pr state, proposed Pfr state, and intermediate state obtained by morphing using chimera). In the future, a manifold based analysis could be used for experimental diffraction snapshots of phytochrome which might reveal the conformational landscape of the phytochrome molecule. This would provide insight into the long searched scientific question related to the mechanism of photoactivation. A single-particle approach combined with the ability to separate conformational states promises a very interesting future for structural biology.



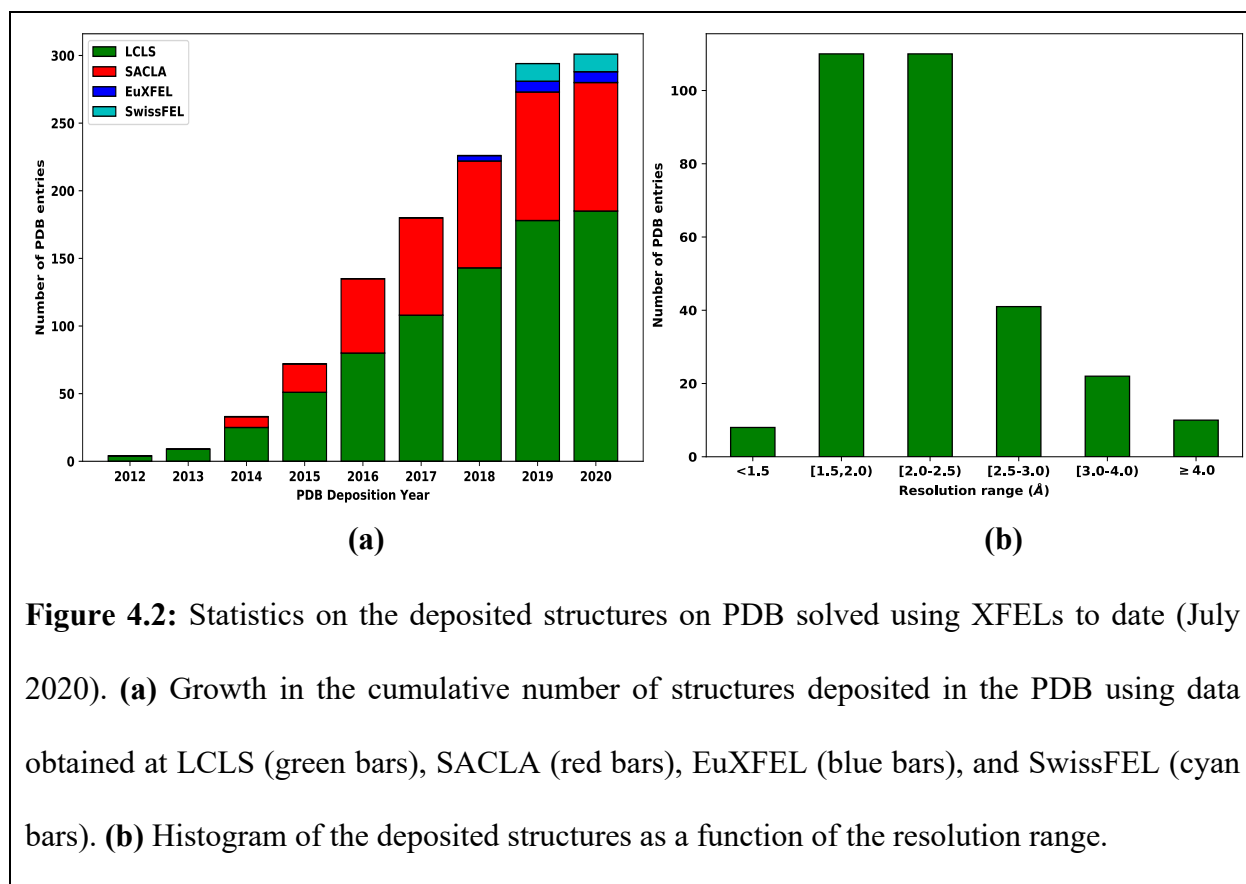
#### 4.4 An Outlook on SFX

SFX field is still relatively young but already led to multiple breakthroughs in structural biology introducing the possibility of overcoming radiation damage and enabling high-resolution, room-temperature structure determination from microcrystals. After the first SFX experiment in 2011 [56], over 300 protein structures have been deposited on PDB (Fig. 4.2 (a)). These also include targets potentially important for pharmacology, which used to be inaccessible for conventional analytical techniques. Most structures (around 80%) are solved at a resolution within



the (1.5-3.0) Å (Fig. 4.2(b)). The highest resolution structure is that of proteinase K [145] using 13 keV photon energy at SACLA, solved to 1.20Å, pushing the atomic resolution limit [146]. One of the most successful applications of SFX has been the structural study of membrane proteins using LCP as the crystallization and the injector system which led to the structure determination of several GPCRs [147–150].

The major application of XFELs in structural biology is to utilize the inherent temporal resolution capacity of XFELs to study reaction intermediates or other structural changes. The foundation of the SFX has also led to the development of a setup for pump-probe TR-SFX [41, 44, 46, 125]. The first successful TR-SFX experiment was conducted on PYP at LCLS [41]. The high-quality difference electron density maps of PYP at 1.6 Å were obtained for time points between 100 fs and 3 ps [41, 44]. This allows the visualization of *cis/trans* isomerization of a chromophore, one of the fastest reactions in biomolecules. The first TR-SFX experiment using MHz X-ray pulses at EuXFEL was also conducted on PYP [46], discussed in this dissertation. The DED maps of excellent quality at 1.6 Å were obtained for picosecond time range. These results connect with the previous results from synchrotrons and XFELs [41, 44, 151]. This pioneering experiment opens the door to a wide range of time-resolved studies at the EuXFEL. Besides these fast light-triggered reactions, slower processes, such as substrate-triggered biological reactions were also studied by mix-and-inject serial crystallography [89, 152–154]. Binding of an antibiotic to  $\beta$ -lactamase from *Mycobacterium tuberculosis* was demonstrated using this approach. This enables the time-resolved study of enzymatic and other ligand driven biological processes.



**Figure 4.2:** Statistics on the deposited structures on PDB solved using XFELs to date (July 2020). **(a)** Growth in the cumulative number of structures deposited in the PDB using data obtained at LCLS (green bars), SACLA (red bars), EuXFEL (blue bars), and SwissFEL (cyan bars). **(b)** Histogram of the deposited structures as a function of the resolution range.

#### 4.5 Photoactive Proteins: *cis-trans* Isomerizations

The ability to convert light into molecular action is vital to many forms of life. Photoactive proteins play a critical role in biological success, for example, the vision in animals, and phototaxis response observed in organisms. Generally, light activation often triggers the geometrical isomerization of a prosthetic chromophore group [151]. When the chromophore is free from host constraints, in the gas and solution phases, the isomerization pathway is thought to proceed via the one-bond-flip mechanism [155]. But when the chromophore is embedded in a cavity within a photoreceptor, the isomerization cannot occur through the standard one-bond-flip mechanism [155, 156]. Understanding the relationship of the chromophore with its local protein environment may provide the ability to inhibit or promote photoisomerization in the biological process. Many

signaling proteins exhibit a *cis*-to-*trans* isomerization. PYP has been used as a model system for investigating structure-function relations in photoactive proteins. Its chromophore, pCA, can be photoexcited by absorbing a photon in the blue region of the spectrum. Upon photon absorption, PYP enters a photocycle involving numerous intermediates (Fig. 1.5 (b)). The hydrogen bond network of the pCA chromophore with Tyr42, Cys69, and Glu46 residues (Fig. 1.5 (a)) play an integral role in establishing the barriers of activation and influencing both the ground and excited-state structures. The fastest time scale achievable at synchrotrons is  $\sim 100$  ps, limited by the pulse width of X-rays. Synchrotron studies on PYP [42, 43] captured all reaction intermediates except for those of the *trans*-*cis* isomerization, which is the fastest bimolecular reactions and needs a shorter time-scale. PYP has also been studied on the fs time scale with time-resolved spectroscopic techniques such as transient absorption spectroscopy [157], fluorescence spectroscopy [158], and Raman spectroscopy [159]. Although these methods reveal the limited structural information [160], they provide insights on protein dynamics that can guide and assist fs structural studies at XFELs. The structural changes of the PYP during *trans*-*cis* isomerization were first visualized using XFEL source [44]. This is marked as one of the important time-resolved experiment at XFELs. Since then, there have been several studies on various proteins such as myoglobin [161], bacteriorhodopsin [162, 163], photosystem II [164–167], phytochrome [168], which show that it is possible to follow cyclic and non-cyclic reactions at the XFEL. The photocycle of PYP has been extensively investigated from femtoseconds to seconds. The time range between 1 ps -100 ps has not been investigated in detail so far with at least one more process observed by spectroscopy [159]. Our pioneering TR-SFX study at EuXFEL has revealed the structures at 10 ps, 30 ps, and 80 ps after the laser excitation.

#### 4.6 Optimum Power for Pump-Probe TR-SFX

Successful pump-probe TR-SFX experiment requires photoactive microcrystals and a short-pulsed laser tuned to the most suitable wavelength which must be synchronized to the X-ray pulse to achieve various time delays. A reaction must be initiated rapidly, uniformly, and nondestructively in the crystal [87]. The large fraction of the molecules needs to be activated without damaging the crystal. The laser power at the sample position must be carefully assessed. Lower power laser pulse will prevent sufficient photo-initiation, whereas high power pulses deposit energy and may damage the crystal. Recently, a pump-power titration was performed with phytochrome crystals which shows that the higher laser fluences do not alter the result of the experiment and rather, positively contribute to stronger DED maps that can be interpreted easily [168]. The advantages of TR-SFX lie in the use of microcrystals. Microcrystals being smaller have lower optical density, which allows a more homogeneous excitation of molecules within microcrystals. Experimental data collected with a serial approach are less sensitive to the accumulated X-ray and pump laser-induced damaged. Laser pulse energy densities that have been used for ultrafast pump-probe experiments are on the order of 0.5-5  $mJ/mm^2$  [41, 44, 161–163]. The first time-resolved experiment with PYP was initiated with *ns* laser pulses (450 *nm*, 0.8  $mJ/mm^2$ ) [41]. The extent of photoactivation is around 40%. With *ns* pulses, those molecules which initially and very rapidly revert to the dark state can be excited anew, multiple times, which boosts the apparent photoactivation yield. At synchrotrons, the typical reaction initiations are 10-15% [160]. The larger levels of photoinitiation lead to stronger difference signals which ultimately produce more accurate structure determination. The second experiment with PYP at LCLS uses *fs* laser pulses (140 *fs*) [44]. With *fs* pulses, around 20% of the molecules populate the photocycle. This agrees with the values found spectroscopically [169, 170]. These experiments show that *fs*

laser excitation is feasible with microcrystals and open the door for *fs* time-resolved experiments. TR-SFX experiment at EuXFEL with PYP was initiated using laser pulses of 240 *fs* at a wavelength of 420 *nm* with a flux density of 1.6 *mJ/mm<sup>2</sup>*. Population transfer for each time point in this experiment is approximately 7%. This value is lower compared to similar excitation schemes at other XFELs. This is due to illuminating into the flanks of the absorption spectrum. The excitation was achieved at 420 *nm* rather than into the central absorption peak at 450 *nm*. Still, excellent data are collected because the laser penetration depth matches the micrometer crystal size [41], leading to the uniform sample excitation. The most appropriate optical conditions for a TR-SFX experiment is a trade-off between power density and peak power. The goal is to exploit the experimentally available time resolution and maximize the levels of detectable populations.

#### 4.7 SPI and Crystallographic Data Analysis: A Comparison

Both SPI and crystallographic data analysis follow a similar broad pipeline. Initially, the hits (when X-ray interacts with the sample) are identified from the total collected frames during an experiment. Diffraction patterns are then preprocessed which includes correction of detector artifacts, estimate and subtract photon background, etc. The orientation of each diffraction pattern must be determined relative to the direction of the X-ray beam. For crystallographic data, the known orientation is then used for assigning Miller indices of each Bragg spot (partial). Intensities from each diffraction pattern are scaled and merged. This produces a file that covers reflection intensities in 3D reciprocal space. Diffraction patterns of a single molecule with known orientations can directly be merged into a 3D diffraction volume. The structure factor amplitudes are generated from the intensities. Using appropriate phasing methods, the phases of the structure factor are retrieved. This results in the electron density map which shows the distribution of

electrons at each point which may then be interpreted to find coordinates for each atom in the molecule. The reconstructed electron densities are often judged by a single factor: resolution, the level of detail a map shows. Resolution is straightforward to ascertain in crystallography, but not in SPI.

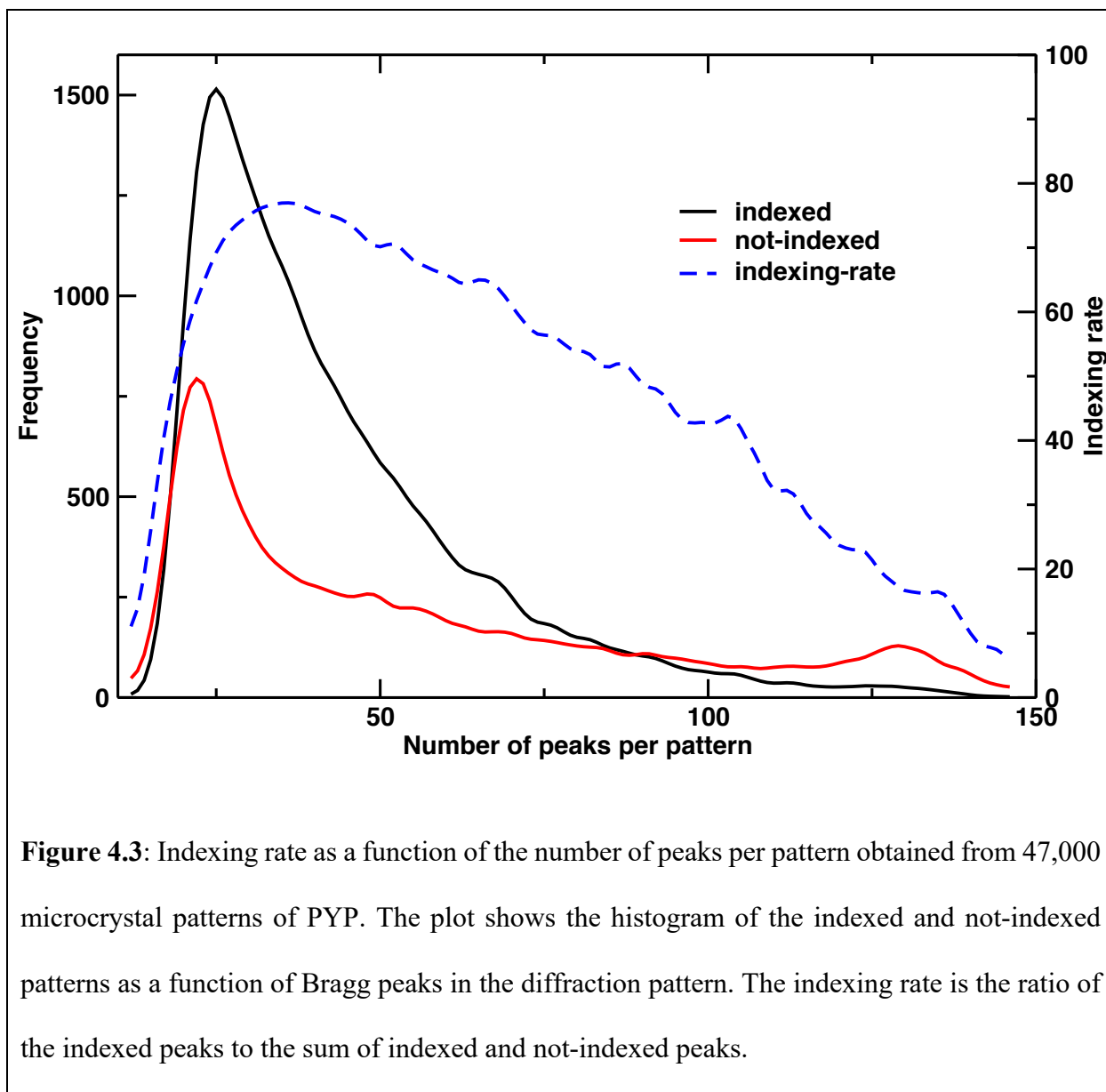
Crystallographic data can be preprocessed using one of the available programs specifically designed or adapted for SFX, such as Cheetah [103], CASS [171], psocake [172, 173]. Near real-time feedback is critical for minimizing protein consumption and minimizing the number of time-delays recorded in a time-series during a TR-SFX experiment. Experimental progress could be monitored online using OnDA [174] for serial crystallography and also to get real-time feedback. The autoindexing program like MOSFLM [107], DirAx [108], XDS [109], asdf [113], and XGandalf [110] are used for determining the orientation and indexing the diffraction pattern. For SFX data these autoindexing programs are included in CrystFEL [106] which is also used for merging the intensities. After a complete intensity dataset is obtained, further calculations can be performed by using CCP4 programs [114]. Most structures obtained from XFEL to date have been solved using the molecular replacement (MR) method [175]. In MR, a structural model of a homologous protein is used to solve the phase problem and obtain the initial phase estimates. When the templates of sufficient quality for MR are not available *de novo* phasing is required [176–178].

Crystallographic data-processing software are robust, and their development have reduced the required amount of data and improved data quality. However, sometimes the highly automated nature of the indexing is less successful than would be expected in a conventional situation where human intervention would be applied in difficult cases. The indexing rate is typically lower when the unit-cell parameters are very large. This gives the spots with very small separations on the

detector. Also, the indexing rate is lower when the detector distance is too short, resulting spots are confined to the middle of the detector instead of spreading over its entire surface. The best indexing rate is found for patterns containing a few tens of Bragg peaks (Fig. 4.3). The development of the auto-indexing algorithm for sparse patterns would greatly increase SFX data utilization. An algorithm like SPIND (sparse-pattern indexing) [179] has been tested for SFX data with very few Bragg reflections.

Data from single-particle diffraction can be preprocessed using the Hummingbird [180], Cheetah [103], Psocake [172] software. The intensity information of each diffraction pattern is converted to the number of photon counts. The important step after this is the identification of single-particle snapshots. For this, there are no available robust programs until now. The development of robust algorithms for determining single hits is critical to the success of SPI. The commonly used method for identifying single particle hits includes principal component analysis with spectral clustering [181], and manifold based approach [21]. Manifold based machine learning algorithms have also been applied to reveal the concerted structural changes exercised by the molecule [21]. Preprocessed single-particle snapshots are then used for orientation determination. Determination of orientations for single-particle snapshots is a key challenge because the signal-to-noise ratio in each snapshot is poor. A recent study of SPI reported successful reconstruction of electron density from diffraction patterns at a signal level of fewer than 100 photons on the average pattern [182]. Multiple algorithmic methods of orientation recovery have been developed to assign orientations to single-particle X-ray diffraction patterns like the common-line method [68], Generative Topographic Mapping (GTM) [69], EMC [70], and correlation maximization [71]. The EMC algorithm has been implemented in Dragonfly software [72]. The source code for this software package can be downloaded online. As non-crystalline objects produce continuous diffraction

patterns, the phases can be directly recovered in an iterative process by sequentially enforcing constraints in reciprocal and real space as discussed in Chapter 3. Electron density maps can be validated using the Fourier Shell Correlation, a method widely used by the cryo-EM community [183], and phase retrieval transfer function [184]. But researchers are still calling for better methods for validating electron density [185].





## 4.8 Future of Time-Resolved Experiments

New XFELs with superconducting accelerators produce a faster pulse X-ray repetition rate, which significantly shortens the data collection time [46, 186]. This allows for multiple sets of data collection within a single beamtime shift. Time-resolved experimental datasets represent the mixture of different short-lived intermediates. With the help of methods like the SVD [160, 187–189] the structures of intermediates can be determined from the time-dependent electron density maps. Reaction pathways and rate coefficients connecting the reaction intermediates can be determined. With the continuous development of sample delivery methods, an increase in beam intensity, instrumentation, and data processing over the past few years, in the future, remote and automated experiments will be of a common practice.

The structural biology community has begun to adopt the serial crystallographic approach at synchrotron radiation (SR) facilities. The advancement in beamline optics, detectors, and high throughput sample delivery methods have made room-temperature serial crystallography possible also at SR sources. Over the last few years, several serial millisecond crystallography (SMX) experiments have been conducted [190–194] at synchrotron beamlines equipped with high-viscosity (toothpaste) injectors [148]. More recently, a polychromatic ('pink') beam has also been successfully used with microcrystals [195, 196]. The abundance of synchrotron beamtime will facilitate time-resolved investigations on a large number of biological macromolecules in the near future [197, 198]. The time-resolved could be the mainstream structural technique for understanding biological functions [125].

Time-resolved cryo-EM has also been used for imaging of short-lived states (10-1000 *ms*) of biological molecules, particularly using the mixing-spraying method [199, 200]. In this method, mixing and reacting are achieved in a monolithic silicon chip, a microfluidic device with two solution inlets, and one spray outlet. For time-resolved cryo-EM studies, a biological molecule is stopped at multiple time points by fast-freezing or chemical fixation. The trapped biological complexes are then visualized by conventional transmission electron microscopy followed by 3D reconstructions. The time-resolved experiment without freezing or staining of biological molecules at ambient temperature is possible with the SPI method. With the continuous effort of the SPI initiative, the 'holy grail' of XFEL technology, the dream to solve atomic resolution structures without the need for crystals, will be achieved in the near future. The scientific community will then explore the possibilities with time-resolved SPI. This could help in understanding biological functions previously difficult with other techniques. For instance, the full-length bacterial phytochrome with histidine kinase could be mixed with ATP to study the structural dynamics of the histidine kinase autophosphorylation complex. Difference electron density maps between the Pr and Pfr states (Fig. 3.8 (b)) could help in understanding signaling mechanisms of the bacteriophytochromes. The ultrashort pulse length of the EuXFEL gives the unique possibility to study the dynamics of a very short-lived complexes and transient states of dynamic molecules. The EuXFEL can probe dynamics many orders of magnitude faster than what is possible with cryo-EM, which is limited to *ms* timescale. With time-resolved SPI, different well-populated reaction intermediates could be identified with the help of classification programs like manifold embedding. It can also show how the molecule population redistributes into different states as time progresses. It can be argued that SPI can capture functionally meaningful states as

the reconstructions obtained depict ensemble averages of free-standing molecules, unhindered by intermolecular contacts.

## 5. References

1. Nobel Prize in Chemistry. <https://www.nobelprize.org/uploads/2018/06/popular-chemistryprize2014.pdf> [accessed March 20, 2020].
2. Hell SW. Far-field optical nanoscopy. *Science* 2007. DOI: 10.1126/science.1137395.
3. Als-Nielsen J, McMorrow D. *Elements of Modern X-ray Physics: Second Edition*. 2011. DOI: 10.1002/9781119998365.
4. Bernal JD, Crowfoot D. X-ray photographs of crystalline pepsin [3]. *Nature* 1934. DOI: 10.1038/133794b0.
5. Strandberg, Dickerson, Rossmann. 50 Years of Protein Structure Analysis. *Journal Of Molecular Biology* 2009. DOI: 10.1016/j.jmb.2009.05.024.
6. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of Hæmoglobin: A three-dimensional fourier synthesis at 5.5- resolution, obtained by X-ray analysis. *Nature* 1960. DOI: 10.1038/185416a0.
7. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 1958. DOI: 10.1038/181662a0.
8. Rossmann MG. John C. Kendrew and His Times. *Journal of Molecular Biology* 2017. DOI: 10.1016/j.jmb.2017.03.012.
9. Garman EF. Developments in X-ray crystallographic structure determination of biological macromolecules. *Science* 2014; 343(6175): 1102–1108. DOI: 10.1126/science.1247829.
10. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallographica Section D: Biological Crystallography* 2002. DOI: 10.1107/S0907444902003451.
11. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! *BioRxiv* 2020. DOI: 10.1101/2020.05.21.106740.
12. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, et al. Single-particle cryo-EM at atomic resolution. *BioRxiv* 2020. DOI: 10.1101/2020.05.22.110189.
13. Rupp B. *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science; 2009.
14. Sayre D. Some implications of a theorem due to Shannon. *Acta Crystallographica* 1952. DOI: 10.1107/s0365110x52002276.
15. Shannon CE. Communication in the Presence of Noise. *Proceedings of the IRE* 1949. DOI: 10.1109/JRPROC.1949.232969.
16. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC. Dynamics of Ligand Binding to Myoglobin. *Biochemistry* 1975. DOI: 10.1021/bi00695a021.
17. Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben IE, Sauke TB, et al. Protein states and proteinquakes. *Proceedings of the National Academy of Sciences of the United States of America* 1985. DOI: 10.1073/pnas.82.15.5000.
18. Frauenfelder H, Parak F, Young RD. Conformational substates in proteins. *Annual Review of Biophysics and Biophysical Chemistry* 1988. DOI: 10.1146/annurev.bb.17.060188.002315.
19. Kruschel D, Zagrovic B. Conformational averaging in structural biology: Issues, challenges and computational solutions. *Molecular BioSystems* 2009. DOI: 10.1039/b917186j.
20. Tinoco I, Gonzalez RL. Biological mechanisms, one molecule at a time. *Genes and*

- Development 2011. DOI: 10.1101/gad.2050011.
21. Hosseinizadeh A, Mashayekhi G, Copperman J, Schwander P, Dashti A, Sepehr R, et al. Conformational landscape of a virus by single-particle X-ray scattering. *Nature Methods* 2017; 14(9): 877–881. DOI: 10.1038/nmeth.4395.
  22. Ekeberg T, Svenda M, Abergel C, Maia FRNC, Seltzer V, Claverie JM, et al. Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Physical Review Letters* 2015; 114(9): 1–6. DOI: 10.1103/PhysRevLett.114.098102.
  23. Seibert MM, Ekeberg T, Maia FRNC, Svenda M, Andreasson J, Jönsson O, et al. Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* 2011. DOI: 10.1038/nature09748.
  24. Brandt S, Von Stetten D, Günther M, Hildebrandt P, Frankenberg-Dinkel N. The fungal phytochrome FphA from *Aspergillus nidulans*. *Journal of Biological Chemistry* 2008. DOI: 10.1074/jbc.M805506200.
  25. Auldridge ME, Forest KT. Bacterial phytochromes: More than meets the light. *Critical Reviews in Biochemistry and Molecular Biology* 2011. DOI: 10.3109/10409238.2010.546389.
  26. Jiang ZY, Swem LR, Rushing BC, Devanathan S, Tollin G, Bauer CE. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* 1999. DOI: 10.1126/science.285.5426.406.
  27. Wagner JR, Brunzelle JS, Forest KT, Vierstra RD. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* 2005. DOI: 10.1038/nature04118.
  28. Burgie ES, Bussell AN, Walker JM, Dubiel K, Vierstra RD. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proceedings of the National Academy of Sciences* 2014; 111(28): 10179–10184. DOI: 10.1073/pnas.1403096111.
  29. Woitowich NC, Halavaty AS, Waltz P, Kupitz C, Valera J, Tracy G, et al. Structural basis for light control of cell development revealed by crystal structures of a myxobacterial phytochrome. *IUCrJ* 2018; 5. DOI: 10.1107/S2052252518010631.
  30. Sanchez JC, Carrillo M, Pandey S, Noda M, Aldama L, Feliz D, et al. High-resolution crystal structures of a myxobacterial phytochrome at cryo and room temperatures. *Structural Dynamics* 2019. DOI: 10.1063/1.5120527.
  31. Essen LO, Mailliet J, Hughes J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proceedings of the National Academy of Sciences of the United States of America* 2008. DOI: 10.1073/pnas.0806477105.
  32. Takala H, Björling A, Berntsson O, Lehtivuori H, Niebling S, Hoernke M, et al. Signal amplification and transduction in phytochrome photosensors. *Nature* 2014. DOI: 10.1038/nature13310.
  33. Burgie ES, Zhang J, Vierstra RD. Crystal Structure of *Deinococcus* Phytochrome in the Photoactivated State Reveals a Cascade of Structural Rearrangements during Photoconversion. *Structure* 2016; 24(3): 448–457. DOI: 10.1016/j.str.2016.01.001.
  34. Gourinchas G, Ettl S, Göbl C, Vide U, Madl T, Winkler A. Long-range allosteric signaling in red light-regulated diguanylyl cyclases. *Science Advances* 2017; 3(3): e1602498. DOI: 10.1126/sciadv.1602498.
  35. Björling A, Berntsson O, Lehtivuori H, Takala H, Hughes AJ, Panman M, et al. Structural photoactivation of a full-length bacterial phytochrome. *Science Advances* 2016. DOI:

- 10.1126/sciadv.1600920.
36. Jadzinsky PD, Calero G, Ackerson CJ, Bushnell DA, Kornberg RD. Structure of a thiol monolayer-protected gold nanoparticle at 1.1 Å resolution. *Science* 2007. DOI: 10.1126/science.1148624.
  37. Meyer TE. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *BBA - Bioenergetics* 1985. DOI: 10.1016/0005-2728(85)90094-5.
  38. Hellingwerf KJ, Hendriks J, Gensch TH. On the configurational and conformational changes in photoactive yellow protein that leads to signal generation in *Ectothiorhodospira halophila*. *Journal of Biological Physics* 2002. DOI: 10.1023/A:1020360505111.
  39. Purwar N, Tenboer J, Tripathi S, Schmidt M. Spectroscopic studies of model photo-receptors: Validation of a nanosecond time-resolved micro-spectrophotometer design using photoactive yellow protein and  $\alpha$ -phycoerythrocyanin. *International Journal of Molecular Sciences* 2013. DOI: 10.3390/ijms140918881.
  40. Schmidt M. A short history of structure based research on the photocycle of photoactive yellow protein. *Structural Dynamics* 2017. DOI: 10.1063/1.4974172.
  41. Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, Frank M, et al. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* 2014. DOI: 10.1126/science.1259357.
  42. Ihee H, Rajagopal S, Srajer V, Pahl R, Anderson S, Schmidt M, et al. Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proceedings of the National Academy of Sciences of the United States of America* 2005. DOI: 10.1073/pnas.0409035102.
  43. Schotte F, Cho HS, Kaila VRI, Kamikubo H, Dashdorj N, Henry ER, et al. Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proceedings of the National Academy of Sciences of the United States of America* 2012. DOI: 10.1073/pnas.1210938109.
  44. Pande K, Hutchison CDM, Groenhof G, Aquila A, Robinson JS, Tenboer J, et al. Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* 2016. DOI: 10.1126/science.aad5081.
  45. Kort R, Vonk H, Xu X, Hoff WD, Crielaard W, Hellingwerf KJ. Evidence for trans-cis isomerization of the p-coumaric acid chromophore as the photochemical basis of the photocycle of photoactive yellow protein. *FEBS Letters* 1996. DOI: 10.1016/0014-5793(96)00149-4.
  46. Pandey S, Bean R, Sato T, Poudyal I, Bielecki J, Cruz Villarreal J, et al. Time-resolved serial femtosecond crystallography at the European XFEL. *Nature Methods* 2019: 1–33. DOI: 10.1038/s41592-019-0628-z.
  47. Madey JMJ. Stimulated emission of bremsstrahlung in a periodic magnetic field. *Journal of Applied Physics* 1971. DOI: 10.1063/1.1660466.
  48. Pellegrini C, Marinelli A, Reiche S. The physics of x-ray free-electron lasers. *Reviews of Modern Physics* 2016. DOI: 10.1103/RevModPhys.88.015006.
  49. Schmöser P (Peter), Dohlus M, Rossbach J, Behrens C, Revision of Schmöser P (Peter). Free-electron lasers in the ultraviolet and x-ray regime : physical principles, experimental results, technical realization.
  50. Robinson I, Gruebel G, Mochrie S. X-ray beams with high coherence. *New Journal of Physics* 2010. DOI: 10.1088/1367-2630/12/3/035002.

51. Feldhaus J, Saldin EL, Schneider JR, Schneidmiller EA, Yurkov M V. Possible application of X-ray optical elements for reducing the spectral bandwidth of an X-ray SASE FEL. *Optics Communications* 1997. DOI: 10.1016/S0030-4018(97)00163-6.
52. Henderson R. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Quarterly Reviews of Biophysics* 1995; 28(2): 171–193. DOI: 10.1017/S003358350000305X.
53. Garman EF, Schneider TR. Macromolecular Cryocrystallography. *Journal of Applied Crystallography* 1997. DOI: 10.1107/S0021889897002677.
54. Solem JC, Baldwin GC. Microholography of living organisms. *Science* 1982. DOI: 10.1126/science.218.4569.229.
55. Neutze R, Wouts R, van der Spoel D, Weckert E, Haidu J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* 2000; 406(August): 752–757.
56. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, et al. Femtosecond X-ray protein nanocrystallography. *Nature* 2011. DOI: 10.1038/nature09750.
57. Chapman HN, Barty A, Bogan MJ, Boutet S, Frank M, Hau-Riege SP, et al. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics* 2006. DOI: 10.1038/nphys461.
58. Cowley JM. *Diffraction Physics*. Third Revi. North Holland: Elsevier; 1995.
59. Wilson AJC, Prince E. *International Tables for Crystallography Vol. C, Mathematical, Physical and Chemical Tables*. Dordrecht, Holland 1999; 1.
60. Kirian RA. Structure determination through correlated fluctuations in x-ray scattering. *Journal of Physics B: Atomic, Molecular and Optical Physics* 2012. DOI: 10.1088/0953-4075/45/22/223001.
61. Drenth J, Mesters J. *Principles of protein X-ray crystallography: Third edition*. 2007. DOI: 10.1007/0-387-33746-6.
62. Shoemake K. *Uniform random rotations*. *Graphics Gems III (IBM Version)*, Elsevier; 1992.
63. Kuipers JB, others. *Quaternions and rotation sequences*. vol. 66. Princeton university press Princeton; 1999.
64. Munke A, Andreasson J, Aquila A, Awel S, Ayyer K, Barty A, et al. Coherent diffraction of single Rice Dwarf virus particles using hard X-rays at the Linac Coherent Light Source. *Scientific Data* 2016. DOI: 10.1038/sdata.2016.64.
65. Pedersen JS. Analysis of small-angle scattering data from colloids and polymer solutions: Modeling and least-squares fitting. *Advances in Colloid and Interface Science* 1997. DOI: 10.1016/S0001-8686(97)00312-6.
66. Joseph W. Goodman. *Statistical Optics*. John Wiley & Sons.; 2000.
67. Hantke MF, Hasse D, Maia FRNC, Ekeberg T, John K, Svenda M, et al. High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nature Photonics* 2014. DOI: 10.1038/nphoton.2014.270.
68. Shneerson VL, Ourmazd A, Saldin DK. Crystallography without crystals. I. The common-line method for assembling a three-dimensional diffraction volume from single-particle scattering. *Acta Crystallographica Section A* 2008; 64(2): 303–315.
69. Fung R, Shneerson V, Saldin DK, Ourmazd A. Structure from fleeting illumination of faint spinning objects in flight. *Nature Physics* 2009. DOI: 10.1038/nphys1129.
70. Loh NTD, Elser V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2009. DOI: 10.1103/PhysRevE.80.026705.

71. Tegze M, Bortel G. Atomic structure of a single large biomolecule from diffraction patterns of random orientations. *Journal of Structural Biology* 2012. DOI: 10.1016/j.jsb.2012.04.014.
72. Ayyer K, Lan TY, Elser V, Loh ND. Dragonfly: An implementation of the expand-maximize-compress algorithm for single-particle imaging. *Journal of Applied Crystallography* 2016. DOI: 10.1107/S1600576716008165.
73. Hosseinizadeh A, Schwander P, Dashti A, Fung R, D'Souza RM, Ourmazd A. High-resolution structure of viruses from random diffraction snapshots. *Phil Trans R Soc B* 2014; 369(1647): 20130326.
74. Miao J, Sayre D. On possible extensions of X-ray crystallography through diffraction-pattern oversampling. *Acta Crystallographica Section A: Foundations of Crystallography* 2000. DOI: 10.1107/S010876730001031X.
75. Miao J, Hodgson KO, Sayre D. An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. *Proceedings of the National Academy of Sciences of the United States of America* 2001. DOI: 10.1073/pnas.111083998.
76. Gerchberg RW, Saxton WO. PRACTICAL ALGORITHM FOR THE DETERMINATION OF PHASE FROM IMAGE AND DIFFRACTION PLANE PICTURES. *Optik (Stuttgart)* 1972.
77. Fienup JR. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters* 1978; 3(1): 27–29.
78. Fienup JR, Crimmins TR, Holsztynski W. RECONSTRUCTION OF THE SUPPORT OF AN OBJECT FROM THE SUPPORT OF ITS AUTOCORRELATION. *Journal of the Optical Society of America* 1982. DOI: 10.1364/JOSA.72.000610.
79. Fienup JR, Wackerman CC. Phase-retrieval stagnation problems and solutions. *Journal of the Optical Society of America A* 1986. DOI: 10.1364/josaa.3.001897.
80. Fienup JR. Phase retrieval algorithms: a comparison. *Applied Optics* 1982. DOI: 10.1364/ao.21.002758.
81. Marchesini S. A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments* 2007. DOI: 10.1063/1.2403783.
82. Marchesini S, He H, Chapman HN, Hau-Riege SP, Noy A, Howells MR, et al. X-ray image reconstruction from a diffraction pattern alone. *Physical Review B* 2003; 68(14): 140101.
83. Oszlányi G, Süto A. Ab initio structure solution by charge flipping. *Acta Crystallographica Section A: Foundations of Crystallography* 2004. DOI: 10.1107/S0108767303027569.
84. Aquila A, Barty A, Bostedt C, Boutet S, Carini G, Deponte D, et al. The linac coherent light source single particle imaging road map. *Structural Dynamics* 2015; 2(4): 41701. DOI: 10.1063/1.4918726.
85. Hosseinizadeh A, Dashti A, Schwander P, Fung R, Ourmazd A. Single-particle structure determination by X-ray free-electron lasers: Possibilities and challenges. *Structural Dynamics* 2015; 2(4): 41601. DOI: 10.1063/1.4919740.
86. Moffat K. Time-Resolved Crystallography. *Acta Crystallographica Section A: Foundations of Crystallography* 1998. DOI: 10.1107/S0108767398010605.
87. Moffat K. Time-resolved biochemical crystallography: A mechanistic perspective. *Chemical Reviews* 2001. DOI: 10.1021/cr990039q.
88. Schmidt M. Structure Based Kinetics by Time-Resolved X-ray Crystallography. In: Braun M., Gilch P. ZW, editor. *Ultrashort Laser Pulses in Biology and Medicine. Biological and Medical Physics, Biomedical Engineering*, Springer, Berlin, Heidelberg; 2008. DOI:



- [https://doi.org/10.1007/978-3-540-73566-3\\_9](https://doi.org/10.1007/978-3-540-73566-3_9).
89. Olmos JL, Pandey S, Martin-Garcia JM, Calvey G, Katz A, Knoska J, et al. Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC Biology* 2018; 16(1). DOI: 10.1186/s12915-018-0524-5.
  90. Bourgeois D, Weik M. Kinetic protein crystallography: A tool to watch proteins in action. *Crystallography Reviews* 2009. DOI: 10.1080/08893110802604868.
  91. DePonte DP, Weierstall U, Schmidt K, Warner J, Starodub D, Spence JCH, et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *Journal of Physics D: Applied Physics* 2008. DOI: 10.1088/0022-3727/41/19/195505.
  92. Nogly P, Panneels V, Nelson G, Gati C, Kimura T, Milne C, et al. Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography. *Acta Crystallographica Section A Foundations and Advances* 2016. DOI: 10.1107/s2053273316099368.
  93. Kupitz C, Olmos JL, Holl M, Tremblay L, Pande K, Pandey S, et al. Structural enzymology using X-ray free electron lasers. *Structural Dynamics* 2017. DOI: 10.1063/1.4972069.
  94. Calvey GD, Katz AM, Schaffer CB, Pollack L. Mixing injector enables time-resolved crystallography with high hit rate at X-ray free electron lasers. *Structural Dynamics* 2016. DOI: 10.1063/1.4961971.
  95. Oberthuer D, Knoška J, Wiedorn MO, Beyerlein KR, Bushnell DA, Kovaleva EG, et al. Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Scientific Reports* 2017. DOI: 10.1038/srep44628.
  96. Sugahara M, Mizohata E, Nango E, Suzuki M, Tanaka T, Masuda T, et al. Grease matrix as a versatile carrier of proteins for serial crystallography. *Nature Methods* 2014. DOI: 10.1038/nmeth.3172.
  97. Sugahara M, Song C, Suzuki M, Masuda T, Inoue S, Nakane T, et al. Oil-free hyaluronic acid matrix for serial femtosecond crystallography. *Scientific Reports* 2016. DOI: 10.1038/srep24484.
  98. Sugahara M, Motomura K, Suzuki M, Masuda T, Joti Y, Numata K, et al. Viscosity-adjustable grease matrices for serial nanocrystallography. *Scientific Reports* 2020. DOI: 10.1038/s41598-020-57675-7.
  99. Hatsui T, Graafsma H. X-ray imaging detectors for synchrotron and XFEL sources. *IUCrJ* 2015. DOI: 10.1107/S205225251500010X.
  100. Blaj G, Caragiulo P, Carini G, Carron S, Dragone A, Freytag D, et al. X-ray detectors at the Linac Coherent Light Source. *Journal of Synchrotron Radiation* 2015. DOI: 10.1107/S1600577515005317.
  101. Allahgholi A, Becker J, Delfs A, Dinapoli R, Goettlicher P, Greiffenberg D, et al. The Adaptive Gain Integrating Pixel Detector at the European XFEL. *Journal of Synchrotron Radiation* 2019. DOI: 10.1107/S1600577518016077.
  102. Boutet S, Foucar L, Barends TRM, Botha S, Doak RB, Koglin JE, et al. Characterization and use of the spent beam for serial operation of LCLS. *Journal of Synchrotron Radiation* 2015. DOI: 10.1107/S1600577515004002.
  103. Barty A, Kirian R a., Maia FRNC, Hantke M, Yoon CH, White T a., et al. Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography* 2014. DOI: 10.1107/S1600576714007626.
  104. Hunter MS, Segelke B, Messerschmidt M, Williams GJ, Zatsepin NA, Barty A, et al. Fixed-target protein serial microcrystallography with an X-ray free electron laser. *Scientific*

- Reports 2014. DOI: 10.1038/srep06026.
105. Martin AV, Morgan AJ, Ekeberg T, Loh ND, Maia FRNC, Wang F, et al. The extraction of single-particle diffraction patterns from a multiple-particle diffraction pattern. *Optics Express* 2013. DOI: 10.1364/oe.21.015102.
  106. White TA, Kirian RA, Martin A V., Aquila A, Nass K, Barty A, et al. CrystFEL: A software suite for snapshot serial crystallography. *Journal of Applied Crystallography* 2012. DOI: 10.1107/S0021889812002312.
  107. Leslie AGW. The integration of macromolecular diffraction data. *Acta Crystallographica Section D: Biological Crystallography*, 2006. DOI: 10.1107/S0907444905039107.
  108. Duisenberg AJM. Indexing in single-crystal diffractometry with an obstinate list of reflections. *Journal of Applied Crystallography* 1992. DOI: 10.1107/S0021889891010634.
  109. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography* 2010. DOI: 10.1107/S0907444909047374.
  110. Gevorkov Y, Yefanov O, Barty A, White TA, Mariani V, Brehm W, et al. XGANDALF - Extended gradient descent algorithm for lattice finding. *Acta Crystallographica Section A: Foundations and Advances* 2019. DOI: 10.1107/S2053273319010593.
  111. Kirian RA, Wang X, Weierstall U, Schmidt KE, Spence JCH, Hunter M, et al. Femtosecond protein nanocrystallography—data analysis methods. *Optics Express* 2010. DOI: 10.1364/oe.18.005713.
  112. Brehm W, Diederichs K. Breaking the indexing ambiguity in serial crystallography. *Acta Crystallographica Section D: Biological Crystallography* 2014. DOI: 10.1107/S1399004713025431.
  113. White TA, Mariani V, Brehm W, Yefanov O, Barty A, Beyerlein KR, et al. Recent developments in CrystFEL. *Journal of Applied Crystallography* 2016. DOI: 10.1107/S1600576716004751.
  114. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 2011. DOI: 10.1107/S0907444910045749.
  115. Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science* 2012. DOI: 10.1126/science.1218231.
  116. Genick UK, Borgstahl GEO, Ng K, Ren Z, Pradervand C, Burke PM, et al. Structure of a protein photocycle intermediate by millisecond time- resolved crystallography. *Science* 1997. DOI: 10.1126/science.275.5305.1471.
  117. Wiedorn MO, Oberthür D, Bean R, Schubert R, Werner N, Abbey B, et al. Megahertz serial crystallography. *Nature Communications* 2018. DOI: 10.1038/s41467-018-06156-7.
  118. Sayre D, Chapman HN, Miao J. On the Extendibility of X-ray Crystallography to Noncrystals. *Acta Crystallographica Section A: Foundations of Crystallography* 1998. DOI: 10.1107/S0108767397015572.
  119. Strüder L, Epp S, Rolles D, Hartmann R, Holl P, Lutz G, et al. Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 2010. DOI: 10.1016/j.nima.2009.12.053.
  120. Van Heel M, Schatz M. Fourier shell correlation threshold criteria. *Journal of Structural Biology* 2005; 151(3): 250–262.

121. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology* 2003; 333(4): 721–745. DOI: 10.1016/j.jmb.2003.07.013.
122. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 2004. DOI: 10.1002/jcc.20084.
123. Zumbahlen H. *Linear Circuit Design Handbook*. 2008. DOI: 10.1016/B978-0-7506-8703-4.X0001-6.
124. Cochran W. Some properties of the (Fo-Fc)-synthesis. *Acta Crystallographica* 1951; 4(5): 408–411.
125. Pandey S, Poudyal I, Malla TN. Pump-Probe Time-Resolved Serial Femtosecond Crystallography at X-Ray Free Electron Lasers. *Crystals* 2020; 10(7): 628. DOI: 10.3390/cryst10070628.
126. Ziaja B, Jurek Z, Medvedev N, Saxena V, Son SKK, Santra R. Towards Realistic Simulations of Macromolecules Irradiated under the Conditions of Coherent Diffraction Imaging with an X-ray Free-Electron Laser. *Photonics* 2015; 2(1): 256–269. DOI: 10.3390/photonics2010256.
127. Reddy HKN, Yoon CH, Aquila A, Awel S, Ayyer K, Barty A, et al. Coherent soft X-ray diffraction imaging of coliphage PR772 at the Linac coherent light source. *Scientific Data* 2017. DOI: 10.1038/sdata.2017.79.
128. Rose M, Bobkov S, Ayyer K, Kurta RP, Dzhigaev D, Kim YY, et al. Single-particle imaging without symmetry constraints at an X-ray free-electron laser. *IUCrJ* 2018. DOI: 10.1107/S205225251801120X.
129. Maia FRNC. The coherent X-ray imaging data bank. *Nature Methods* 2012. DOI: 10.1038/nmeth.2110.
130. Lundholm I V., Sellberg JA, Ekeberg T, Hantke MF, Okamoto K, Van Der Schot G, et al. Considerations for three-dimensional image reconstruction from experimental data in coherent diffractive imaging. *IUCrJ* 2018. DOI: 10.1107/S2052252518010047.
131. Boutet S, Bogan MJ, Barty A, Frank M, Benner WH, Marchesini S, et al. Ultrafast soft X-ray scattering and reference-enhanced diffractive imaging of weakly scattering nanoparticles. *Journal of Electron Spectroscopy and Related Phenomena* 2008. DOI: 10.1016/j.elspec.2008.06.004.
132. Lan TY, Li PN, Lee TK. Method to enhance the resolution of x-ray coherent diffraction imaging for non-crystalline bio-samples. *New Journal of Physics* 2014; 16(3): 033016. DOI: 10.1088/1367-2630/16/3/033016.
133. Ayyer K. Reference-enhanced x-ray single-particle imaging. *Optica* 2020; 7(6): 593. DOI: 10.1364/OPTICA.391373.
134. Safer D, Bolinger L, Leigh JS. Undecagold clusters for site-specific labeling of biological macromolecules: simplified preparation and model applications. *Journal of Inorganic Biochemistry* 1986. DOI: 10.1016/0162-0134(86)80001-0.
135. Hainfeld JF, Furuya FR. A 1.4-nm gold cluster covalently attached to antibodies improves immunolabeling. *Journal of Histochemistry and Cytochemistry* 1992. DOI: 10.1177/40.2.1552162.
136. Ackerson CJ, Jadzinsky PD, Jensen GJ, Kornberg RD. Rigid, specific, and discrete gold nanoparticle/antibody conjugates. *Journal of the American Chemical Society* 2006. DOI: 10.1021/ja0555668.

137. Dahan I, Sorrentino S, Boujemaa-Paterski R, Medalia O. Tiopronin-Protected Gold Nanoparticles as a Potential Marker for Cryo-EM and Tomography. *Structure* 2018. DOI: 10.1016/j.str.2018.06.009.
138. Kehoe DM, Grossman AR. Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* 1996. DOI: 10.1126/science.273.5280.1409.
139. Hughes J, Lamparter T, Mittmann F, Hartmann E, Gartner W, Wilde A, et al. A prokaryotic phytochrome [6]. *Nature* 1997. DOI: 10.1038/386663a0.
140. Yeh KC, Wu SH, Murphy JT, Lagarias JC. A cyanobacterial phytochrome two-component light sensory system. *Science* 1997. DOI: 10.1126/science.277.5331.1505.
141. Giannakis D, Schwander P, Ourmazd A. The symmetries of image formation by scattering I Theoretical framework. *Optics Express* 2012. DOI: 10.1364/oe.20.012799.
142. Schwander P, Giannakis D, Yoon CH, Ourmazd A. The symmetries of image formation by scattering II Applications. *Optics Express* 2012. DOI: 10.1364/oe.20.012827.
143. Schwander P, Fung R, Ourmazd A. Conformations of macromolecules and their complexes from heterogeneous datasets. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2014. DOI: 10.1098/rstb.2013.0567.
144. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the National Academy of Sciences of the United States of America* 2005. DOI: 10.1073/pnas.0500334102.
145. Masuda T, Suzuki M, Inoue S, Song C, Nakane T, Nango E, et al. Atomic resolution structure of serine protease proteinase K at ambient temperature. *Scientific Reports* 2017. DOI: 10.1038/srep45604.
146. Johansson LC, Stauch B, Ishchenko A, Cherezov V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends in Biochemical Sciences* 2017. DOI: 10.1016/j.tibs.2017.06.007.
147. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, et al. Serial femtosecond crystallography of G protein-coupled receptors. *Science* 2013. DOI: 10.1126/science.1244142.
148. Weierstall U, James D, Wang C, White TA, Wang D, Liu W, et al. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications* 2014. DOI: 10.1038/ncomms4309.
149. Fenalti G, Zatsepin NA, Betti C, Giguere P, Han GW, Ishchenko A, et al. Structural basis for bifunctional peptide recognition at human  $\delta$ -opioid receptor. *Nature Structural and Molecular Biology* 2015. DOI: 10.1038/nsmb.2965.
150. Zhang H, Qiao A, Yang D, Yang L, Dai A, De Graaf C, et al. Structure of the full-length glucagon class B G-protein-coupled receptor. *Nature* 2017. DOI: 10.1038/nature22363.
151. Jung YO, Lee JH, Kim J, Schmidt M, Moffat K, Šrajter V, et al. Volume-conserving trans-cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. *Nature Chemistry* 2013. DOI: 10.1038/nchem.1565.
152. Kupitz C, Olmos JL, Holl M, Tremblay L, Pande L, Pandey S, et al. Structural enzymology using X-ray free electron lasers. *Structural Dynamics* 2017; 4(4). DOI: 10.1063/1.4972069.
153. Schmidt M. Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography. *Advances in Condensed Matter Physics* 2013. DOI: 10.1155/2013/167276.
154. Schmidt M. Reaction initiation in enzyme crystals by diffusion of substrate. *Crystals* 2020.

- DOI: 10.3390/cryst10020116.
155. Liu RSH, Asato AE. The primary process of vision and the structure of bathorhodopsin: A mechanism for photoisomerization of polyenes. *Proceedings of the National Academy of Sciences of the United States of America* 1985. DOI: 10.1073/pnas.82.2.259.
  156. Warshel A, Barboy N. Energy Storage and Reaction Pathways in the First Step of the Vision Process. *Journal of the American Chemical Society* 1982. DOI: 10.1021/ja00370a003.
  157. Devanathan S, Pacheco A, Ujj L, Cusanovich M, Tollin G, Lin S, et al. Femtosecond spectroscopic observations of initial intermediates in the photocycle of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Biophysical Journal* 1999. DOI: 10.1016/S0006-3495(99)76952-3.
  158. Nakamura R, Hamada N, Ichida H, Tokunaga F, Kanematsu Y. Coherent oscillations in ultrafast fluorescence of photoactive yellow protein. *Journal of Chemical Physics* 2007. DOI: 10.1063/1.2802297.
  159. Creelman M, Kumauchi M, Hoff WD, Mathies RA. Chromophore dynamics in the PYP photocycle from femtosecond stimulated raman spectroscopy. *Journal of Physical Chemistry B* 2014. DOI: 10.1021/jp408584v.
  160. Schmidt M, Srajer V, Henning R, Ihee H, Purwar N, Tenboer J, et al. Protein energy landscapes determined by five-dimensional crystallography. *Acta Crystallographica Section D: Biological Crystallography* 2013. DOI: 10.1107/S0907444913025997.
  161. Barends TRM, Foucar L, Ardevol A, Nass K, Aquila A, Botha S, et al. Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* 2015. DOI: 10.1126/science.aac5492.
  162. Nango E, Royant A, Kubo M, Nakane T, Wickstrand C, Kimura T, et al. A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* 2016. DOI: 10.1126/science.aah3497.
  163. Nogly P, Weinert T, James D, Carbajo S, Ozerov D, Furrer A, et al. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* 2018. DOI: 10.1126/science.aat0094.
  164. Kern J, Alonso-Mori R, Tran R, Hattne J, Gildea RJ, Echols N, et al. Simultaneous femtosecond x-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* 2013. DOI: 10.1126/science.1234273.
  165. Kupitz C, Basu S, Grotjohann I, Fromme R, Zatsepin NA, Rendek KN, et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* 2014. DOI: 10.1038/nature13453.
  166. Young ID, Ibrahim M, Chatterjee R, Gul S, Fuller FD, Koroidov S, et al. Structure of photosystem II and substrate binding at room temperature. *Nature* 2016. DOI: 10.1038/nature20161.
  167. Michihiro S, Fusamichi A, Michihiro S, Minoru K, Yoshiki N, Takanori N, et al. Light-induced structural changes and the site of O=O bond formation in PSII caught by XFEL. *Nature* 2017. DOI: 10.2210/pdb5ws5/pdb.
  168. Claesson E, Wahlgren WY, Takala H, Pandey S, Castillon L, Kuznetsova V, et al. The primary structural photoresponse of phytochrome proteins captured by a femtosecond x-ray laser. *ELife* 2020. DOI: 10.7554/eLife.53514.
  169. Lincoln CN, Fitzpatrick AE, Thor JJV. Photoisomerisation quantum yield and non-linear cross-sections with femtosecond excitation of the photoactive yellow protein. *Physical Chemistry Chemical Physics* 2012. DOI: 10.1039/c2cp41718a.

170. Hutchison CDM, Kaucikas M, Tenboer J, Kupitz C, Moffat K, Schmidt M, et al. Photocycle populations with femtosecond excitation of crystalline photoactive yellow protein. *Chemical Physics Letters* 2016. DOI: 10.1016/j.cplett.2016.04.087.
171. Foucar L, Barty A, Coppola N, Hartmann R, Holl P, Hoppe U, et al. CASS - CFEL-ASG software suite. *Computer Physics Communications* 2012. DOI: 10.1016/j.cpc.2012.04.023.
172. Yoon CH. Psocake: GUI for Making Data Analysis a Piece of Cake, 2020. DOI: 10.1142/9789811204579\_0010.
173. Shin H, Kim S, Yoon CH. Data Analysis using Psocake at PAL-XFEL. *Journal of the Korean Physical Society* 2018. DOI: 10.3938/jkps.73.16.
174. Mariani V, Morgan A, Yoon CH, Lane TJ, White TA, O'grady C, et al. OnDA: Online data analysis and feedback for serial X-ray imaging. *Journal of Applied Crystallography* 2016. DOI: 10.1107/S1600576716007469.
175. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallographica Section D: Biological Crystallography* 2010. DOI: 10.1107/S0907444909042589.
176. Barends TRM, Foucar L, Botha S, Doak RB, Shoeman RL, Nass K, et al. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* 2014. DOI: 10.1038/nature12773.
177. Nakane T, Song C, Suzuki M, Nango E, Kobayashi J, Masuda T, et al. Native sulfur/chlorine SAD phasing for serial femtosecond crystallography. *Acta Crystallographica Section D: Biological Crystallography* 2015. DOI: 10.1107/S139900471501857X.
178. Yamashita K, Pan D, Okuda T, Sugahara M, Kodan A, Yamaguchi T, et al. An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Scientific Reports* 2015. DOI: 10.1038/srep14017.
179. Li C, Li X, Kirian R, Spence JCH, Liu H, Zatsepin NA. SPIND: A reference-based auto-indexing algorithm for sparse serial crystallography data. *IUCrJ* 2019. DOI: 10.1107/S2052252518014951.
180. Daurer BJ, Hantke MF, Nettelblad C, Maia FRNC. Hummingbird: Monitoring and analyzing flash X-ray imaging experiments in real time. *Journal of Applied Crystallography* 2016. DOI: 10.1107/S1600576716005926.
181. Yoon CH, Schwander P, Abergel C, Andersson I, Andreasson J, Aquila A, et al. Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering. *Optics Express* 2011. DOI: 10.1364/oe.19.016542.
182. Ayyer K, Morgan AJ, Aquila A, DeMirci H, Hogue BG, Kirian RA, et al. Low-signal limit of X-ray single particle diffractive imaging. *Optics Express* 2019; 27(26): 37816. DOI: 10.1364/OE.27.037816.
183. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, et al. Outcome of the first electron microscopy validation task force meeting. *Structure*, 2012. DOI: 10.1016/j.str.2011.12.014.
184. Chapman HN, Barty A, Marchesini S, Noy A, Hau-Riege SP, Cui C, et al. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *Journal of the Optical Society of America A* 2006. DOI: 10.1364/josaa.23.001179.
185. Shen Z, Teo CZW, Ayyer K, Loh ND. An encryption-decryption framework for validating single-particle imaging 2020.
186. Mancuso AP, Aquila A, Batchelor L, Bean RJ, Bielecki J, Borchers G, et al. The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european XFEL: Initial installation. *Journal of Synchrotron Radiation* 2019. DOI:

- 10.1107/S1600577519003308.
187. Schmidt M, Rajagopal S, Ren Z, Moffat K. Application of singular value decomposition to the analysis of time-resolved macromolecular x-ray data. *Biophysical Journal* 2003. DOI: 10.1016/S0006-3495(03)75018-8.
  188. Rajagopal S, Schmidt M, Anderson S, Ihee H, Moffat K. Analysis of experimental time-resolved crystallographic data by singular value decomposition. *Acta Crystallographica Section D: Biological Crystallography* 2004. DOI: 10.1107/S0907444904004160.
  189. Schmidt M, Pahl R, Srajer V, Anderson S, Ren Z, Ihee H, et al. 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* 2004. DOI: 10.1073/pnas.0305983101.
  190. Botha S, Nass K, Barends TRM, Kabsch W, Latz B, Dworkowski F, et al. Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallographica Section D: Biological Crystallography* 2015. DOI: 10.1107/S1399004714026327.
  191. Nogly P, James D, Wang D, White TA, Zatsepin N, Shilova A, et al. Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* 2015. DOI: 10.1107/S2052252514026487.
  192. Martin-Garcia JM, Conrad CE, Nelson G, Stander N, Zatsepin NA, Zook J, et al. Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. *IUCrJ* 2017. DOI: 10.1107/S205225251700570X.
  193. Weinert T, Olieric N, Cheng R, Brünle S, James D, Ozerov D, et al. Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nature Communications* 2017. DOI: 10.1038/s41467-017-00630-4.
  194. Berntsen P, Hadian Jazi M, Kusel M, Martin A V., Ericsson T, Call MJ, et al. The serial millisecond crystallography instrument at the Australian Synchrotron incorporating the “lipidico” injector. *Review of Scientific Instruments* 2019. DOI: 10.1063/1.5104298.
  195. Meents A, Wiedorn MO, Srajer V, Henning R, Sarrou I, Bergtholdt J, et al. Pink-beam serial crystallography. *Nature Communications* 2017. DOI: 10.1038/s41467-017-01417-3.
  196. Martin-Garcia JM, Zhu L, Mendez D, Lee MY, Chun E, Li C, et al. High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. *IUCrJ* 2019. DOI: 10.1107/S205225251900263X.
  197. Schulz EC, Mehrabi P, Müller-Werkmeister HM, Tellkamp F, Jha A, Stuart W, et al. The hit-and-return system enables efficient time-resolved serial synchrotron crystallography. *Nature Methods* 2018. DOI: 10.1038/s41592-018-0180-2.
  198. Mehrabi P, Schulz EC, Agthe M, Horrell S, Bourenkov G, von Stetten D, et al. Liquid application method for time-resolved analyses by serial synchrotron crystallography. *Nature Methods* 2019. DOI: 10.1038/s41592-019-0553-1.
  199. Chen B, Frank J. Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. *Microscopy (Oxford, England)* 2016. DOI: 10.1093/jmicro/dfv344.
  200. Lu Z, Shaikh TR, Barnard D, Meng X, Mohamed H, Yassin A, et al. Monolithic microfluidic mixing-spraying devices for time-resolved cryo-electron microscopy. *Journal of Structural Biology* 2009. DOI: 10.1016/j.jsb.2009.08.004.

## 6. Appendices

### Appendix A: Justification for Calculating Joint Probability

For the calculation of the joint probability  $\tilde{P}$  we regarded the statistics for each voxel to be independent of all other voxels, as it is the case for independent dice. However, as a diffraction pattern samples a set of voxels, entirely determined by the orientation of the molecule, the assumption of independence is not granted. To estimate the effect on our formalism, we calculated the joint probability  $\tilde{P}$  by simulation as follows. We recorded the actual number of visits for each voxel by merging 4323 diffraction patterns of the phytochrome at 10Å (see Table 3.1), where the number of patterns was estimated by Eq. (2.23) and Eq. (2.24), based on independent voxels with  $M = 39$ . Out of a total of 27 trials, we found 15 instances with all voxels visited at least  $M$  times. This corresponds to a probability of 0.55. As the statistical error expected from the number of trials is about  $\pm 0.1$ , this estimation is within the value predicted by the assumption of independent voxels. We, therefore, conclude that our calculation of the joint probability, as given by Eq. (2.24), is sufficient for the purpose of the present work.

### Appendix B: Python Code to Estimate the Exact Number of Snapshots

```
import numpy as np

def logFactorial(n):
    if n < 20 :
        value = np.log(np.math.factorial(n))
    else:
        value = 0.5*np.log(2*np.pi*n) + n*np.log(n/np.e)
    return value

def pnm(p,N,M):
    #this function gives the probability to observe a voxel at least M times
```



```

#from an ensemble of N snapshots
# p = probability to hit a voxel for a single snapshot
# N = Number of Snapshots
# M = Redundancy
if N < M:
    s = 1
else:
    s = 0
    lp = np.log(p)
    lmp = np.log(1-p)
    for k in np.arange(M):
        s = s + np.exp(logFactorial(N) - logFactorial(N-k) - logFactorial(k) + k*lp + (N-k)*lmp)
return np.maximum(1-s,0)

def numberOfSnapShots(d,D,nPhotons,SNR,P_tilde):
    #nS number of Snapshots
    #d is the Resolution
    # D is the Diameter of a Molecule
    #nPhotons is the Number of Photons
    #P_tilde is the combined probability
    #Number of Resolution elements
    R = D/d
    #number of voxels at Resolution Shell
    nV_Shell = 16*np.pi*R**2

    #probability per Shannon voxel
    p = 1./(4*R)
    M = np.ceil(SNR**2/nPhotons)

    # P -> Probability to observe a voxel at least M times from an ensemble of nS snapshot
    # obtained from given P_tilde
    P = np.exp(2*np.log(P_tilde)/nV_Shell)

    nSmax = 1e12#
    step = 2**10#

    nS0 = M
    while step > 1:
        for nS in np.arange(nS0,nSmax,step):
            if pnm(p,nS,M) > P:
                break
        nS0 = nS - step
        step = step /2
    return nS

```

## Appendix C: Random Phasor Approximation

The complex-valued random variables that arise as a sum of many small elementary complex-valued contributions can be approximated by a random phaser approximation [66].

Consider a large number  $N$  of complex phasors, the  $k^{\text{th}}$  phasor having random length  $\alpha_k$ , and random phase  $\phi_k$ . The resultant phasor, with length  $\alpha$  and phase  $\phi$ , is defined as

$$\mathbf{A} = \alpha e^{i\phi} = \sum_{k=1}^N \alpha_k e^{i\phi_k} . \quad (5.1)$$

For simplicity, consider the following assumptions about the statistical properties of the elementary phasors composing the sum.

1. The  $\alpha_k$  and phase  $\phi_k$  of the  $k^{\text{th}}$  elementary phasor are statistically independent of each other and of the amplitudes and phases of all other elementary phasors.
2. The random variables  $\alpha_k$  are identically distributed for all  $k$ .
3. The phases  $\phi_k$  are uniformly distributed on  $(-\pi, +\pi)$ .

The mean values of the real ( $r$ ) and imaginary parts ( $i$ ) are calculated as follows:

$$\bar{r} = \sum_{k=1}^N \overline{\alpha_k \cos \phi_k} = \sum_{k=1}^N \bar{\alpha}_k \overline{\cos \phi_k} = N \bar{\alpha} \overline{\cos \phi} \quad (5.2)$$

$$\bar{i} = \sum_{k=1}^N \overline{\alpha_k \sin \phi_k} = \sum_{k=1}^N \bar{\alpha}_k \overline{\sin \phi_k} = N \bar{\alpha} \overline{\sin \phi} . \quad (5.3)$$

Here we have used the facts that  $\alpha_k$  and  $\phi_k$  are independent and identically distributed for all  $k$ .

With the assumption that the phases  $\phi_k$  are uniformly distributed on  $(-\pi, +\pi)$ , then  $\overline{\sin \phi} = \overline{\cos \phi} = 0$ . With this

$$\bar{r} = \bar{i} = 0 \quad (5.4)$$

Similarly, the second moments  $\overline{r^2}$  and  $\overline{i^2}$  can be written as

$$\overline{r^2} = \sum_{j=1}^N \sum_{k=1}^N \bar{\alpha}_j \bar{\alpha}_k \overline{\cos \phi_j \cos \phi_k} \quad (5.5)$$

$$\bar{i}^2 = \sum_{j=1}^N \sum_{k=1}^N \bar{\alpha}_j \bar{\alpha}_k \overline{\sin\phi_j \sin\phi_k} \quad (5.6)$$

Using,

$$\overline{\cos\phi_j \cos\phi_k} = \overline{\sin\phi_j \sin\phi_k} = \begin{cases} 0 & k \neq j \\ \frac{1}{2} & k = j, \end{cases}$$

$$\bar{r}^2 = \bar{i}^2 = N \frac{\bar{\alpha}^2}{2}. \quad (5.7)$$

Eq. (5.1) can be written as the sum of the real and the imaginary parts.

$$\mathbf{A} = \alpha e^{i\phi} = \text{Re}(\mathbf{A}) + \text{Im}(\mathbf{A}).$$

With this

$$\begin{aligned} \left| \sum_{k=1}^N \alpha_k e^{i\phi_k} \right|^2 &= \overline{(\text{Re}(\mathbf{A}) + \text{Im}(\mathbf{A}))^2} = \overline{(r + i)^2} \\ &= \overline{(r^2 + 2ri + i^2)} \\ &= \bar{r}^2 + 2\bar{r}\bar{i} + \bar{i}^2 \\ &= \bar{r}^2 + \bar{i}^2 = N\bar{\alpha}^2. \end{aligned}$$

Using this approximation, we can write

$$\left| \sum_{j=1}^{N_c} \exp(2\pi i \mathbf{q} \cdot \mathbf{X}_j) \right|^2 = N_c.$$

This assumes that the positions of atoms are completely uncorrelated, which is approximately true at resolutions approaching to atomic scale.

# Curriculum Vitae

## Ishwor Poudyal

University of Wisconsin-Milwaukee, Department of Physics, Kenwood Interdisciplinary Research Complex (KIRC), 3135 North Maryland Ave., Milwaukee, WI, 53211

## EDUCATION

### PhD in Physics

2014-2020

*University of Wisconsin-Milwaukee*

*Dissertation: Macromolecular Structure Determination at XFELs from Single-Particle Imaging to Time-Resolved X-ray Crystallography*

### MSc in Physics

2010-2013

*Central Department of Physics, Tribhuvan University, Kirtipur Nepal*

*Dissertation: Molecular Dynamics Study of Diffusion of Carbon Monoxide in Water at Different Temperatures*

### BSc in Physics (Major: Physics, Minor: Mathematics, Statistics)

2007-2010

*Trichandra Multiple Campus, Tribhuvan University, Kathmandu, Nepal*

## PUBLICATIONS

1. S. Pandey\*, **I. Poudyal\***, T.N. Malla\* (2020) “*Pump-Probe Time-Resolved Serial Femtosecond Crystallography at X-ray Free Electron Lasers.*” *Crystals* 10.7, 628. (\*corresponding author)
2. **I. Poudyal**, M. Schmidt, and P. Schwander (2020). “*Single-particle imaging by x-ray free electron lasers – How many snapshots are needed?*” *Structural Dynamics* 7.2, 024102.
3. S. Pandey #, R. Bean#, T. Sato#, **I. Poudyal**, J. Bielecki, J. Cruz Villarreal, O. Yefanov, V. Mariani, T. A. White, C. Kupitz, M. Hunter, M. H. Abdellatif, S. Bajt, V. Bondar, A. Echelmeier, D. Doppler, M. Emons, M. Frank, R. Fromme, Y. Gevorkov, G. Giovanetti, M. Jiang, D. Kim, Y. Kim, H. Kirkwood, A. Klimovskaia, J. Knoska, F. H. M. Koua, R. Letrun, S. Lisova, L. Maia, V. Mazalova, D. Meza, T. Michelat, A. Ourmazd, G. Palmer, M. Ramilli, R. Schubert, P. Schwander, A. Silenzi, J. Sztuk-Dambietz, A. Tolstikova, H. N. Chapman, A. Ros, A. Barty, P. Fromme, A. P. Mancuso, *M. Schmidt* (2019) *Time-Resolved Serial Femtosecond Crystallography at the European XFEL. Nature Methods* 17.1, 73-78. (# equal contribution)
4. J. Olmos, S. Pandey, J. Martin-Garcia, G. Calvey, A. Katz, J. Knoska, C. Kupitz, M. S. Hunter, M. Liang, D. Oberthuer, O. Yefanov, M. Wiedorn, M. Heyman, M. Holl, K. Pande, A. Barty, M. D. Miller, S. Stern, S. Roy-Chowdhury, J. Coe, N. Nagaratnam, J. Zook, J. Verburgt, T. Norwood, **I. Poudyal**, D. Xu, J. Koglin, M. H. Seaberg, Y. Zhao, S. Bajt, T. Grant, V. Mariani, G. Nelson, G. Subramanian, E. Bae, R. Fromme, R. Fung, P.

Schwander, M. Frank, T. A. White, U. Weierstall, N. Zatsepin, J. Spence, P. Fromme, H. N. Chapman, L. Pollack, L. Trembley, A. Ourmazd, G. Phillips, *M. Schmidt* (2018) *Authentic Enzyme Intermediates Captured “on-the-fly” by Mix-and-Inject Serial Crystallography*, *BMC-Biology* 16 (1), 59.

5. N. C. Woitowich, A. S. Halavati, P. Waltz, C. Kupitz, J. Valera, G. Tracy, K.D. Gallagher, E. Claesson, T. Nakana, S. Pandey, G. Nelson, R. Tanaka, E. Nango, E. Mizohata, S. Owada, K. Tono, Y. Joti, A.C. Nugent, H. Patel, A. Mapara, J. Hopkins, P. Duong, D. Bizhga, S.E. Kovaleva, R. St. Peter, C.N. Hernandez, W.B. Ozarowski, S. Roy-Chowdhuri, J-H Yang, P. Edlund, H. Takala, J. Ihalainen, J. Brayshaw, T. Norwood, **I. Poudyal**, P. Fromme, J.C.H. Spence, K. Moffat, S. Westenhoff, *M. Schmidt\**, E. Stojkovic\* (2018) *Structural Basis for Light Control of Cell Development as Revealed by Crystal Structures of a Myxobacterial Phytochrome*, *IUCrJ* 5 (5).
6. C. Kupitz, J. L. Olmos Jr., M. Holl, L. Tremblay, K. Pande, S. Pandey, D. Oberthür, M. Hunter, M. Liang, A. Aquila, J. Tenboer, G. Calvey, A. Katz, Y. Chen, M.O. Wiedorn, J. Knoska, A. Meents, V. Mariani, T. Norwood, **I. Poudyal**, T. Grant, M. D. Miller, W. Xu, A. Tolstikova, A. Morgan, M. Metz, J. Martin-Gracia, J.D. Zook, S. Roy-Chowdhury, J. Coe, N. Nagaratnam, D. Meza, R. Fromme, S. Basu, M. Frank, T. White, A. Barty, S. Bajt, O. Yefanov, H.N. Chapman, N. Zatsepin, G. Nelson, U. Weierstall, J. Spence, P. Schwander, L. Pollack, P. Fromme, A. Ourmazd, G.N. Phillips Jr., *M. Schmidt* (2017) *Structural Enzymology Using X-ray Free Electron Lasers*, *Structural Dynamics* 4 (4), 044003.
7. P. Edlund, H. Takala, E. Claesson, L. Henry, R. Dods, H. Lehtivuori, M. Panman, K. Pande, T. White, T. Nakane, O. Berntsson, E. Gustavsson, P. Bath, V. Modi, S. Roy-Chowdhury, J. Zook, P. Berntsen, S. Pandey, **I. Poudyal**, J. Tenboer, C. Kupitz, A. Barty, P. Fromme, J. D. Koralek, T. Tanaka, J. Spence, M. Liang, M. S. Hunter, S. Boutet, E. Nango, K. Moffat, G. Groenhof, J. Ihalainen, E. A. Stojkovic, *M. Schmidt\**, S. Westenhoff\* (2016), *The room temperature crystal structure of a bacterial phytochrome determined by serial femtosecond crystallography*, *Scientific reports* 6, 35279.
8. **I. Poudyal**, N.P. Adhikari (2014) *Temperature dependence of diffusion coefficient of carbon monoxide in water: A molecular dynamics study*, *Journal of Molecular Liquids* 194, 77-8.

## CONFERENCES ATTENDED AND/OR POSTER PRESENTATION

- Participant/attendee at the NSF sponsored BioXFEL STC 7<sup>th</sup> Annual International Conference; 28-30 January 2020, San Juan, Puerto Rico.
- **Poster Presentation:** “Pushing XFEL Single-Particle Imaging towards Near-Atomic Resolution” at BioXFEL STC 7<sup>th</sup> Annual International Conference: 28-30 January 2020; San Juan; Puerto Rico.

- Participant/attendee at the Nature Conference *Functional dynamics – visualizing molecules in action*; 6-8 November 2019, Tempe, AZ.
- **Poster Presentation:** “Proteins and Enzymes: How many Single-Particle Snapshots we Need?” at Nature Conference: 6-8 November 2019; Tempe, AZ.
- Participant/attendee at the NSF sponsored BioXFEL STC 6<sup>th</sup> Annual International Conference;  
12 -14 February 2019; San Diego, CA.
- Participant / attendee at the NSF sponsored BioXFEL STC 5<sup>th</sup> Annual International Conference;  
13-15 February 2018; New Orleans, LA.
- **Poster Presentation:** “Single Particle Time-Resolved Diffraction with Bacterial Phytochromes” at BioXFEL STC 5<sup>th</sup> Annual International Conference: 13-15 February 2018; New Orleans; LA.
- Participant / attendee at the European XFEL Data Analysis Workshop; 16 February 2018; New Orleans, LA.
- Participant / attendee at the NSF sponsored BioXFEL STC 4<sup>th</sup> Annual International Conference;  
10-12 January 2017; Las Vegas, NV.
- **Poster Presentation:** “X-ray Free Electron Laser (XFEL) in Protein Crystallography” at BioXFEL STC 4<sup>th</sup> Annual International Conference: 10-12 January 2017; Las Vegas; NV
- Participant / attendee at the BioXFEL Single Particle Sample Delivery Workshop; 9 January 2107; Las Vegas, NV.
- Participant / attendee at the UWM BioXFEL Fall Workshop; 29 September 2015; Milwaukee, WI.