

August 2017

# Difference Structures from Time-resolved SAXS/ WAXS

Prakash Nepal

*University of Wisconsin-Milwaukee*

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



Part of the [Physics Commons](#)

---

## Recommended Citation

Nepal, Prakash, "Difference Structures from Time-resolved SAXS/WAXS" (2017). *Theses and Dissertations*. 1675.  
<https://dc.uwm.edu/etd/1675>

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).

DIFFERENCE STRUCTURES FROM  
TIME-RESOLVED SAXS/WAXS

by

Prakash Nepal

A Dissertation Submitted in  
Partial Fulfillment of the  
Requirements for the Degree of

Doctor of Philosophy  
in Physics

at

The University of Wisconsin-Milwaukee

August 2017

ABSTRACT

# DIFFERENCE STRUCTURES FROM TIME-RESOLVED SAXS/WAXS

by

Prakash Nepal

The University of Wisconsin-Milwaukee, 2017  
Under the Supervision of Professor Dilano Kerzaman Saldin

It has been thought that there is not enough information to recover a solution to the problem of inverting time-resolved Small Angle X-ray Scattering (SAXS) / Wide Angle X-ray Scattering (WAXS) difference data. In this context, this study presents a new idea by extending the concept of “difference Fourier method” which aims to overcome the above limitation in the idealized case of simulated data.

If the structural changes in a typical time-resolved studies are generally confined to a relatively localized region, they can be found from difference SAXS/WAXS curves and knowledge of the atomic coordinates of one of the states. The method has some similarities to the difference Fourier method of crystallography, which may not be directly applicable to SAXS/WAXS analysis.

This study demonstrates the identification of the displaced residues once this pro-

posed idea/method is applied to structural changes involved in Photoactive Yellow Protein (PYP), which has been used as a test case in time-resolved studies.

# TABLE OF CONTENTS

<b>Abstract</b>	<b>ii</b>
<b>Table of Contents</b>	<b>iii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>x</b>
<b>Acknowledgments</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Time-resolved SAXS/WAXS and XFEL . . . . .	1
<b>2 Basic Theory</b>	<b>13</b>
2.1 X-ray Scattering from an Atom . . . . .	13
2.2 Scattering from a Molecule . . . . .	15
2.3 Debye Formula for Scattering Profiles . . . . .	18
2.4 Spherical Harmonics . . . . .	19
2.5 Guiniers Law in SAXS . . . . .	22
2.6 Porod's Law in SAXS . . . . .	24
2.7 SAXS Intensity from Scattered Patterns . . . . .	25
2.8 Spherical Bessel Function . . . . .	26

<b>3</b>	<b>Theory for Structural Changes</b>	<b>28</b>
3.1	Rapid direct method for extraction of difference electron density . . . . .	28
3.2	Matrix Inversion with SVD (Singular Value Decompostion) . . . . .	33
3.3	Simulated Annealing . . . . .	34
<b>4</b>	<b>Results</b>	<b>38</b>
4.1	Photoactive Yellow Protein(PYP) . . . . .	38
4.2	Output of the rapid direct method . . . . .	40
4.3	Results with Simulated Annealing . . . . .	52
4.4	Consideration of Poisson Noise . . . . .	59
4.5	Experimental SAXS/WAXS data . . . . .	66
4.6	Limitations of this proposed method . . . . .	66
<b>5</b>	<b>Conclusion and Outlook</b>	<b>69</b>
	<b>References</b>	<b>73</b>
	<b>Appendices</b>	<b>84</b>
<b>A</b>	<b>Protein Data Bank Format</b>	<b>85</b>
<b>B</b>	<b>Pseudocode</b>	<b>87</b>
	<b>Curriculum Vitae</b>	<b>89</b>

# LIST OF FIGURES

1.1	(a) The photosynthetic reaction center of RCvir <i>Blastochloris viridis</i> showing a complex of four subunits H, L, M C, reaction center cofactors (BPh, BCh, P960, QA, QB, Hemes) and membrane boundaries specified by blue lines, (b) Time-resolved SAXS/WAXS difference data recorded as a function of time delay, obtained from [12] . . . . .	9
1.2	(a) Representation of Mb structure showing haem, covalently bounded helix F with haem, parts of helices H,G,E,C and A. The green wave is the representation of light which induces photolysis of the bond between protein and ligand. The yellow arrows showing initiation of structural changes after photolysis. The polypeptide chain segments associated with collective mode of a $\sim 4 ps$ period after ligand release are highlighted by red. (b) Time-resolved SAXS/WAXS difference data recorded as a function of time delay for MbCO in solution, obtained from [32] . . . . .	10
1.3	Time dependency of Mb structural parameters(change in Radius of gyration and change in volume of protein), obtained from [32]. The red curve showing a damped oscillation with a $3.6 ps$ time period and $6 ps$ decay time. The two perpendicular lines within $2 ps$ time period shows that the increase in protein volume is delayed with respect to $R_g$ increase. . . . .	11
2.1	X-ray Scattering . . . . .	14

2.2	Plot of atomic form factor for carbon and oxygen . . . . .	16
2.3	Example of data from protein data bank in pdb format . . . . .	17
2.4	Spherical harmonics plot characterized with different quantum numbers .	21
2.5	Figure showing experimental scattered pattern along with SAXS curve . .	25
2.6	Plot of first three spherical bessel function of first kind. . . . .	27
4.1	Unexcited PYP structure with chromophore . . . . .	39
4.2	(a) Theoretical(Simulated) SAXS/WAXS Intensities for unexcited and excited PYP structure with chromophore swing (b)Logarithmic plot of theoretical(simulated) SAXS/WAXS Intensities for unexcited and excited PYP structure with chromophore swing (c) Difference in theoretical SAXS/WAXS Intensity for chromophore swing case. $q$ values are in crystallographic covention with unit $\text{\AA}^{-1}$ . . . . .	41
4.3	Electron density differences between the photoexcited PYP [67] assumed and the “dark” structure of PYP(PDB ID 2phy)[68] from step1 i.e. rapid direct method of algorithm. In this step only the highest and lowest values of the density differences is used so the contour level is chosen so as to show only these. Consistent with the convention in time-resolved crystallography, the positive difference electron density (towards which the chromophore moves) is indicated by the green lobe and the negative electron density (from which the chromophore has moved) is indicated by red. . . .	42
4.4	Slices through the difference in density(top to bottom) obtained from matrix inversion during chromophore swing, red as postive difference density and blue as negative difference density. . . . .	43
4.5	Slices through the difference in density(top to bottom) during chromophore swing obtained from matrix inversion when trucated at another mode, red as postive difference density and blue as negative difference density. . . . .	44

4.6	Slices through the difference in density(top to bottom) during chromophore swing obtained from matrix inversion when truncated at other mode different than earlier two, red as postive difference density and blue as negative difference density. . . . .	45
4.7	Difference in Theoretical(simulated) SAXS/WAXS Intensity for ARG124 Change. $q$ values are in crystallographic covention with unit $\text{\AA}^{-1}$ . . . . .	46
4.8	Same as Fig. 4.3 except that the same difference electron density is found for the movement of another residue by step 1 i.e. rapid direct method described in this study. . . . .	47
4.9	Slices through the differnce in density(top to bottom) obtained from matrix inversion during Arginine shift, red as positive difference density and blue as negative difference density. . . . .	48
4.10	Slices through the difference in density(top to bottom) during Arginine torsional shift obtained from matrix inversion truncated at another mode, red as positive difference density and blue as negative difference density. . .	49
4.11	Slices through the differnce in density(top to bottom) during Arginine torsional shift obtained from matrix inversion truncated at other mode different than earlier two, red as positive difference density and blue as negative difference density. . . . .	50
4.12	(a)Same as Fig. 4.3, but after two-step approach of this study i.e. rapid direct method and Simulated Annealing algorithm, (b) same as (a) except from the difference Fourier method of classical time-resolved crystallography. . . . .	53
4.13	(a) Same as Fig. 4.8 except that the difference electron density is now recovered after two-step approach of this study i.e. rapid direct method and Simulated Annealing algorithm, (b) same as (a) except from the difference Fourier method of classical time-resolved crystallography. . . . .	54

4.14	Figure showing atomic coordinates of the chromophore in unperturbed, perturbed states, and recovered by simulated annealing . . . . .	57
4.15	Figure showing atomic coordinates of the ARG124 in unperturbed, perturbed states, and recovered by simulated annealing . . . . .	58
4.16	Figure showing outer resolution ring . . . . .	60
4.17	Difference in Simulated SAXS/WAXS intensity after Poisson noise treatment for different snapshots. One photon per Shannon pixel is considered in this case. The noise free difference SAXS/WAXS intensity values are weighted by a factor of $10^{-3}$ . . . . .	60
4.18	Zoomed plot of Fig 4.17 for difference in Simulated SAXS/WAXS intensity after Poisson noise treatment for different snapshots. The noise free difference SAXS/WAXS intensity values are weighted by a factor of $10^{-3}$ . . . . .	61
4.19	(a) For 13,000,000 photons per Shannon Pixel corresponding to billion particles and with 10,000 snapshots, (b) For 130,000 photons per Shannon Pixel corresponding to ten million particles and with 10,000 snapshots. . . . .	62
4.20	(a) For 13,000 photons per Shannon Pixel corresponding to million particles and with 10,000 snapshots, (b) For 1300 photons per Shannon Pixel corresponding to hundred thousand particles and with 10,000 snapshots. . . . .	63
4.21	(a) For 13,000 photons per Shannon Pixel corresponding to million particles and with 1 million snapshots, (b) For 1300 photons per Shannon Pixel corresponding to hundred thousand particles and with 1 million snapshots. . . . .	64
4.22	For 130 photons per Shannon Pixel corresponding to ten thousand particles and with 1 million snapshots. . . . .	65
4.23	Comparison of Simulated SAXS/WAXS intensity (panel a) to experimentally measured SAXS/WAXS Intensity (panel b) . . . . .	67

# LIST OF TABLES

2.1	Table of Cromer-Mann coefficients . . . . .	15
4.1	Peaks for Chromophore swing case in terms of root mean square deviation(r.m.s.d.) of map . . . . .	51
4.2	Peaks for ARG124 shift case in terms of root mean square deviation(r.m.s.d.) of map . . . . .	52
4.3	Atom coordinates of the chromophore of the unperturbed (u) structure. . .	55
4.4	Atom coordinates of the chromophore of the perturbed (p) structure. . . .	56
4.5	First set of atomic coordinates of the chromophore recovered by simulated annealing, starting from the known unperturbed (u) structure. A total of ten sets of chromophore atomic coordinates recovered by simulated annealing are shown in Figure 4.14 . . . . .	56
4.6	Atom coordinates of the ARG124 of the unperturbed (u) structure. . . . .	56
4.7	Atom coordinates of the ARG124 of the perturbed (p) structure. . . . .	56
4.8	First set of atomic coordinates of the ARG124 recovered by simulated annealing, starting from the known unperturbed (u) structure. A total of nine sets of atomic coordinates of the ARG124 recovered by simulated annealing are shown in figure 4.15 . . . . .	57
A.1	Explanation of the format of pdb file [73] . . . . .	86

# ACKNOWLEDGMENTS

My time in the Physics Department at the University of Wisconsin-Milwaukee has provided many wonderful experiences and opportunities and I am truly appreciative of being able to complete my studies and research here! There are many people that I would like to acknowledge and thank, including:

Distinguished Professor Dilano K. Saldin (my PhD advisor/committee chair) – overall and mostly for his guidance and support. Prof. Saldin helped me to develop this method for difference structures in SAXS / WAXS, which I am confident will make a positive impact in this field. To the other members of my PhD committee: Professor Marius Schmidt, for encouraging me to work diligently, which ultimately helped me to be more competitive and for his suggestions to the work for this dissertation; Professor Peter Schwander for his thoughtfulness, suggestions and hints relative to pushing my research forward and for being available to answer my numerous questions; and to Distinguished Professor Michael Weinert and Professor Prasenjit Guptasarma (also the Department Chair), for their willingness to join my PhD committee and for providing very helpful remarks which contributed greatly to the final success of this dissertation. Also, I am grateful to Distinguished Professor Abbas Ourmazd for his openness and suggestions regarding my research and plans for the future.

Many former and current UWM Physics graduate students, post docs, professors and staff members have also helped me, including: Professor Patrick R Brady, Dr. Richard O'Shaughnessy (former research associate and now professor at Rochester Institute of Technology) and Dr. Justin Ellis (former PhD student), all of whom were in a completely different research area (gravity/ cosmology/astrophysics) but who helped me to be a more

independent researcher; Dr. Sandi Wibowo and Dr. Sung Soon Kim (both who also received their PhD from UWM Physics) for their help during my academic career and for the friendship and helpful suggestions in many other things not related to my academics; and finally to Ms. Kate Valerius (UWM Physics Graduate Program Assistant) for her unwavering commitment to helping me from the very beginning, when I applied to the program, to my arrival to the US and the department and now through the completion of my PhD. She has always been encouraging, has taught me to be proactive and has given me much confidence in moving forward.

To my parents . . . Baburam Sharma (my father) and Sita Devi Upadhyani (my mother), I cannot express enough gratitude for your continuous support and trust in me! Your hard work ethics and ability to foresee positive things has had a profound impact on me. Also, to Suman (my brother), for his care and support, especially during my UG/MSc college years in Nepal.

And finally, to my beloved homeland of Nepal and its people, which I miss each and every day. I dedicate my work and this dissertation to you.

# Chapter 1

## Introduction

### 1.1 Time-resolved SAXS/WAXS and XFEL

Small angle X-ray scattering(SAXS) is an elastic scattering of X-rays by a sample recorded at low angles. This technique is widely used to get the information about the shape and size of biomolecules/proteins between 50 Å to 250 Å , multi-molecular assemblies and even in partially ordered systems of up to 1500 Å [1]. The X-ray source can be synchrotron light or a laboratory source capable of providing higher X-ray flux. The scattering pattern is formed by scattered X-rays recorded at 2D detector. One of the major problems in SAXS instrumentation is the separation of the weak scattered intensity from the strong main beam [2].

Biological molecules normally scatter weakly and the measured solutions are dilute. The random positions and orientations of particles lead to intensity distribution which for non-interfering particles is proportional to the scattering from a single particle averaged over all orientations. At low angles the intensity curve is rapidly decaying function of the scattering vector. Wide angle X-ray scattering (WAXS) technique refers to the study of scattered radiation at wide angles typically  $\geq 10^\circ$  . The detector to sample distance

is smaller and diffraction maxima at larger angle are observed. With a special measuring instrument, it is possible to perform SAXS and WAXS experiments in a single run.

The method for calculating scattering profile was initially proposed by Debye in 1915[3]. The main principles of SAXS were developed by Guinier following his studies on metallic alloys in the late 1930s[4]. In the 1960s, the method became important in the study of biomolecules/proteins in solution as it helped one to get a low-resolution structural envelope. After the availability of synchrotron radiation in the 1970's there was a breakthrough in SAXS experiments. The main challenge of SAXS as a biophysical technique is to extract information about the three-dimensional structure of the object from one-dimensional experimental data.

Previously, overall only structural parameters like radius of gyration, volume of the macromolecules were obtained directly from the experimental data, while the analysis in terms of three-dimensional models was limited to simple geometrical bodies(e.g. spherical, cylindrical, etc.) or may be performed on trial-and-error-basis. In a good SAXS experiment, a number of solutions with varying concentrations of biomolecules under study are measured. With the extrapolated scattered curves measured at different concentrations to zero concentration, it is possible to obtain a scattering curve at infinite dilution.

Another breakthrough in SAXS data analysis is in 1990's which opened the way for reliable ab initio modelling of the shape and domain structure of biological complexes. The project called Small-Angle X-ray scattering Initiative for Europe (SAXIER) created a unified European SAXS infrastructure with the aim to combine advanced SAXS methods along with analytical techniques to create reliable software to rapidly analyse large quantities of data. Several SAXS analysis software have been intensely developed and freely available nowadays.

In the ab initio modelling [5], an angular envelope function of the particle is described by a series of spherical harmonics. The low resolution shape can be obtained by the coefficients of the series which fit the scattering data. The ab initio approach with some refinement is implemented in the computer program called SASHA (Small Angle Scattering Shape Determination) [6]. Under certain conditions it was shown that a unique envelope can be obtained from scattering data [7]. This approach is relevant for globular particles with relatively simple shapes. SAXS data can also work as a structural constraint on techniques utilized to determine the conformational fold of the protein [8].

The amount of data from solution X-ray scattering is confined to that from a single curve of intensity as a function of scattering angle,  $I(q)$ . Two regions, small-angle X-ray scattering (SAXS) and wide-angle X-ray scattering (WAXS) lend themselves to different types of analysis and interpretation. Generally SAXS spectra are good for obtaining a radius of gyration and perhaps a relatively low-resolution envelope of the molecule and WAXS data provide a very sensitive measure of structural changes in the system.

As has been shown by Stuhrmann, [5, 9] and Svergun and Stuhrmann, [10] there is enough information in an experimental SAXS spectrum, and methods have been standardized, to obtain at least a low resolution structure of the molecule being studied. According to Svergun and Stuhrmann's opinion, the width of a data element in a SAXS spectrum is approximately the size of a Shannon pixel,  $1/2D$ , where  $D$  is the diameter of the molecule. Given the span of a typical SAXS curve this normally translates to about twenty pieces of information, enough for the extraction of the spherical harmonic expansion coefficients of the electron density of the molecule up to the equivalent of an angular momentum quantum number of perhaps  $l=3$ . This then describes a low resolution structure and not much detail.

Time-resolved small angle/wide angle x-ray scattering (TR-SAXS/WAXS) studies of protein structural dynamics is an active and progressing field of research which utilizes both synchrotron radiation and x-ray free electron radiation [11]. The beauty of TR-SAXS/WAXS is that it allows different time-scales and a number of independent conformational states to be observed following a rapid reaction triggering event (exclusively a laser-flash to date). Since TR-SAXS/WAXS does not require crystals, the method also overcomes many of the limitations of time-resolved crystallography.

TR SAXS/WAXS is highly relevant for following the dynamics of any system where two data sets exist and one wants a detailed comparison. One example would be the “protein quake” [12] postulated to explain fast laser light absorption events. The postulate is that the extra energy is dissipated by the formation of a transient protein quake within the molecule. For these reasons, TR-SAXS/WAXS has become an important biophysical technique and is expected to become more widely applied in the future.

Time-resolved protein crystallography [13] is based on the difference Fourier approximation in which the phases of a perturbed ( $p$ ) state are approximated by the phases of a similar unperturbed ( $u$ ) structure, and can include extrapolated calculations in the complex plane of the structure factors [14]. Since the amplitudes are found by experiment, this enables the difference electron density to be found by three dimensional Fourier transformation and hence the time variation of a structure. The difference Fourier method is not directly applicable to SAXS/WAXS as the basic experimental data consists of the averages of intensities over all orientations.

A severe limitation of TR-SAXS/WAXS is that the structural information accessible to the experimentalist is only one-dimensional (the change in X-ray scattering,  $\delta I(q)$ ),

versus the X-ray scattering length,  $q$ ). It has been assumed by virtually all researchers in the field to imply that there is no unique solution to the structure of the changed conformation, and there has been increasingly structural results from the basis of recurring structural changes extracted by fitting an ensemble average (typically generated from molecular dynamics simulations [12]) to the difference X-ray scattering data; by extracting and applying candidate movements from the analysis of deposited X-ray structures; or a combination of both.

The theory presented in this dissertation is an attempt to develop the extraction of structural changes from time-resolved SAXS/WAXS. The difference Fourier method in crystallography has proved powerful in revealing small structural changes in a biological macromolecule. These structural changes usually arise either from chemical differences (e.g. +/- a ligand or heavy atom) or temporal differences (e.g. as its structure evolves after absorption of a photon by a photoreceptor). The method makes use of the (assumed known) X-ray phases and amplitudes of the unperturbed structure and the difference between the experimental structure amplitudes in the perturbed and unperturbed states,  $\delta F$ .

The goal of this dissertation is to extend the concept of “difference Fourier” to SAXS/WAXS data. Highly significant experimental differences in  $I(q)$  to large  $q$  exist e.g. of time dependent solution scattering data after a perturbation of PYP (Photoactive Yellow Protein). Objective interpretation of these differences is difficult since it has been assumed that there is no direct, invertible relationship between continuous scattering from a solution and the differences in structure.

In this context, this dissertation presents theoretical framework for a new idea which tries to overcome this limitation in the idealized case of simulated data and under the

assumption that the movements are small and localized. This may break away from conventional thinking that there is not sufficient data to recover a unique solution to the problem of inverting TR-SAXS/WAXS difference data. The main attempt is to show the mathematics showing how, in the approximation of a small perturbation, “the additional knowledge of the reference atomic structure from other methods can thus be brought into play to improve the extraction of information using this method.”

The method may be exact for very small structural changes enough for the change in the intensities to be linear must happen at some stage. It is only then the changes in the intensity are proportional to the change in the structure that the theory described in this dissertation is applicable. The method then becomes similar to the difference Fourier method of X-ray crystallography, an approach that might not previously thought of as being directly applicable to SAXS/WAXS analysis.

Locating the regions of the moved substructures by the method presented in this study exploits the Singular Value Decomposition(SVD) technique for matrix inversion. This allows one to apply the method to SAXS/WAXS data where the number of data points is much less than the number of real space points at which the electron density is sought. This dissertation illustrates the identification of the displaced residues in the result section with an application to conformational changes in photoactive yellow protein, which has been used as a test case in time-resolved crystallography [15, 16, 17, 18, 19].

Although it is quite easy to arrange a finely spaced grid of data points in crystallography at which one seeks the electron density, it is much more difficult in SAXS/WAXS due to the paucity of the data. This study suggest here that knowledge of the crystal structure can be used as an approximation to extract knowledge from scattering curves measured from solutions. Even if one can only obtain a resolution to the level of identify-

ing the general region in which atoms have changed their positions, it would be significant achievement. Since the average residue in a protein has only a relatively small number of atoms, it will not be too difficult to vary only the positions of these atoms with a global optimization algorithm such as simulated annealing [20] to get a reasonably accurate solution [21]. This type of atomic level fitting has, in fact been used to compare small changes in WAXS spectra [12].

Linac Coherent Light Source(LCLS)[22, 23] is the world's first free electron laser for hard X-rays. It could produce ultra-short pulses[24, 25, 26] of tens of femtosecond ( $fs$ ) at a rate of 120 pulses per second [27]. X-rays of ten billion times more powerful than what was available before can reveal a lot of unknown queries. Important vital proteins are very difficult to crystallize but the XFEL can be utilized with its immense brightness to obtain diffraction patterns of individual uncrystallized biomolecules. The advantage of immense brightness is that peak brightness happens only over a period of femtoseconds enabling the scattering of X-rays happening before molecule is destroyed. The diffraction patterns of undamaged samples can be recorded [28, 29]. Moreover, the advent of XFEL's such as LCLS has paved the way for ultrafast time-resolved studies [30, 31].

Researchers from different fields are benefitted greatly from XFELs. Different instruments of LCLS namely AMO( Atomic, Molecular Optical Science), CXI(Coherent X-ray Imaging), XPP(X-ray Pump Probe), MEC (Matter in Extreme Conditions), MFX (Macromolecular Femtosecond Crystallography, SXR (Soft X-ray Material Sciece) and XCS (X-ray Correlation Spectroscopy) and with the features described above, XFELs are providing new ways to reveal images of biomolecules, atomic and molecular motions, properties of materials and other fundamental information. The next step is the LCLS -II to ensure for advanced research in science. LCLS-II will provide a significant jump in capability from 120 pulses per second to 1 million pulses per second. Researchers from a

wide range of fields will be able to perform experiments that are now impossible.

In order to track biomolecules/proteins structural dynamics with high spatial and time resolution, several time-resolved SAXS/WAXS experiments have been performed at LCLS [12, 32]. A femtosecond laser pulse excites proteins and triggers a structural change. A suitably delayed X-ray pulse (according to the nature of studied object) passes the sample and time-resolved X-ray scattering images are recorded on the detector. The recorded two-dimensional scattering images with detector are azimuthally averaged to give one-dimensional scattering patterns. The scattering angle  $2\theta$  is converted to momentum transfer  $q$  with the formula in crystallographic convention as

$$q = (2 \sin \theta) / \lambda \tag{1.1}$$

with  $\lambda$  is the X-ray wavelength.

One of the celebrated experiment is by Neutze et al.[12] on *Blastochloris viridis* photosynthetic reaction center to measure ultrafast protein-structural changes using time-resolved small angle/wide-angle X-ray scattering at an XFEL. Although the “protein quake” phenomenon was already proposed, the first direct evidence of protein quake at LCLS X-ray free electron laser is observed with ultrafast conformational changes within picoseconds, and preceding the propagation of heat through protein. TR-SAXS/WAXS is directly related to protein structure and helps directly visualizing ultrafast protein motions and it is possible to derive conformational changes against difference SAXS/WAXS data.

Neutze et al. were able to show the carbon atoms of the protein helices increase their distance from the interior of the photoreaction centre on the picosecond time scale. With

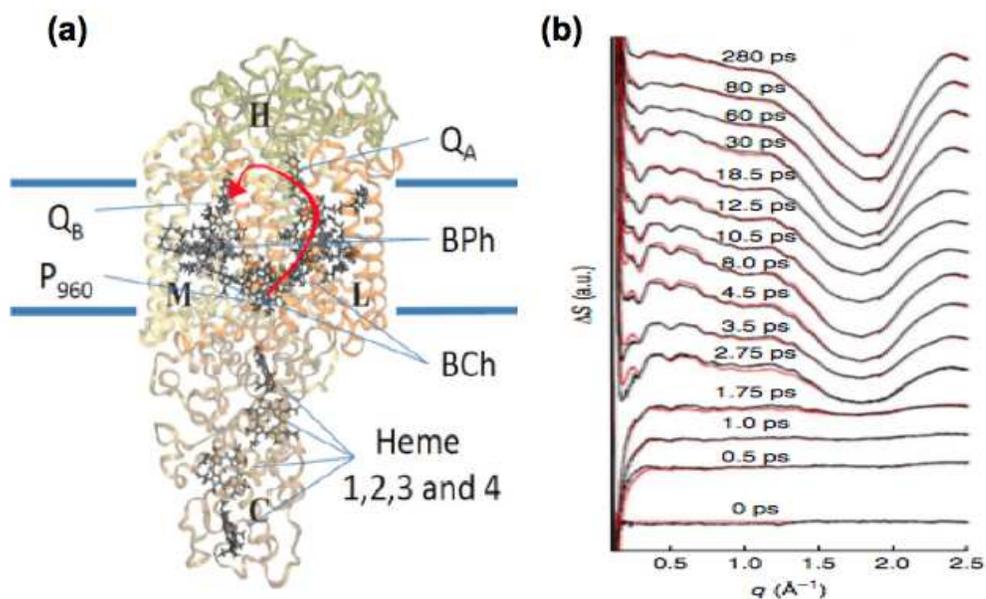


Figure 1.1: (a) The photosynthetic reaction center of RCvir *Blastochloris viridis* showing a complex of four subunits H, L, M C, reaction center cofactors (BPh, BCh, P960, QA, QB, Hemes) and membrane boundaries specified by blue lines, (b) Time-resolved SAXS/WAXS difference data recorded as a function of time delay, obtained from [12]

the help of four components from spectral decomposition: an ultrafast component (C1), a second component identified as protein conformational changes (C2), a non-equilibrated component (C3) and an equilibrated component (C4), they concluded that the structural deformation propagated away from the epicenter of the protein quake before thermal diffusion.

Another important experiment is by Levantino et al.[32] on Myoglobin (Mb) reported the second direct evidence of protein-quake through femtosecond X-ray solution scattering measurements at LCLS X-ray free electron laser. The photo-induced structural rearrangement around the chromophore occurs in the femtosecond time scale accompanied by quake-like intramolecular motion through the protein. The radius of gyration of myoglobin increases within a picosecond timescale and is followed by protein expan-

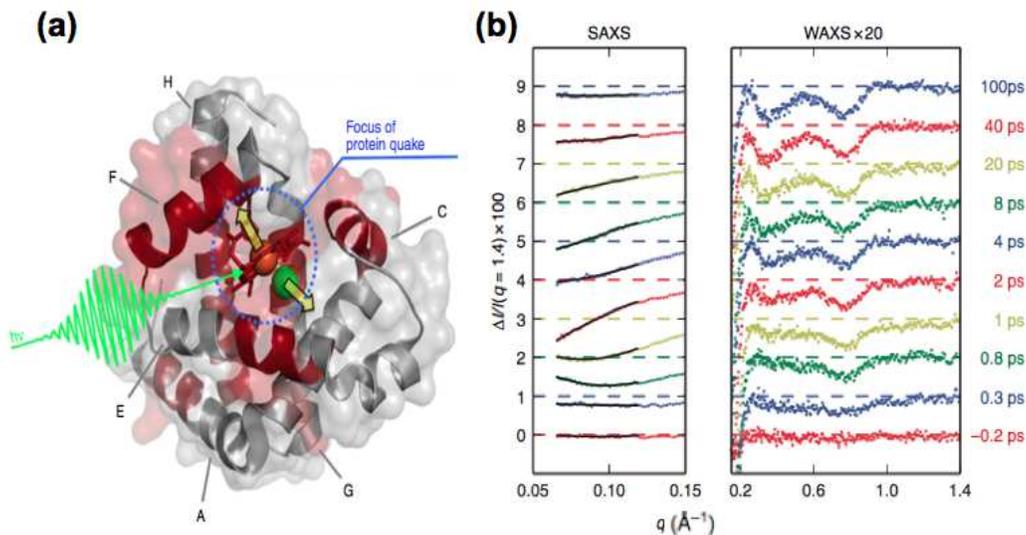


Figure 1.2: (a) Representation of Mb structure showing haem, covalently bounded helix F with haem, parts of helices H,G,E,C and A. The green wave is the representation of light which induces photolysis of the bond between protein and ligand. The yellow arrows showing initiation of structural changes after photolysis. The polypeptide chain segments associated with collective mode of a  $\sim 4$  ps period after ligand release are highlighted by red. (b) Time-resolved SAXS/WAXS difference data recorded as a function of time delay for MbCO in solution, obtained from [32]

sion. The initial localized chemical changes propagate and accompany the global protein confirmation in the picosecond time scale. The initial ultrafast rearrangement of haem molecule excites a series of deformation starting from the aminoacids nearby haem to entire polypeptide chain and solvent hydration layer. The significant global structural rearrangements attributed by the relative motion of helices is revealed by changes in WAXS signal and volume change is revealed by the changes in SAXS signal.

The structural response of Myoglobin (Mb) after photolysis propagates to global protein structural changes is concurrent with the timescale set by the acoustic speed of sound and confirm the proteinquake hypothesis. The ultrafast studies are capable of showing the

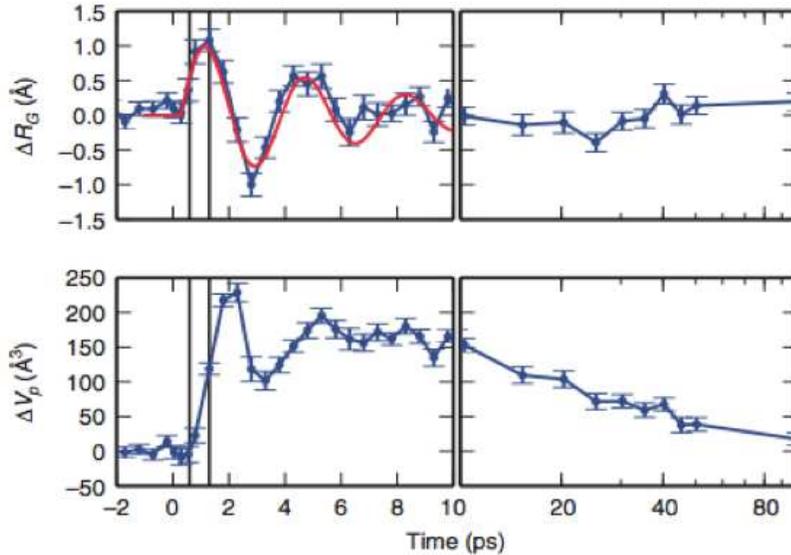


Figure 1.3: Time dependency of Mb structural parameters (change in Radius of gyration and change in volume of protein), obtained from [32]. The red curve showing a damped oscillation with a 3.6 ps time period and 6 ps decay time. The two perpendicular lines within 2 ps time period shows that the increase in protein volume is delayed with respect to  $R_g$  increase.

intrinsic ballistic-like nature of protein motion which is normally not revealed in ensemble measurements at longer timescales. By analysing both SAXS and WAXS regions, it was possible to extract directly the time evolution of relevant structural parameters. The Mb structure oscillates with a period of few picosecond (ps) as shown by time dependence of  $R_g$  (radius of gyration) and  $V_p$  (volume of protein). The work so far demonstrated that the femtosecond time resolution available at XFEL sources and structural sensitivity of X-ray solution scattering techniques can be utilized to know the intrinsic nature of protein elementary motions. Levantino et al. [32] were able to monitor both regions of SAXS and WAXS and extracted the time evolution of relevant structural parameters.

Because of the improved capabilities of X-ray sources, the SAXS/WAXS technique is

extended to monitor transient protein structures with a better time resolution even of the femtosecond range. The solution scattering studies in proteins reveal structural dynamics without the constraint imposed by crystal contacts as well as complement the results obtained from time-resolved diffraction techniques [33, 34, 35]. There are a lot of interesting future applications of XFEL based TR SAXS/WAXS like search for phase-coherent nuclear motions in photosynthetic reaction centers [36], photoactive yellow protein [37], phytochromes [38], photoreceptive complexes and bacterial rhodopsins [39, 40], sensory rhodopsin, visual rhodopsin, heme containing proteins such as hemoglobin [41, 42] and myoglobin [43, 44]. TR SAXS/WAXS can also be applied to trace both reversible as well as irreversible reactions such as protein folding induced by ligand dissociation and electron transfer, enhancing its applicability to various protein reactions such as a drug binding to target proteins [41].

# Chapter 2

## Basic Theory

### 2.1 X-ray Scattering from an Atom

The figure 2.1 shows incident waves, scattered waves with phase difference. The scattered waves having wave vector  $\vec{k}$  and incoming wave vector  $\vec{k}_0$ . With the approximation that observer is very far, the scattered waves are parallel. Scattered waves have a phase difference as they are scattered at different positions. As each scattered waves travel a different length, phase difference arises.

In figure , the second scattered wave travels longer than the first scattered wave, the difference in path length is

$$\text{Path difference} = AO' + O'B. \quad (2.1)$$

The total phase difference becomes  $\exp(2\pi i \vec{r} \cdot \vec{q})$ . The amplitude of the scattered wave is multiplied by a phase factor  $\exp(2\pi i \vec{r} \cdot \vec{q})$ . As the proton is heavier than an electron the nucleus is considered stationary during the scattering process. X-rays are scattered by electrons when they interact with atom. The total number of electrons of the atom

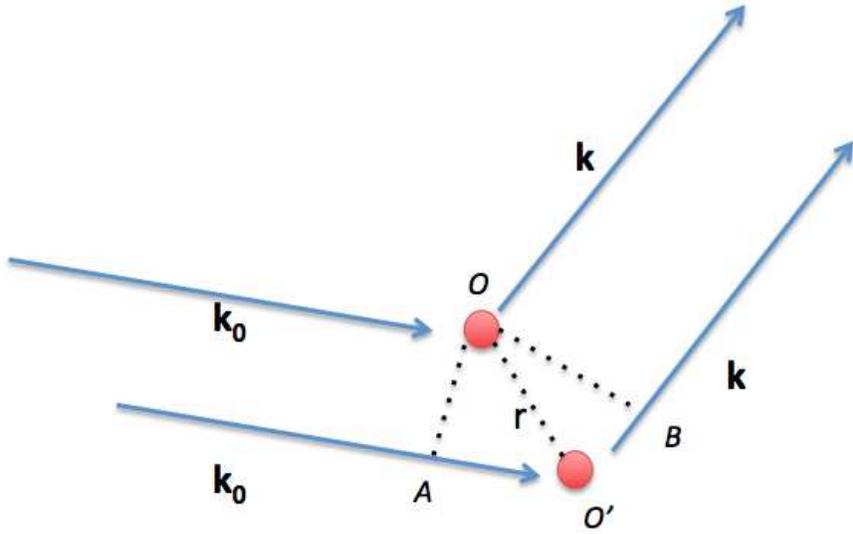


Figure 2.1: X-ray Scattering

influence the strength with which the electrons scatter. The scattering power of an atom is represented by an atomic form factor. The scattering angle also affects the atomic form factor, higher scattering angle decreases the atomic form factor significantly. With many electrons having density  $\rho(\vec{r})$ , at particular  $\vec{q}$  will then gives

$$F(\vec{q}) = \int \rho(\vec{r}) \exp(2\pi i \vec{q} \cdot \vec{r}) d\vec{r}. \quad (2.2)$$

It seems that Fourier transform of the electron density is structure factor. In experiments, the measurable quantity is intensity which is expressed as

$$I(\vec{q}) \propto |F(\vec{q})|^2 \quad (2.3)$$

## 2.2 Scattering from a Molecule

The structure of the molecule is considered as a collection of individual atoms. Although atoms inside the molecule are of same type but they differ in positions only. Being able to know the Fourier transform of a single atom, we could find the total structure factor as the sum over all contribution of the Fourier transform of atoms at all positions. The atomic form factor is the Fourier transform of a single atom. With the work done by Cromer and Mann [45], the Hartree-Fock approximation can be utilized to obtain empirical parameters to approximate atomic form factors. The way they determined the parameters was by fitting 9 parameters in a Gaussian's series to a normalized scattering curves. Those parameters are available from the international Tables of Crystallography [46] and the Gaussian function is shown in equation 2.4.

Atom	$a_1$	$a_2$	$a_3$	$a_4$	$b_1$	$b_2$	$b_3$	$b_4$	$c$
C	2.31	1.02	1.589	0.865	20.84	10.21	0.569	51.65	0.216
N	12.213	3.132	2.013	1.166	0.006	9.893	28.997	0.583	-11.529
O	3.049	2.287	1.546	0.867	13.277	5.701	0.324	32.909	0.251
S	7.070	5.340	2.236	1.512	1.366	19.828	0.092	55.228	-0.159

Table 2.1: Table of Cromer-Mann coefficients

The parameters for different type of atoms are listed in Table 2.1. There are 9 parameters for each atom and above table shows only for four atoms namely carbon, oxygen, nitrogen, and sulfur. The atomic form factor can be calculated using Gaussian function with the knowledge of 9 parameters.

$$f(\sin(\theta)/\lambda) = \sum_{i=1}^4 a_i \exp(-b_i(\sin(\theta)/\lambda)^2) + c \quad (2.4)$$

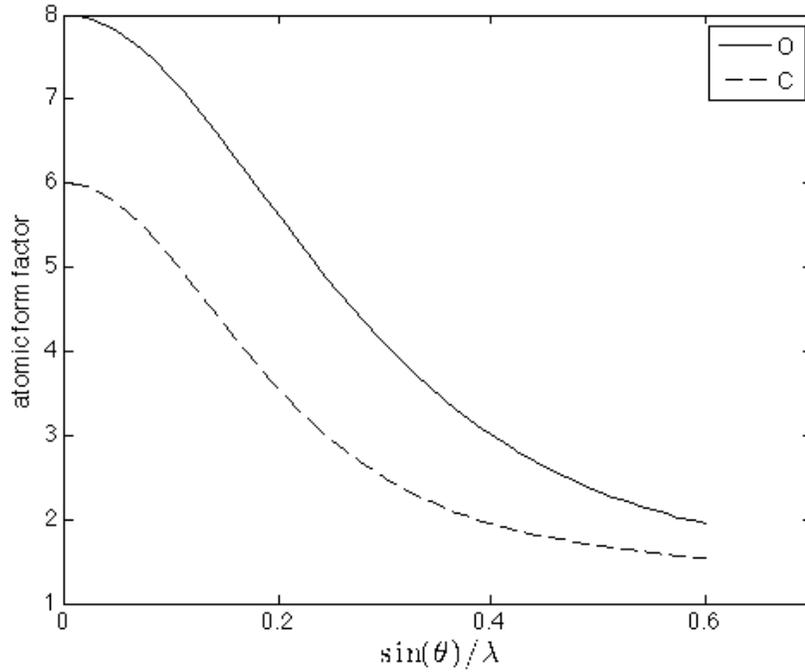


Figure 2.2: Plot of atomic form factor for carbon and oxygen

A plot of the atomic form factor for carbon and oxygen is shown in figure. It is shown in the plot that the value of the atomic form factor goes to their atomic number when  $\sin(\theta)/\lambda$  close to zero.

With the approximation of atomic form factor, the structure factor can be calculated in a simpler way. Once the atomic form factor is calculated, the structure factor can be obtained faster for all atoms. Then the expression for the structure factor in terms of the atomic form factors is described as

$$F^M(\vec{q}) = \sum_{j=1}^N f_j \exp(2\pi\vec{q}\cdot\vec{R}_j) \quad (2.5)$$

with the sum over all  $N$  atoms of molecule at position  $\vec{R}_j$  signifying  $f_j$  as atomic form factor of  $j$ th atom.

To find the structure factor of a molecule above equation is used. As long as a molecule is listed as a collection of atoms in different positions, above equation can be used to simulate the structure factor. Some structures of molecules have been solved using crystallographic methods and their structures are available in the pdb. The pdb file describes a molecule as a list of atom type as well as their positions. Thus, one can simulate a structure factor by using above equation where the entry is from the pdb file.

Fig below is a brief overview of the pdb file. One needs to understand thoroughly the format[47] and the convention of the file in order to read the information from pdb file.

ATOM	1	N	THR	A	12	67.946	33.337	16.826	1.00102.16	N
ATOM	2	CA	THR	A	12	66.915	34.315	17.328	1.00102.16	C
ATOM	3	C	THR	A	12	66.570	35.274	16.204	1.00102.16	C
ATOM	4	O	THR	A	12	66.803	36.487	16.305	1.00102.16	O
ATOM	5	CB	THR	A	12	67.518	35.046	18.502	1.00102.16	C
ATOM	6	OG1	THR	A	12	66.587	35.986	19.020	1.00102.16	O
ATOM	7	CG2	THR	A	12	68.790	35.815	18.139	1.00102.16	C
ATOM	8	N	MET	A	13	66.002	34.684	15.140	1.00 89.79	N
ATOM	9	CA	MET	A	13	65.641	35.451	13.882	1.00 89.79	C
ATOM	10	C	MET	A	13	64.509	36.438	14.049	1.00 89.79	C
ATOM	11	O	MET	A	13	64.222	37.242	13.151	1.00 89.79	O
ATOM	12	CB	MET	A	13	65.134	34.499	12.528	1.00102.16	C
ATOM	13	CG	MET	A	13	64.734	35.183	11.118	1.00102.16	C
ATOM	14	SD	MET	A	13	64.160	34.185	9.729	1.00102.16	S
ATOM	15	CE	MET	A	13	63.807	35.229	8.332	1.00102.16	C
ATOM	16	N	ARG	A	14	63.810	36.459	15.148	1.00 50.58	N
ATOM	17	CA	ARG	A	14	62.716	37.444	15.257	1.00 50.58	C
ATOM	18	C	ARG	A	14	63.265	38.810	15.654	1.00 50.58	C
ATOM	19	O	ARG	A	14	62.901	39.839	15.065	1.00 50.58	O
ATOM	20	CB	ARG	A	14	61.689	37.045	16.311	1.00102.16	C

Figure 2.3: Example of data from protein data bank in pdb format

First, the pdb file has row entries where each row is a single atom in particular position together with additional information. It consists of multiple columns where each column has particular information. There are 27 columns and all data is in a text file in ASCII format. For getting structure factors, only atom types and their positions are needed. There are four pieces of information needed, namely atom type, position-x, position-y, and position-z. The position-x is shown between columns 31-38, position-y is shown between columns 39-46. The position-z is shown between 47-54. The structure factor can be

simulated using equation above with the information provided in pdb file. The detailed explanation of the pdb file is given in the Appendix A.

## 2.3 Debye Formula for Scattering Profiles

Debye formula [3] is usually utilized for model calculations. Particles can be thought of as discrete sum of different scatterers i.e. atoms. The spherically averaged intensity can be expressed in terms of sum of form factors of atoms.

$$I(q) = \sum_{j=1}^N \sum_{k=1}^N f_j(q) f_k(q) \frac{\sin(qr_{jk})}{qr_{jk}} \quad (2.6)$$

where  $r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|$

The derivation of above equation is as follows. Proteins are composed of aminoacids which in turn are formed from atoms differing by side chain arrangement. For the intensity, averaged over all orientations,

$$I(q) = \left\langle \sum_{j=1}^N \sum_{k=1}^N f_j f_k^* \exp(iq(r_k - r_j)) \right\rangle \quad (2.7)$$

If the diffracted intensity is contributed only by atoms, the phase factor reduces to one. The relative distance  $r_j - r_k$  contribute the interference between the atoms.

Splitting the sum gives,

$$I(q) = \sum_{j=1}^N I_j(q) + 2 \langle f_j(q) f_k(q) \exp(i(qr_{jk} + \phi_k - \phi_j)) \rangle \quad (2.8)$$

As molecules are in different orientations (no fixed origin), the phase, which is depen-

dent on origin vanishes after averaging over all orientations. The structure factor thus independent of orientation. Moreover, the spherical averaging of all orientations is

$$\langle \exp(iqr_{jk}) \rangle = \frac{\sin(qr_{jk})}{qr_{jk}} \quad (2.9)$$

Debye factor is the representation of spherical averaging. The Debye formula is then

$$I(q) = \sum_{j=1}^N I_j(q) + 2 \sum_{j \neq k}^n f_j(q) f_k(q) \frac{\sin(qr_{jk})}{qr_{jk}} \quad (2.10)$$

In a combined form (with interference between atoms and atomic diffraction), the equation can be written as 2.6 .

The Debye formula is used for getting an intensity profile from atomic coordinates. Since the scattering profile need to be computed from every interacted pairs of atoms resulting high computational cost  $O(N^2)$ , different other models have been developed to compute the intensity instead.

## 2.4 Spherical Harmonics

One of the approaches used to reduce the computational cost of SAXS profile reconstruction is Spherical Harmonics. The scattering body or protein is expressed in terms of series of spherical harmonics. This approach reduces the computational cost effectively from  $O(N^2)$  to  $O(N)$  to obtain scattering intensity by exploiting the orthogonality properties of the basis functions which helps the averaging of the harmonic series.

As the SAXS intensity is angular averaging of reciprocal space intensities,

$$I(q) = \int I(\mathbf{q}) d\hat{\mathbf{q}} \quad (2.11)$$

The intensity is related to structure factor as

$$I(q) = \int |F(\mathbf{q})|^2 d\hat{\mathbf{q}} \quad (2.12)$$

The structure factor can be expressed in spherical harmonic basis, then

$$I(q) = \int \sum_{lm} \sum_{lm} |F_{lm}(q)|^2 |Y_{lm}(\hat{\mathbf{q}})|^2 d\hat{\mathbf{q}} \quad (2.13)$$

Using the orthogonality properties of spherical harmonics,

$$I(q) = 4\pi |Y_{00}|^2 \sum_{lm} |F_{lm}(q)|^2 \quad (2.14)$$

Finally, the SAXS intensity can be written as

$$I(q) = \sum_{lm} |F_{lm}(q)|^2 \quad (2.15)$$

There might have been several studies to obtain difference density from SAXS profiles. The expanded details of this study for getting difference density from difference in SAXS/WAXS intensity are elaborated in Chapter 3.

Spherical harmonics are a set of functions that satisfies such a criterion as well as special functions defined on the surface of sphere. Spherical harmonics is defined in spherical coordinates represented by angles  $\theta$  and  $\phi$ . Two quantum numbers namely  $l$  and  $m$  are characterized by spherical harmonics. The quantum number  $m$  signifies how the function varies with respect to azimuthal angle.

The definition of spherical harmonics is given by

$$Y_{lm}(\theta, \phi) = (-1)^m \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{lm}(\cos \theta) e^{im\phi} \quad (2.16)$$

where the  $P_{lm}(\cos \theta)$  are Legendre polynomials.  $P_{lm}(x)$  can be obtained using Rodrigues formula:

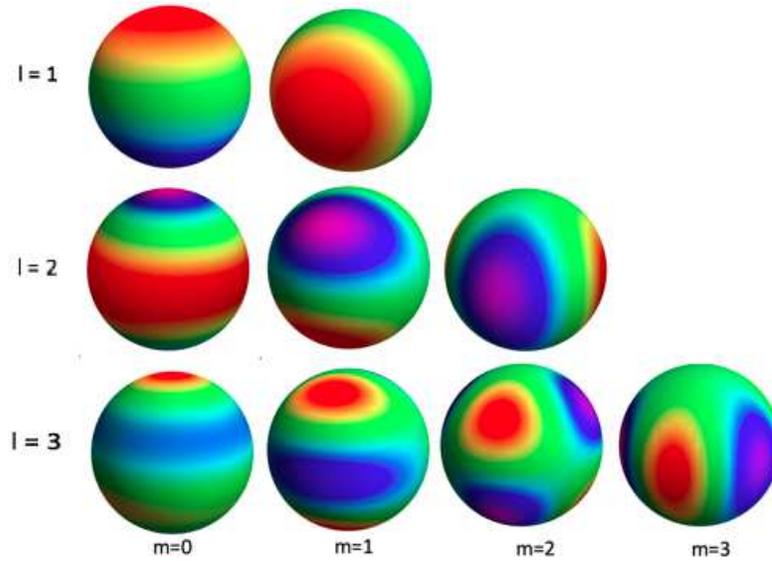


Figure 2.4: Spherical harmonics plot characterized with different quantum numbers

$$P_{lm}(x) = \frac{(-1)^m}{2^l l!} (1-x^2)^{m/2} \frac{d^{l+m}}{dx^{l+m}} [(x^2-1)^l]. \quad (2.17)$$

Spherical harmonics are a set of functions characterized by two quantum numbers. Every single spherical harmonic function with different quantum numbers is orthogonal to each other.

$$\int Y_{lm} Y_{l'm'}^* d\Omega = \delta_{ll'} \delta_{mm'}. \quad (2.18)$$

where  $\delta_{ll'}$ ,  $\delta_{mm'}$  is ia Kronecker delta functions.

Also,

$$Y_{lm}^*(\theta, \phi) = (-1)^m Y_{l,-m}(\theta, \phi) \quad (2.19)$$

## 2.5 Guiniers Law in SAXS

The scattering profile of a particle can be approximated by a Gaussian curve in the neighborhood of the origin.

The structure factor integral is

$$F(\mathbf{q}) = \int \rho(\mathbf{r}) \exp(i\mathbf{q}\cdot\mathbf{r}) d\mathbf{r} \quad (2.20)$$

As  $q \rightarrow 0$ ,

$$F(\mathbf{q}) = \int \rho(\mathbf{r}) \left[ 1 + i\mathbf{q}\cdot\mathbf{r} - \frac{1}{2}(\mathbf{q}\cdot\mathbf{r})^2 \right] d\mathbf{r} \quad (2.21)$$

After integrating,

$$F(q) = N_e \left[ 1 + 0 - \frac{1}{2}q^2 \left\{ \frac{1}{3}q^2 R_g^2 \right\} \right] \quad (2.22)$$

$$F(q) = N_e \exp \left[ -\frac{1}{2} \left\{ \frac{1}{3} q^2 R_g^2 \right\} \right] \quad (2.23)$$

The intensity is then,

$$I(q) = N_e^2 \exp \left[ -\frac{1}{3} q^2 R_g^2 \right] \quad (2.24)$$

$$\ln(I(q)) = 2 \ln N_e - \left( \frac{1}{3} q^2 R_g^2 \right) \quad (2.25)$$

With a plot of  $\ln I(q)$  vs  $q^2$ , Guinier's law helps to determine number of electrons  $N_e$  and radius of gyration  $R_g$ .

Guinier's law can simply be interpreted as following from Debye formula in equation (2.6) .

$$I(q) = \sum_{j=1}^N \sum_{k=1}^N f_j(q) f_k(q) \frac{\sin(qr_{jk})}{qr_{jk}} \quad (2.26)$$

where  $r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|$

In the neighbourhood of  $q \rightarrow 0$ , the use of McLaurin expansion of  $\frac{\sin(qr)}{qr} \approx 1 - \frac{(qr)^2}{3!} + \dots$  into the above equation gives,

$$I(q) = I(0) \left[ 1 - \frac{R_g^2 q^2}{3} + O(q^4) \right] \quad (2.27)$$

$$I(q) \approx I(0) \left[ \exp\left(-\frac{q^2 R_g^2}{3}\right) \right] \quad (2.28)$$

which is the formula used by Guinier in his first SAXS studies on metallic alloys.

## 2.6 Porod's Law in SAXS

The scattering factor for a uniform sphere of radius  $R$  is

$$F(q) = \frac{4\pi R^3}{3} \left\{ \frac{3\sin(qR) - qR\cos(qR)}{(qR)^3} \right\} \quad (2.29)$$

For  $q \rightarrow \infty$

$$F(q) = \frac{4\pi R^3}{3} \left\{ \frac{-\cos(qR)}{(qR)^2} \right\} \quad (2.30)$$

Again,  $I(q)$  at  $q \rightarrow \infty$

$$I(q) = \frac{8\pi^2 R^6}{9 (qR)^4} \quad (2.31)$$

$$I(q) = \frac{2\pi}{9} \frac{A}{q^4} \quad (2.32)$$

$A$  is the surface area of the sphere.

Taking natural log on both sides,

$$\ln I(q) = \ln\left(\frac{2\pi}{9}\right) + \ln A - 4 \ln(q) \quad (2.33)$$

At high  $q$ , ln-ln plot of  $I(q)$  vs  $q$  has a gradient -4 for a sphere and its intercept on the  $\ln(I(q))$  axis gives information about the surface area of the particle. Other gradients for particles of other shapes.

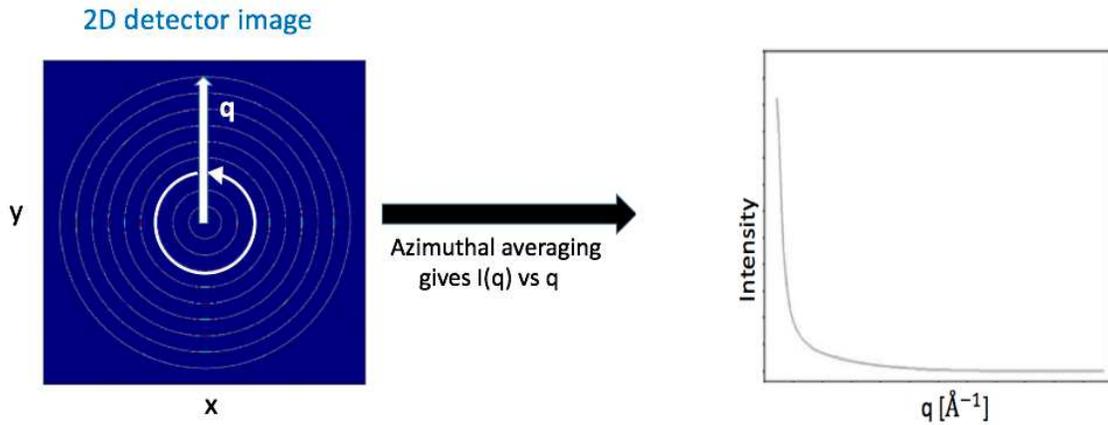


Figure 2.5: Figure showing experimental scattered pattern along with SAXS curve

## 2.7 SAXS Intensity from Scattered Patterns

The 2D scattered images are azimuthally averaged to give one dimensional SAXS/WAXS intensity as a function of scattering length ( $q$ ). Several softwares i.e. pyFAI, ConeX, FIT2D are available nowadays for this purpose.

Azimuthal average is widely used mathematical technique when 2D area detectors are employed for recording SAXS scattering patterns. The technique enables wider solid angle coverage, thus a better collection of X-ray photons.

Input data are  $(x, y)$  and Output data are  $(r, \phi)$  or  $(q, \phi)$  or  $(2\theta, \phi)$

$$r = \sqrt{x^2 + y^2} \quad (2.34)$$

$$\phi = \tan^{-1}(y/x) \quad (2.35)$$

$$q = 2\text{Sin}\theta/\lambda \quad (2.36)$$

$$2\theta = \tan^{-1}(r/d) \quad (2.37)$$

with  $2\theta$  is the scattering angle and  $d$  is the detector distance from sample.

## 2.8 Spherical Bessel Function

The radial function used in spherical geometries satisfy the following differential equation

$$r^2 \frac{d^2 R}{dr^2} + 2r \frac{dR}{dr} + [(kr)^2 - n(n+1)] R = 0 \quad (2.38)$$

Two linearly independent solutions to above equation are the spherical Bessel function  $j_n(kr)$  and spherical Neumann function  $y_n(kr)$ . For problems involving spherical case, only the spherical Bessel function of the first kind is required. The spherical Bessel function is relation to ordinary Bessel function  $J_n$  by

$$j_n(kr) = \sqrt{\frac{\pi}{2kr}} J_{n+1/2}(kr) \quad (2.39)$$

The first three spherical Bessel functions of the first kind are

$$j_0(kr) = \frac{\text{sin}kr}{kr} \quad (2.40)$$

$$j_1(kr) = \frac{\text{sin}kr}{(kr)^2} - \frac{\text{cos}kr}{kr} \quad (2.41)$$

$$j_2(kr) = \frac{3\text{sin}kr}{(kr)^3} - \frac{3\text{cos}kr}{(kr)^2} - \frac{\text{sin}kr}{kr} \quad (2.42)$$

The higher order forms can be generated from the first form as

$$j_l(kr) = \left(\frac{-r}{k}\right)^l \left(\frac{1}{r} \frac{d}{dr}\right)^l j_0(kr) \quad (2.43)$$

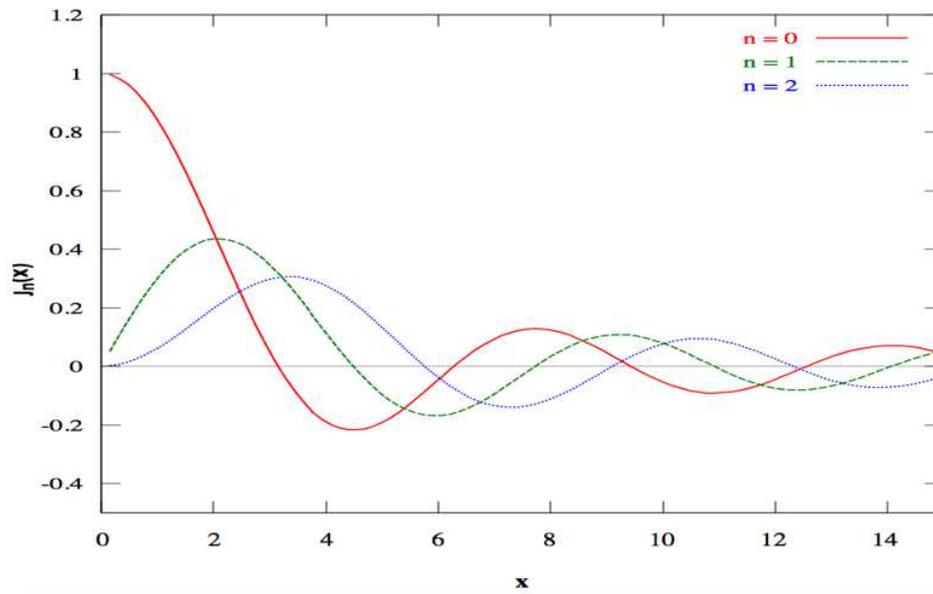


Figure 2.6: Plot of first three spherical Bessel function of first kind.

# Chapter 3

## Theory for Structural Changes

### 3.1 Rapid direct method for extraction of difference electron density

The starting point in this theory is the relation between a SAXS/WAXS spectrum and a set of structure factors of the molecule that is familiar in crystallography. In general a structure factor,  $F_{u,p}(\mathbf{q})$ , of an unperturbed( $u$ ) and a perturbed( $p$ ) structure respectively may be written in terms of their respective electron densities  $\rho_{u,p}(\mathbf{r})$ . By treating the problem as one of finding a perturbed structure from a knowledge of an unperturbed structure one encompasses proposed mix-and-inject experiments [50] and those based on photoexcitation [13, 44]. Given is the general structure factor integral,

$$F_{u,p}(\mathbf{q}) = \int \rho_{u,p}(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} \quad (3.1)$$

An alternative approximation can be calculated from the discrete atomic coordinate data taken from a corresponding crystal structure deposition in the Protein Data Bank

[51] is

$$F_{u,p}(\mathbf{q}) = \sum_n f_n \exp(i\mathbf{q} \cdot \mathbf{r}_n) \quad (3.2)$$

where the summation is over all atoms. The SAXS/WAXS intensity must be integrated over all angles, and is therefore

$$I_{u,p}(q) = \int |F_{u,p}(\mathbf{q})|^2 d\hat{\mathbf{q}} = \int F_{u,p}^*(\mathbf{q}) F_{u,p}(\mathbf{q}) d\hat{\mathbf{q}} \quad (3.3)$$

where the scalar  $q$  is now reduced to a radius.

Consequently, measured quantities in a SAXS/WAXS experiment are not only related to the square moduli of the structure factors, as in X-ray crystallography, they are also related to an angular integral of the square moduli. When considering the relative paucity of data from the experiment, it is not surprising that extraction of electron density directly from the SAXS/WAXS data is impossible. Nevertheless, it is possible to use analogies between the crystallographic difference Fourier method and differences between SAXS curves as shown here to objectively produce the locations of small differences in structure. The quantities one wants to relate are the difference SAXS/WAXS intensities and the difference electron densities. There exists a relationship in this limit of small differences.

First take the variation of (3.3). This gives the equation

$$\delta I_u(q) = \int \{\delta F_u^*(\mathbf{q}) F_u(\mathbf{q}) + F_u^*(\mathbf{q}) \delta F_u(\mathbf{q})\} d\hat{\mathbf{q}} \quad (3.4)$$

The crucial steps are now identifying  $\delta I_u(q)$  with the experimentally determined difference SAXS/WAXS spectrum

$$I_p(q) - I_u(q) \quad (3.5)$$

and a relation of  $\delta F_u(\mathbf{q})$  to a difference electron density. The latter may be found by just taking the variation of (3.1). This gives

$$\delta F_u(\mathbf{q}) = \int \delta \rho_u(\mathbf{r}) \exp(i\mathbf{q} \cdot \mathbf{r}) d\mathbf{r} \quad (3.6)$$

Due to Taylor's expansion for small enough changes it is an exact result that the changes in the SAXS/WAXS curves (accessible to experiment) are linearly related to the change in electron density that is sought in traditional time-resolved spectroscopy as well as here in solution. This theory presents that this linear relationship holds even under experimental conditions in a SAXS/WAXS experiment (as has been found already in low energy electron diffraction (LEED))[52]. This allows us to use a matrix inversion method to solve the inverse problem.

The sum over  $n$  here is over all atoms in the structure. In more details, the angular integral in (3.4) is best done using the orthogonality relation

$$\int Y_{lm}^*(\hat{\mathbf{q}}) Y_{l'm'}(\hat{\mathbf{q}}) d\hat{\mathbf{q}} = \delta_{ll'} \delta_{mm'} \quad (3.7)$$

in spherical coordinates. This is easiest done by substituting the spherical expansion of a plane wave, namely

$$\exp(i\mathbf{q} \cdot \mathbf{r}_n) = 4\pi \sum_{lm} i^l j_l(qr_n) Y_{lm}(\hat{\mathbf{q}}) Y_{lm}^*(\hat{\mathbf{r}}_n) \quad (3.8)$$

into (3.1).

Then one can use the expansion of the structure factors in terms of spherical harmonics [53],

$$F_{u,p}(\mathbf{q}) = \sum_{lm} F_{lm}^{(u,p)}(q) Y_{lm}(\hat{\mathbf{q}}) \quad (3.9)$$

and find from (3.2)

$$F_{lm}^{(u,p)}(q) = 4\pi \sum_n i^l f_n j_l(qr_n) Y_{lm}^*(\hat{\mathbf{r}}_n) \quad (3.10)$$

In the last expression the sum over  $n$  is a sum over atomic coordinates, and this is actually a very convenient way to calculate the SAXS/WAXS spectra *via*

$$I_{u,p}(q) = \sum_{lm} |F_{lm}^{(u,p)}(q)|^2 \quad (3.11)$$

with the angular average done analytically. Note that both  $I_u(q)$  and  $I_p(q)$  are calculated only for the purposes of this study. In practice,  $I_p(q)$  and  $I_u(q)$  are found from the experiment. The SAXS/WAXS intensities are produced in their spherical representation, where the angular average has already been performed.

In a time-resolved experiment one measures multiple diffraction patterns in quick succession. A particular structure is assumed not to change by much in such a short time interval, yet is hoped that a comparison of the diffraction patterns will reveal the time variation of the structure. In time-resolved crystallography [16, 54] this is provided by the difference Fourier method [55]. It is investigated here whether similar information can be obtained from a time-resolved SAXS/WAXS experiment. For the test purposes, the SAXS intensities  $I_u(q)$  and  $I_p(q)$  are both calculated from Eqs. (3.2), (3.10) and (3.11). Representing the intensities and structure factors in spherical coordinates allows the necessity of angular averages for SAXS/WAXS to be done. Substituting (3.2) and (3.6) into (3.4) one may deduce the relationship between the difference SAXS/WAXS spectrum and the difference electron density sought, namely:

$$\delta I_u(q) = \sum_k M_{q,k} \delta \rho_u(\mathbf{r}_k) \quad (3.12)$$

where  $\mathbf{r}_k$  are taken to be a uniform set of Cartesian grid points at which the elec-

tron density of the molecule is required. The definition of small is unclear but may be best investigated by the kind of modeling done here. Simulations performed here suggest that the approximation is valid for the structural changes in PYP that accompany photoexcitation. The elements of the matrix in (3.12) are of the form:

$$M_{q,\mathbf{k}} = \int \sum_k \{F_u(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{r}_k) + c.c.\} d\hat{\mathbf{q}} \quad (3.13)$$

The angular integral is best done by using (3.7). One then obtains

$$M_{q,\mathbf{k}} = 4\pi \sum_{lm} j_l(qr_k) \{i^{-l} F_{lm}^{(u)}(q) Y_{lm}(\hat{\mathbf{r}}_k) + c.c.\} \quad (3.14)$$

As for the spherical harmonic expansion coefficients  $F_{lm}$  of the structure factor  $F$ , they may be calculated from the data in the PDB file by the expression (3.10) above. In the expressions for  $F_u$  (and  $F_p$ ), the sums are over the atoms in the relevant PDB files, and  $j_l$  is a spherical Bessel function of order  $l$ .

Obtaining the pseudo-inverse of Eq.(3.14) one obtains the relationship one seeks, namely

$$\delta\rho_u(\mathbf{r}_k) = M_{\mathbf{k},q}^{-1} \delta I_u(q) \quad (3.15)$$

which calculates an approximate difference density  $\delta\rho_u(\mathbf{r}_k)$  at positions  $\mathbf{r}_k$  by premultiplication of the difference SAXS spectrum  $\delta I_u(q)$  by the elements  $M_{\mathbf{k},q}^{-1}$  of the inverse of the matrix  $M$ .

## 3.2 Matrix Inversion with SVD (Singular Value Decomposition)

The Singular Value Decomposition of a real matrix  $M$  (where  $M$  is  $m \times n$ ) is the product of three decomposed matrices.

$M = USV^t$  where  $U$  is an  $m \times m$  matrix,  $S$  is an  $m \times n$  matrix, and  $V$  is an  $n \times n$  matrix.  $V$  and  $U$  have orthogonal property that

$$V^tV = I \quad (3.16)$$

$$U^tU = I \quad (3.17)$$

and  $S$  has diagonal elements only. The elements are in decreasing order and non-negative such that

$$s_1 \geq s_2 \geq s_3 \dots \geq s_n \geq 0 \quad (3.18)$$

The condition number of  $M$  is the ratio

$$\frac{s_1}{s_n} \quad (3.19)$$

The truncated SVD is used to compute matrix inversion to solve for  $\delta\rho$ . Then the system of equations becomes

$$\delta I_u = M\delta\rho_u \quad (3.20)$$

which is represented as

$$\delta\rho_u = \sum_{i=1}^{n' \leq n} \frac{u_i^\dagger \delta I_u}{s_i} v_i \quad (3.21)$$

The columns matrices of V and U are respectively represented as  $v_i$  and  $u_i$ . The smaller singular values do not decay as fast as the initial bigger singular values but tend to level off and the solution is dominated by the terms in the sum corresponding to the smallest  $s_i$  [56]. Therefore, the truncation is necessary.

### 3.3 Simulated Annealing

The inspiration for name ‘‘Simulated Annealing’’ comes from annealing in metallurgy. Simulated annealing is a probabilistic method for determining the global minimum of a cost function out of several minima’s. The probabilistic approximation algorithm could produce good solution for an optimization problem in large search space.

The neighbouring states and current states are iterated and simulated annealing probabilistically decides between moving the system to neighbouring state or current states. The iteration step is repeated until the system reaches a state that is good enough for the application ( in terms of energy it is minimum energy). The probability function is usually selected in such a way that the probability of acceptance decreases when the difference between terms of the cost function increases i.e. small uphill moves are more likely than large ones.

For energy related problem, the transition probability from one state to another state depends on the energies of the two states and on parameter temperature T. The temperature T contributes important role in the evolution of the state of the system corresponding

to its sensitivity to the variations of system energies. The algorithm starts initially with higher  $T$  and then with the proper annealing schedule ending at  $T = 0$ , the system is expected to search towards a large space containing good solutions with lower energy regions and finally to downhill.

One must specify the energy function, search space, probability function, along with initial and annealing temperature. Considering  $E'$  and  $E''$  as the energy of state  $n'$  and  $n''$ , the acceptance probability  $P(E',E'')$  is defined as 1 if  $E'' < E'$ , and  $\exp(\frac{E'-E''}{T})$  in the formulation of Kirkpatrick et al. [20]. The probability that the annealing algorithm ends with global optimal solution approaches unity as the annealing process is extended [57].

In this study, once one obtains a resolution to the level of identifying the residue in which atoms have changed their positions with the rapid direct method above, since the average residue in a protein has only a relatively small number of atoms, it is expected that it will not be too difficult to vary only the positions of these atoms with a global optimization algorithm such as simulated annealing [20] to get a reasonably accurate solution [21]. In other words the first step is performed only to narrow down the simulated annealing search. This is essential with a method such as simulated annealing which scales in a disadvantageous way as the number of atoms,  $N$ , whose positions are sought (In fact Rous [58] finds evidence that the time for a calculation may vary as  $N^6$ ).

The method of simulated annealing has now become established as a method of global optimization in which one varies the values of certain parameters on which a spectrum depends, according to a general protocol until one obtains the parameters that give best agreement with an experimental spectrum by finding the global minimum of a cost function. Though the global minimum of the cost function may involve temporarily increasing its value as a function of the parameters, this usually works, due to the employment of a

strategy based on the analogy with the physical process of annealing. Normally simulated annealing would not be feasible for problems of this kind, because varying the positions of all the atoms of the sample would not be feasible. Like the latter process it is a method generally capable of overcoming temporary barriers in order to find a global minimum.

The main problem with simulated annealing is that the time taken by the process scales as a high power of the number of parameters,  $N$ . If  $N$  can be kept small, the method becomes feasible. The strategy for keeping  $N$  small is finding the residue which is displaced in the rapid direct method described in the first section of this chapter. Cost function for this purpose i.e.  $[I_{expt} - I_{the}(R_j)]^2$  is minimized by varying the coordinates of the atoms  $R_j$ .  $I_{the}$ , the model intensities, thus vary as a function of the atom coordinates  $R_j$ . Since each atom can normally be described by three coordinates, the number of atoms can be related to the number of parameters by a factor of a maximum of 3 (the number of physical dimensions) although application of chemical constraints will reduce this considerably. If we assume the moved atoms lie within a residue containing these peaks of the “direct” method, the number  $N$  of parameters  $R_j$  can be kept fairly small, and the solution is feasible.

The PYP molecule used as a test case in this study for instance contains about 1100 atoms, but a single residue contains about 10. Thus, assuming the time goes up with  $N^6$  as Rous [58] indicates, this is automatically at least a trillion-fold reduction in the time taken over time to solve the entire molecule. The calculation currently runs few hours on an ordinary laptop. If Rous is right, trying to find the positions of all the atoms in the protein by simulated annealing on an ordinary laptop would take about many many years! Consequently, simulated annealing in this study is not a method for finding a structure from scratch. Rather if an approximate structure can be found by some means, it is excellent as a method of refinement, hence its use in this case in conjunction with a

“direct” method of limited resolution.

# Chapter 4

## Results

### 4.1 Photoactive Yellow Protein(PYP)

In order to investigate the functional behaviour of the biomolecules, a well-understood model system is required. PYP suits on that category which evolves with several intermediates at surrounding temperatures. PYP is relatively small size protein, can be produced easily in more amounts and relatively stable ensures this protein an excellent model system for structure based dynamics investigations. A lot of data collection techniques and computer algorithms are employed for this model system.

PYP was discovered by Meyer [59] in 1985 in cell lysates of a halophilic purple bacterium called *Ectothiorhodospira halophila*, it is a water-soluble protein with a molecular weight of 14kDa. The para-coumaric acid chromophore(pCA) and its nearby residues play a crucial role in the photocycle. The chromophore is covalently attached to Cys69 through a thioester bond and forms hydrogen bonds with amino acid residues Tyr42, Glu46 and with backbone amide of Cys69[60, 61].

PYP enters into a fully reversible photocycle with several intermediates ranging from

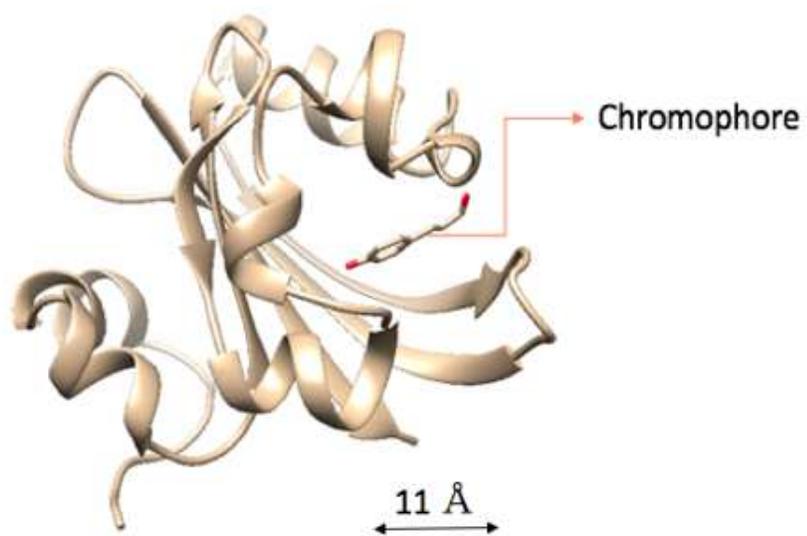


Figure 4.1: Unexcited PYP structure with chromophore

femtoseconds [64, 65] to seconds upon absorption of a blue light[62, 63]. The photocycle has been studied extensively with various time-resolved techniques. The external parameters like pH, X-ray dose, temperature also affects the kinetics of PYP.

The chromophore is an aromatic system in which mesomerically coupled double bonds exist. The valence band electrons are delocalized and can freely move within the bonding network. When a chromophore molecule is illuminated with blue light of certain wavelength, trans/cis configurational changes occur resulting conformational changes of the protein.

More than two decades of structure determination on the PYP photocycle enabled for the first time the determination of a complete structural view of the chemically so important trans to cis isomerization in a biomolecule, starting with fundamental motions on the femtosecond time scale and ending with barriers of activation of the slowest processes in the kinetic mechanism. With the hope of attosecond X-ray sources, more knowledge are expected to be gained. PYP is a versatile molecule capable of covering a wide range of applications and contributing more understanding of biology, chemistry and physics in the case of photoreaction, orbital theory and atomic physics on the attosecond scales [66].

## 4.2 Output of the rapid direct method

The output of the rapid direct method is shown in figures later taking Photoactive Yellow Protein(PYP) as a test case.

The difference electron density maps are produced with Chimera [69]. The tan structure represents the “dark”,known reference state and the atoms moved to generate “ex-

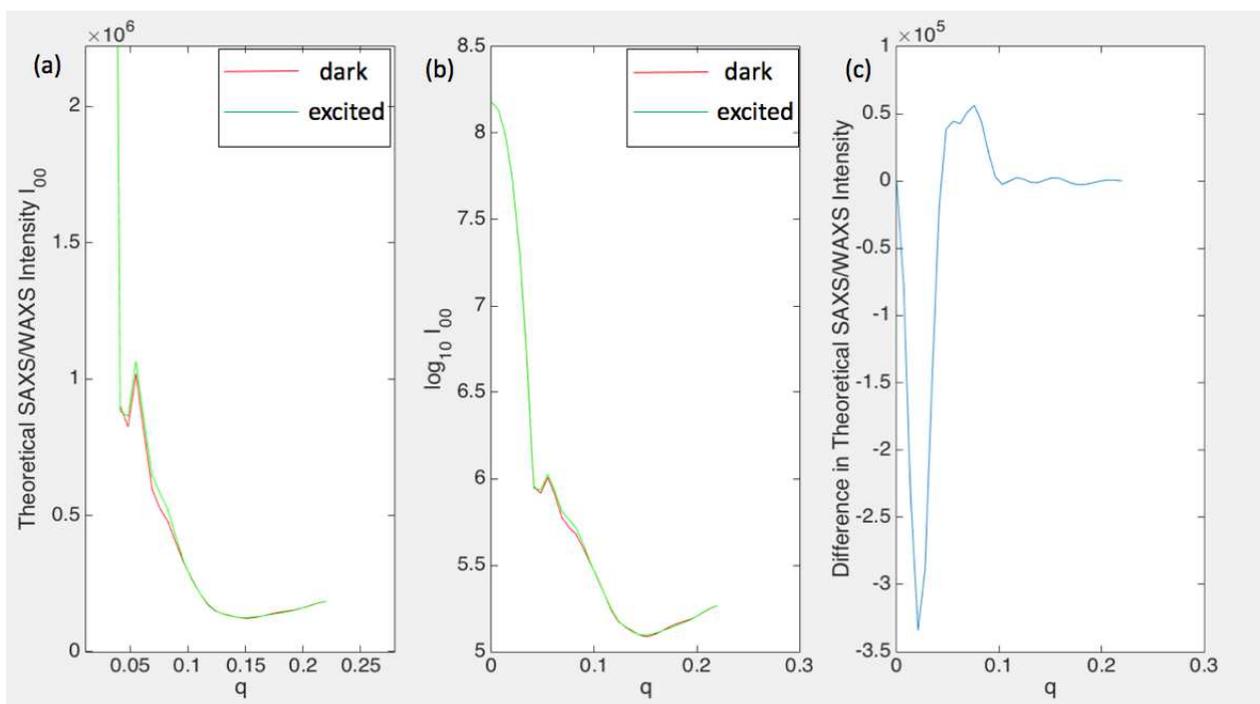


Figure 4.2: (a) Theoretical(Simulated) SAXS/WAXS Intensities for unexcited and excited PYP structure with chromophore swing (b)Logarithmic plot of theoretical(simulated) SAXS/WAXS Intensities for unexcited and excited PYP structure with chromophore swing (c) Difference in theoretical SAXS/WAXS Intensity for chromophore swing case.  $q$  values are in crystallographic covention with unit  $\text{\AA}^{-1}$ .

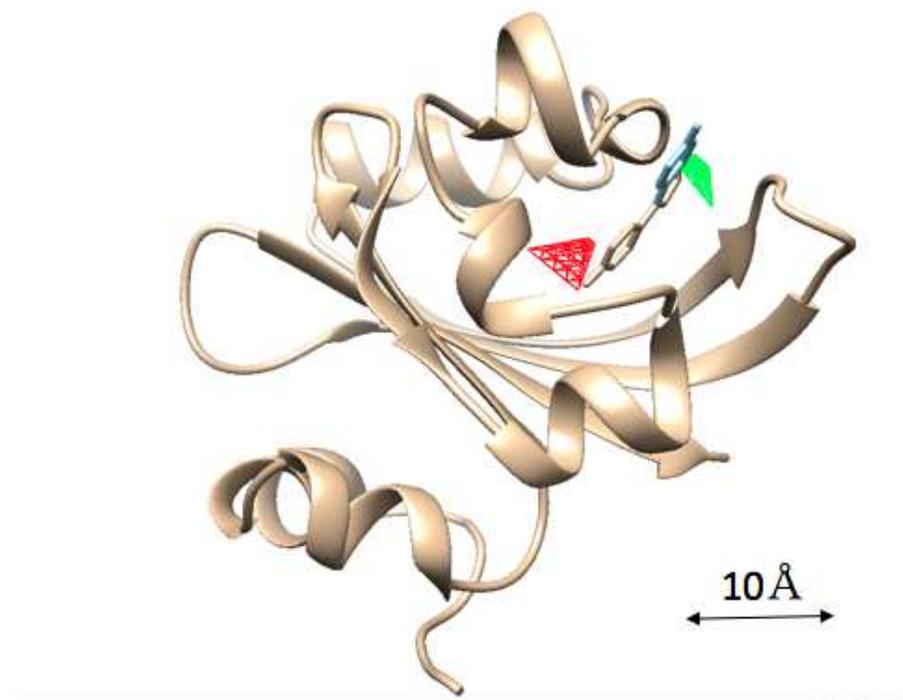


Figure 4.3: Electron density differences between the photoexcited PYP [67] assumed and the “dark” structure of PYP(PDB ID 2phy)[68] from step1 i.e. rapid direct method of algorithm. In this step only the highest and lowest values of the density differences is used so the contour level is chosen so as to show only these. Consistent with the convention in time-resolved crystallography, the positive difference electron density (towards which the chromophore moves) is indicated by the green lobe and the negative electron density (from which the chromophore has moved) is indicated by red.

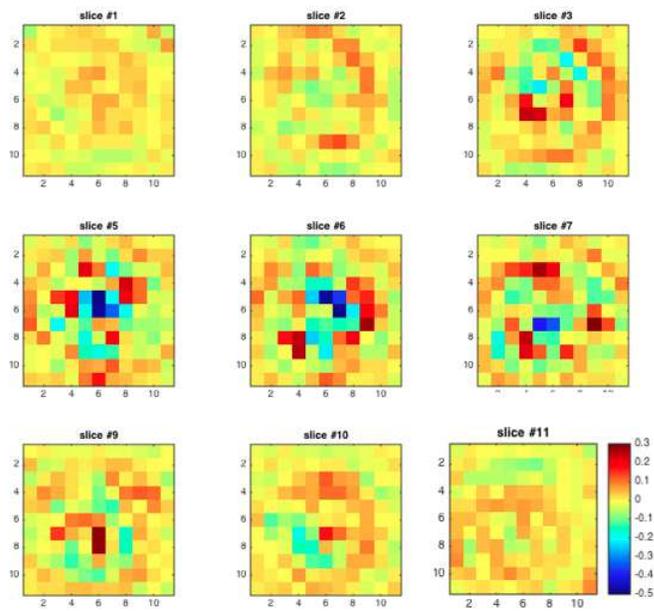


Figure 4.4: Slices through the difference in density(top to bottom) obtained from matrix inversion during chromophore swing, red as positive difference density and blue as negative difference density.

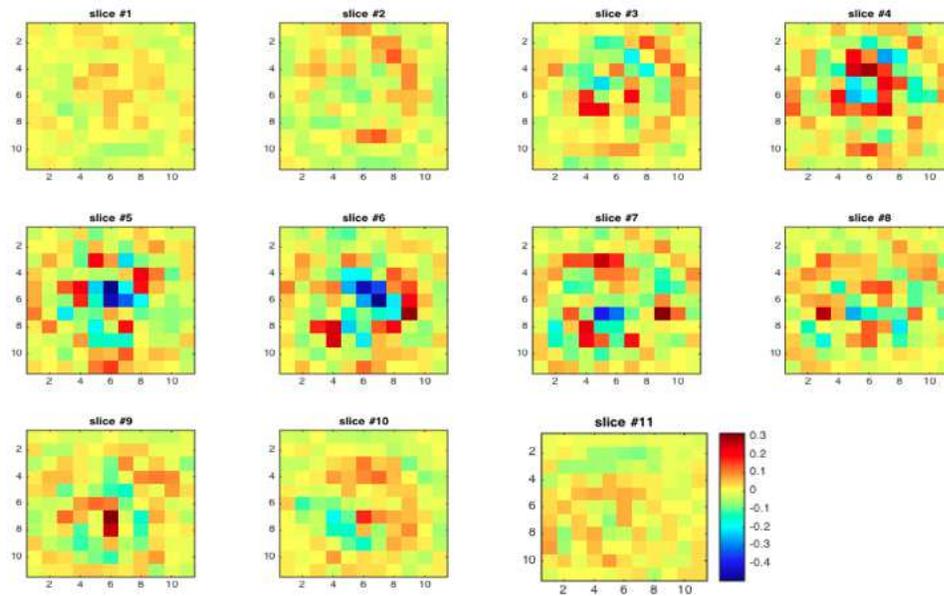


Figure 4.5: Slices through the difference in density(top to bottom) during chromophore swing obtained from matrix inversion when truncated at another mode, red as positive difference density and blue as negative difference density.

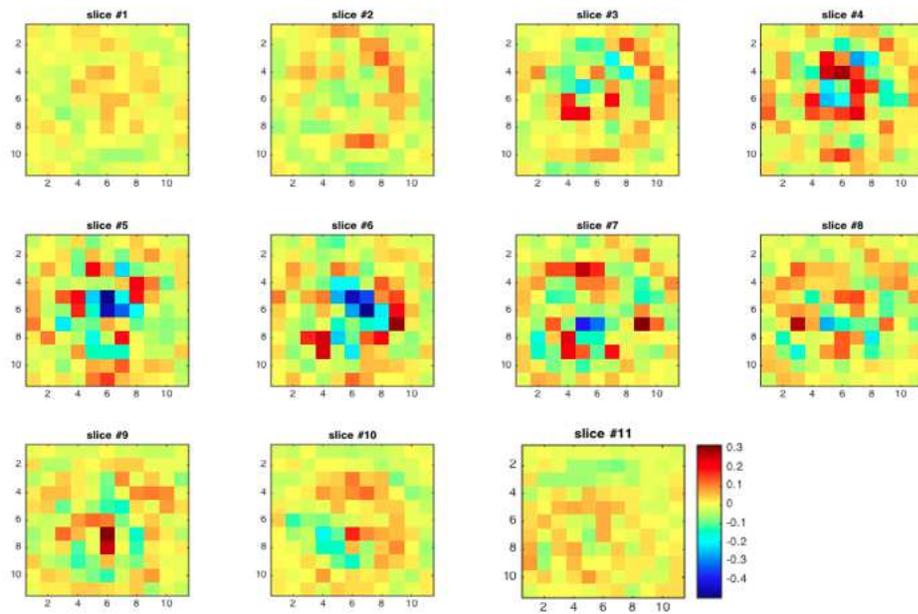


Figure 4.6: Slices through the difference in density(top to bottom) during chromophore swing obtained from matrix inversion when truncated at other mode different than earlier two, red as postive difference density and blue as negative difference density.

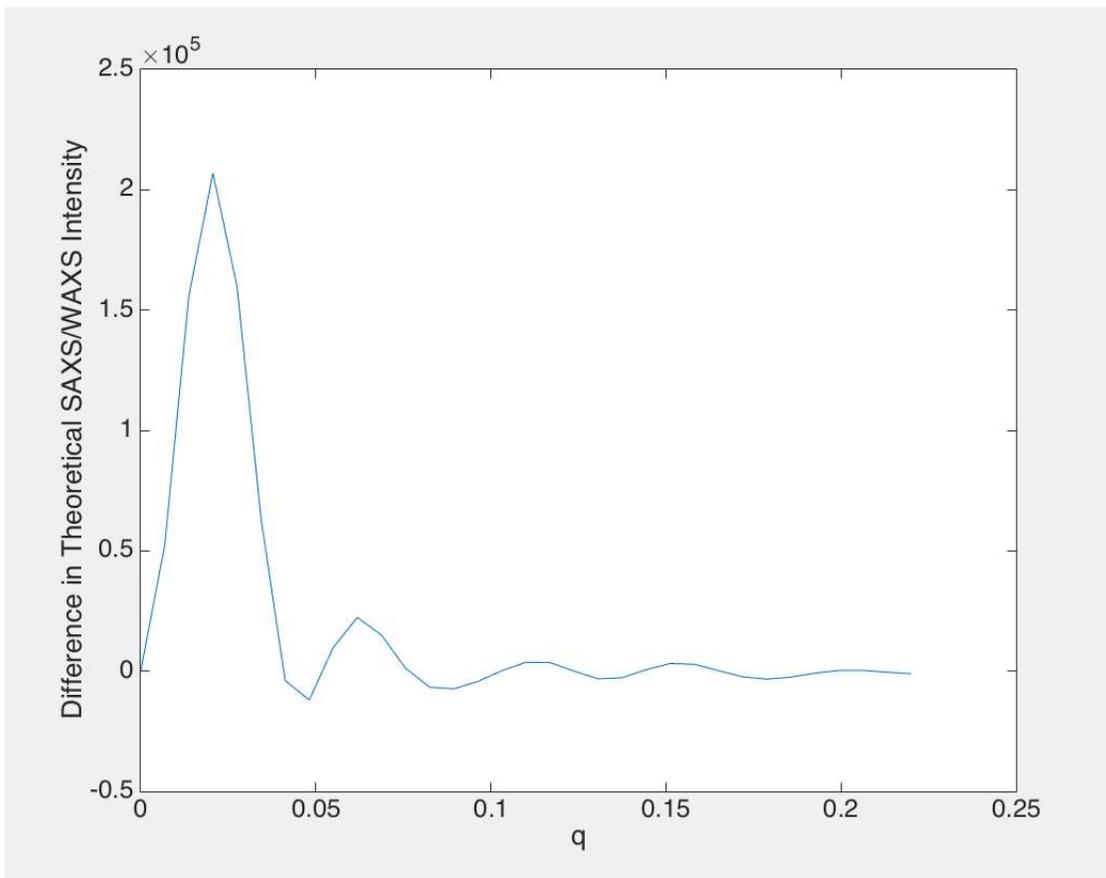


Figure 4.7: Difference in Theoretical(simulated) SAXS/WAXS Intensity for ARG124 Change.  $q$  values are in crystallographic covention with unit  $\text{\AA}^{-1}$ .

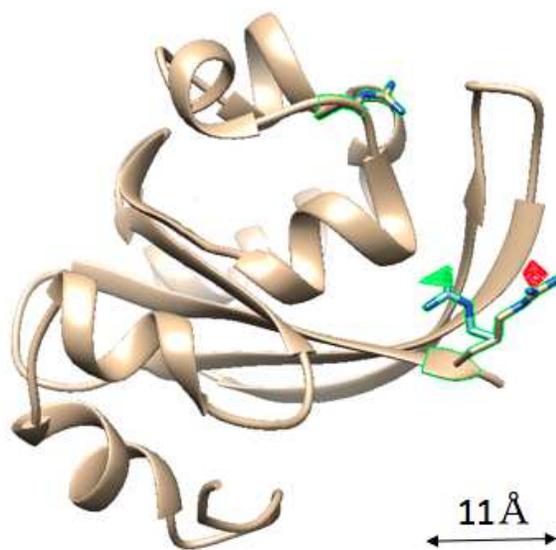


Figure 4.8: Same as Fig. 4.3 except that the same difference electron density is found for the movement of another residue by step 1 i.e. rapid direct method described in this study.

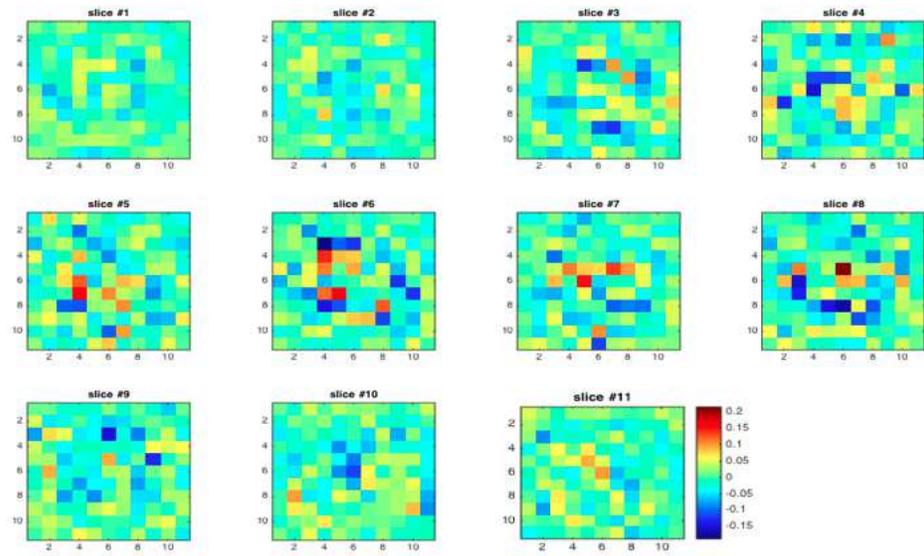


Figure 4.9: Slices through the difference in density (top to bottom) obtained from matrix inversion during Arginine shift, red as positive difference density and blue as negative difference density.

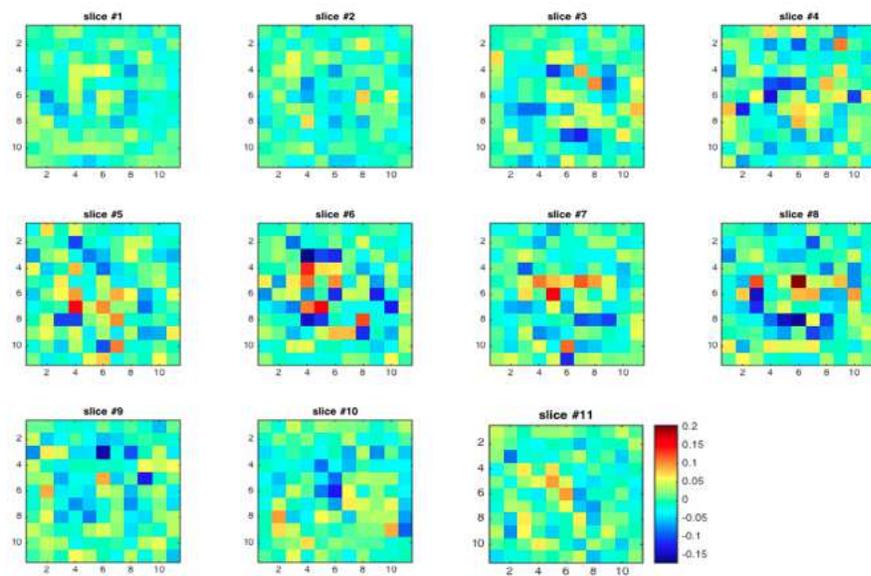


Figure 4.10: Slices through the difference in density(top to bottom) during Arginine torsional shift obtained from matrix inversion truncated at another mode, red as positive difference density and blue as negative difference density.

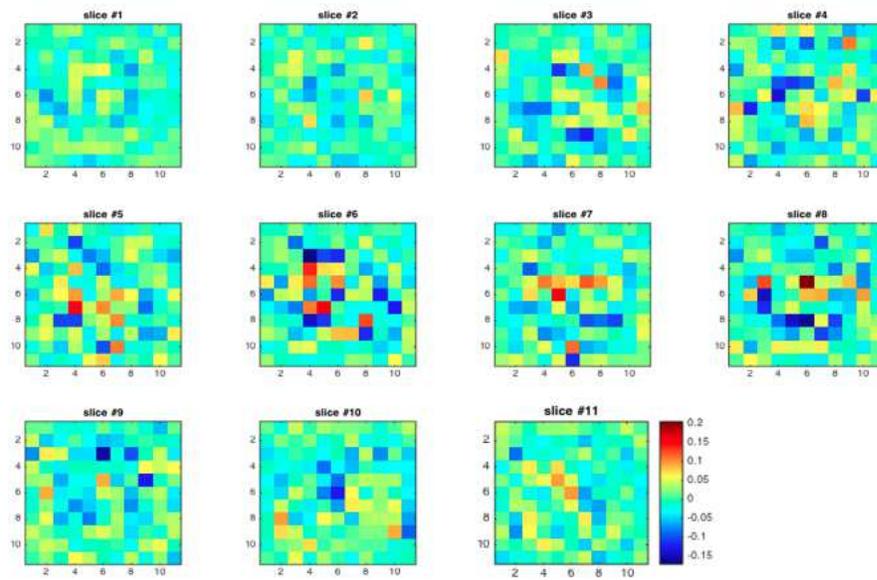


Figure 4.11: Slices through the difference in density(top to bottom) during Arginine torsional shift obtained from matrix inversion truncated at other mode different than earlier two, red as positive difference density and blue as negative difference density.

Table 4.1: Peaks for Chromophore swing case in terms of root mean square deviation(r.m.s.d.) of map

Negative r.m.s.d.	Postive r.m.s.d.
-6.8	+4.3
-6.2	+3.8
-5.1	+3.5
-3.6	+3.3

cited” state used for the calculations are shown in cyan. The difference map calculated from the theory presented in this dissertation is shown in Fig. 4.3 recovers locations from SAXS/WAXS intensity curves and a known reference structure. The greatest magnitude negative (-6.8 r.m.s.d.) and positive(4.3 r.m.s.d) peaks are localized in the region of moving chromophore atoms (Fig. 4.3). r.m.s.d. means root mean squared deviations. The other peaks are shown in Table 4.1

The difference map calculated for a torsion shift in the side chain of ARG-124 shown in Fig. 4.8 also recovers location from the difference in SAXS/WAXS intensity curves and the known reference structure is contoured at -3.6/+3.1 r.m.s.d.. The negative peak (-3.6 r.m.s.d) on at ARG-124 of the reference structure is the most negative feature in the map. However, the positive peak (3.1 r.m.s.d) on the shifted location of ARG-124 is not the first peak in the list. This is near the detection for this method. The other peaks are shown in Table 4.2 . The values for  $q_{max}$  in both cases is  $1.25 \text{ \AA}^{-1}$  (physics convention, corresponding to a resolution of almost  $5 \text{ \AA}$ ) with all maps calculated in a  $50 \times 50 \times 50 \text{ \AA}^3$  box with an  $11 \times 11 \times 11$  grid.

It will be noted that if one uses the information present in the peaks and dips of the electron density, these peaks and dips correctly identify the residues containing the moved atoms, although not much more due to the limitations of SAXS/WAXS data. Although it correctly identifies the residue in which the atomic displacements occur. It is clear on

Table 4.2: Peaks for ARG124 shift case in terms of root mean square deviation(r.m.s.d.) of map

Negative r.m.s.d.	Postive r.m.s.d.
-3.6	+4.0
-3.2	+3.6
-2.8	+3.4
-2.5	+3.1

comparison with Figs 4.3 and 4.12(b) on the one hand and Figs 4.8 and 4.13(b) on the other, that due to the paucity of data in SAXS/WAXS one cannot reconstruct an image of too high a resolution from step 1 i.e. rapid direct method algorithm alone.

In order to overcome this limitation step 2 is embarked on, which uses a simulated annealing method. Such a method is notorious in its bad scaling with the number of parameters varied. This is where we use the fact that the low-resolution maps recovered from step 1 of our algorithm although of low resolution is nevertheless able to recover the residue in which the atoms that are displaced are found.

### 4.3 Results with Simulated Annealing

After the implementation of Simulated Annealing algorithm, it is seen in Fig 4.12(a) that the small red and green lobes of Fig. 4.3 are now expanded out to be somewhat similar to Fig. 4.12(b) from the classical difference Fourier method of time-resolved crystallography. The control map for Fig 4.12(b) is calculated directly for a simulated change to the chromophore and neighboring ARG-52 residue. The red/green contours are at -9.3/+9.1 times the r.m.s of the map. At this resolution, peaks for the shift ARG-52 are not resolved. The value of  $q_{max}$  in both cases is  $1.25 \text{ \AA}^{-1}$  with maps calculated in a  $50 \times 50 \times 50 \text{ \AA}^3$  box with an  $11 \times 11 \times 11$  grid.

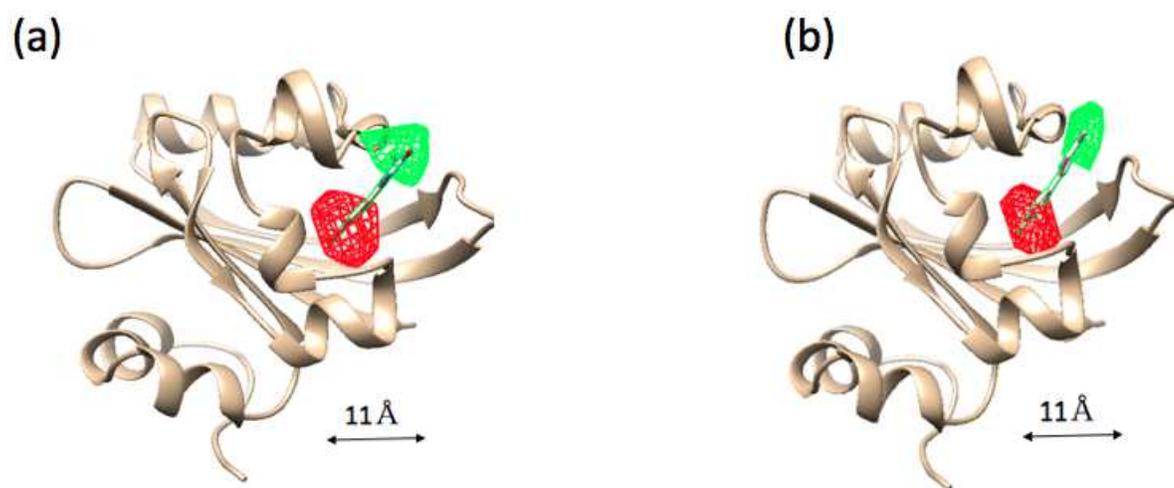


Figure 4.12: (a) Same as Fig. 4.3, but after two-step approach of this study i.e. rapid direct method and Simulated Annealing algorithm, (b) same as (a) except from the difference Fourier method of classical time-resolved crystallography.

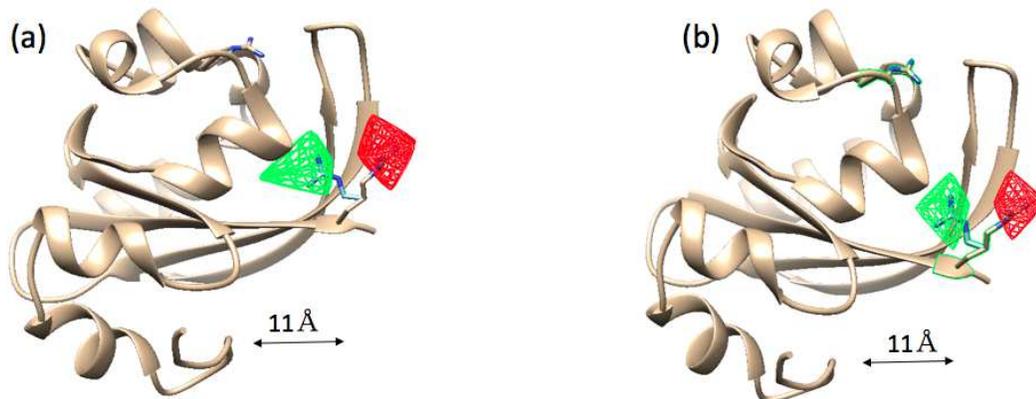


Figure 4.13: (a) Same as Fig. 4.8 except that the difference electron density is now recovered after two-step approach of this study i.e. rapid direct method and Simulated Annealing algorithm, (b) same as (a) except from the difference Fourier method of classical time-resolved crystallography.

The control map for Fig 4.13(b) is calculated directly for a simulated change to the torsion shift in the side chain of ARG-124 contoured at  $-7.3/+7.1$  r.m.s.d.. The value of  $q_{max}$  in both cases is  $1.25 \text{ \AA}^{-1}$  with maps calculated in  $50 \times 50 \times 50 \text{ \AA}^3$  box with an  $11 \times 11 \times 11$  grid.

Although it is obvious from the difference electron density maps that the agreement after the simulated annealing is superior, the relevant atomic coordinates of the unperturbed structure (the starting point) and of the perturbed structure (the ultimate goal) as well as of the coordinates as determined by simulated annealing are also plotted. It will be seen that the simulated annealing finds the correct coordinates to more or less  $2 \text{ \AA}$ . All of the figures are produced with Chimera [69].

Table 4.3: Atom coordinates of the chromophore of the unperturbed (u) structure.

Atom name	$x$	$y$	$z$
C4	-6.4	-1.3	-6.6
C5	-5.5	-2.1	-7.4
C6	-4.5	-2.8	-6.8
C7	-4.3	-2.7	-5.5
C8	-5.1	-2.0	-4.7
C9	-6.1	-1.2	-5.3
O2	-3.3	-3.3	-5.0

Figs. 4.12(a) and 4.12(b) are concerned with the displacements of the chromophore (which is expected on photoexcitation) but Figs. 4.13(a) and 4.13(b) are concerned with the artificial displacement of a residue far from the chromophore in order to test the algorithm. In all cases the equation (3.15) is used to get an approximate electron density and then varied the positions of atoms within the residue by simulated annealing to obtain the exact atomic displacements.

The efficacy of the simulated annealing algorithm may be judged by the results in Tables 4.3 , 4.4 , and 4.5 along with Tables 4.6, 4.7 , and 4.8. The plots of corresponding coordinates along with other sets of coordinates recovered by simulated annealing are also shown. At least in a case where all the atomic displacements are found within the residues identified by the direct method the new method seems to correctly identify the moved atoms as verified by a comparison with the results from the tested difference Fourier method of time-resolved crystallography. The 2nd, 3rd and 4th columns of each Table contain the x- and y- and z- coordinates of the atoms in Å.

Table 4.4: Atom coordinates of the chromophore of the perturbed (p) structure.

Atom name	$x$	$y$	$z$
C4	-7.5	-0.6	-8.8
C5	-6.7	-1.5	-9.5
C6	-6.9	-1.8	-11.0
C7	-7.8	-1.0	-11.5
C8	-8.7	-0.2	-10.9
C9	-8.5	0.5	-9.5
O2	-8.0	-1.3	-12.7

Table 4.5: First set of atomic coordinates of the chromophore recovered by simulated annealing, starting from the known unperturbed (u) structure. A total of ten sets of chromophore atomic coordinates recovered by simulated annealing are shown in Figure 4.14 .

Atom name	$x$	$y$	$z$
C4	-9.1	-1.0	-10.6
C5	-5.8	-1.1	-10.9
C6	-7.1	-1.6	-9.3
C7	-7.2	-0.5	-12.1
C8	-10.1	-0.1	-9.7
C9	-8.9	0.6	-10.2
O2	-6.3	-1.0	-11.7

Table 4.6: Atom coordinates of the ARG124 of the unperturbed (u) structure.

Atom name	$x$	$y$	$z$
CD	-11.6	-9.4	2.1
NE	-12.9	-9.3	1.4
CZ	-13.4	-10.3	0.7
NH1	-12.8	-11.4	0.5
NH2	-14.6	-10.1	0.1

Table 4.7: Atom coordinates of the ARG124 of the perturbed (p) structure.

Atom name	$x$	$y$	$z$
CD	-10.4	-7.9	2.9
NE	-10.2	-8.0	1.4
CZ	-9.1	-8.2	0.8
NH1	-7.9	-8.1	1.4
NH2	-9.1	-8.3	-0.4

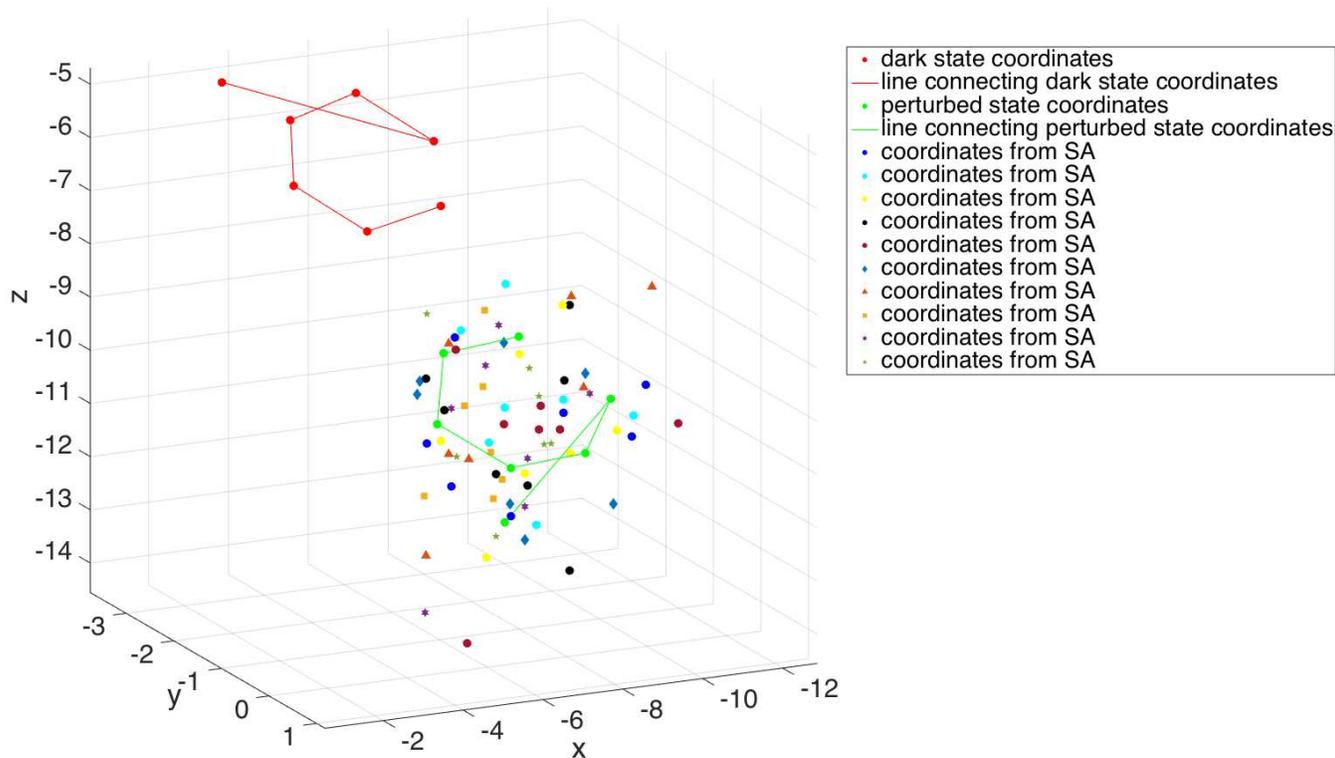


Figure 4.14: Figure showing atomic coordinates of the chromophore in unperturbed, perturbed states, and recovered by simulated annealing

Table 4.8: First set of atomic coordinates of the ARG124 recovered by simulated annealing, starting from the known unperturbed (u) structure. A total of nine sets of atomic coordinates of the ARG124 recovered by simulated annealing are shown in figure 4.15 .

Atom name	$x$	$y$	$z$
CD	-11.6	-7.6	3.0
NE	-7.5	-8.0	-1.2
CZ	-7.2	-9.0	-1.7
NH1	-10.6	-7.8	1.8
NH2	-7.5	-9.2	-0.8

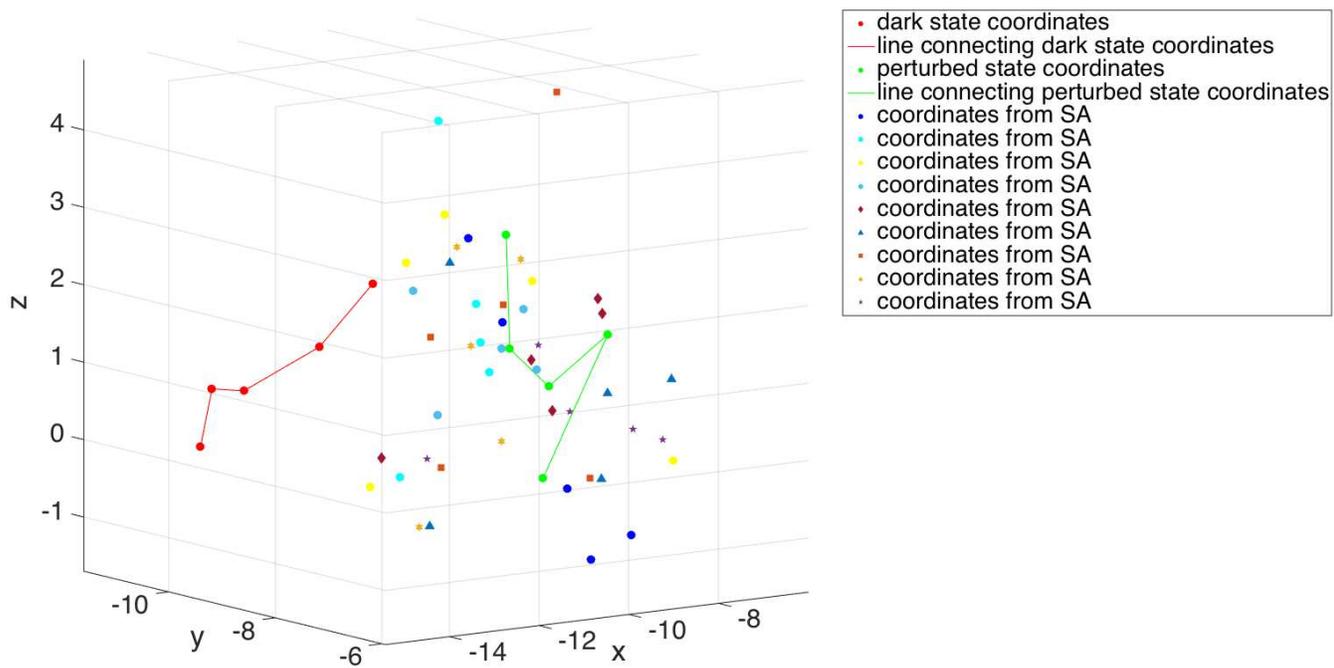


Figure 4.15: Figure showing atomic coordinates of the ARG124 in unperturbed, perturbed states, and recovered by simulated annealing

## 4.4 Consideration of Poisson Noise

The average number of photon counts per Shannon Pixel [70] is introduced as

$$\langle N_p \rangle = F \cdot P \cdot r_e^2 \cdot N_{atom} \cdot f^2 \cdot \Omega_p \quad (4.1)$$

where  $F$  is the photon fluence,  $N_{atom}$  is the number of atoms in the molecule,  $P$  is the polarization factor, and  $f$  is the average atomic scattering factor,  $\Omega_p$  is solid angle and for a particle of width  $w$  is  $\Omega_p \sim \frac{\lambda^2}{4w^2}$ .

Thus,

$$\langle N_p \rangle = F \cdot P \cdot r_e^2 \cdot N_{atom} \cdot f^2 \cdot \frac{\lambda^2}{4w^2} \quad (4.2)$$

Using  $F = \frac{10^{12} \text{photons}}{(100 \text{nm})^2}$  per pulse,  $r_e = 2.8179 \times 10^{-15} \text{m}$  and  $\lambda \sim 2\text{\AA}$ , PYP has around 1100 atoms,  $f$  at  $q_{max} \sim 6.18$ , diameter or width  $50\text{\AA}$  we get,  $\langle N_p \rangle \sim 0.013$  photons per Shanon pixel.

If the liquid jet containing the molecules can be reduced to  $0.1\mu\text{m}$  in width then the volume of liquid illuminated will be  $\sim \pi \cdot 5^2 \cdot 100 \text{nm}^3$  [71]. If the volume illuminated by XFEL beam could be arranged with suitable experimental conditions to contain around 80 molecules in random positions and random orientations along the beam, the above  $\langle N_p \rangle$  could be possibly improved to about 1 photon/Shanon Pixel.

The number of Shannon Pixels in outer resolution ring is

$$\langle N_{sp} \rangle_O = \frac{2\pi \cdot q_{max} \cdot \frac{1}{2w}}{\left(\frac{1}{2w}\right)^2} \quad (4.3)$$

If 1 photon/Shanon Pixel is taken then estimated number of photons in outermost

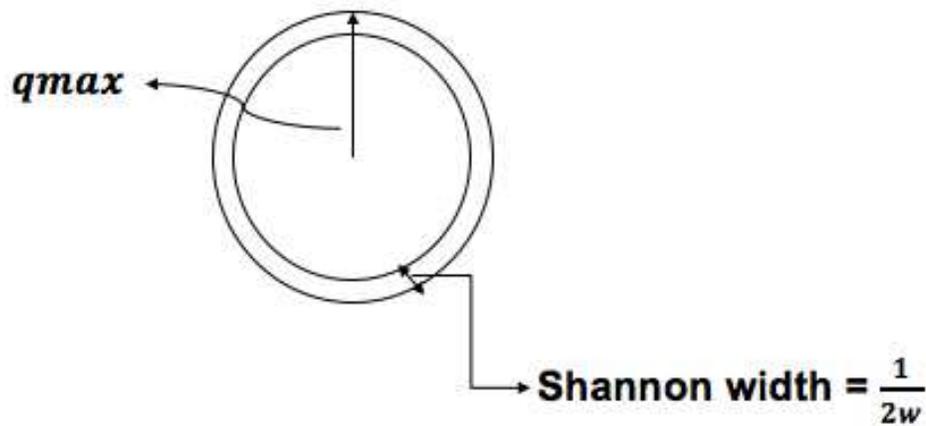


Figure 4.16: Figure showing outer resolution ring

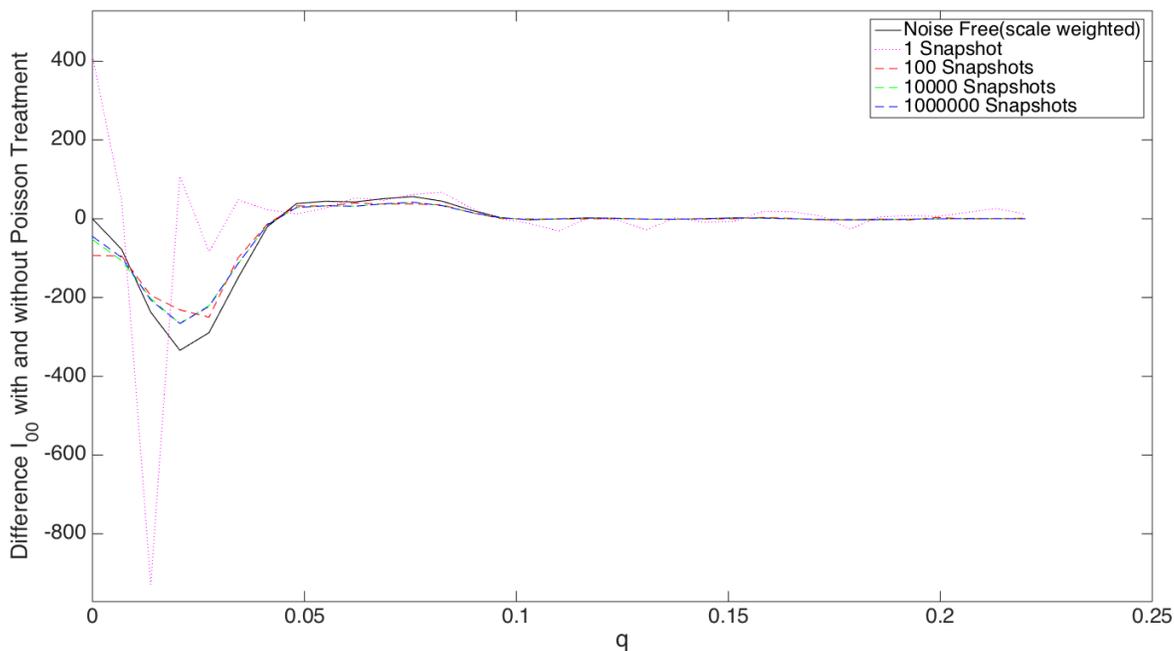


Figure 4.17: Difference in Simulated SAXS/WAXS intensity after Poisson noise treatment for different snapshots. One photon per Shannon pixel is considered in this case. The noise free difference SAXS/WAXS intensity values are weighted by a factor of  $10^{-3}$ .

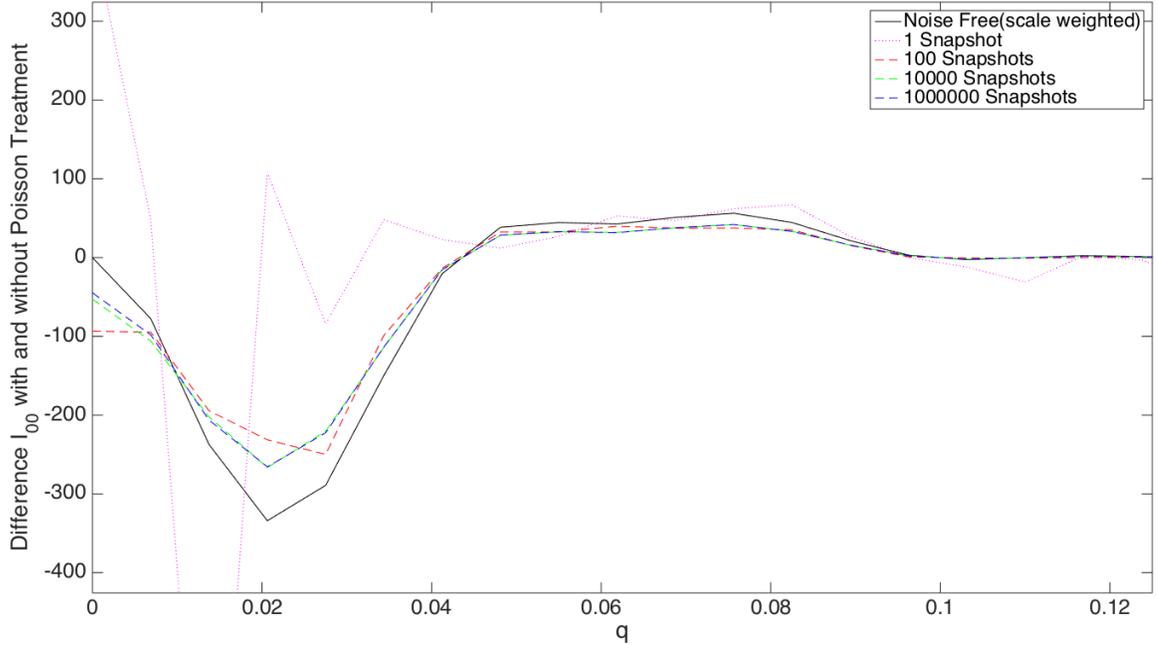


Figure 4.18: Zoomed plot of Fig 4.17 for difference in Simulated SAXS/WAXS intensity after Poisson noise treatment for different snapshots. The noise free difference SAXS/WAXS intensity values are weighted by a factor of  $10^{-3}$ .

resolution ring corresponding to resolution about  $5\text{\AA}$  is around 138. Then theoretical SAXS/WAXS intensity could be converted into experimentally realizable SAXS/WAXS intensity in terms of no. of photons.

In order to study the effect of noise on the difference in density, introduction of shot noise (following Poisson distribution[72]) to experimentally realizable intensity is needed. Few cases of noise implementation and their results are shown in Fig 4.19, Fig 4.20 and Fig 4.21.

Intially 10,000 snapshots are considered and high number of molecules i.e one billion and ten million are taken. The difference map calculated recovers locations from noise implemented SAXS/WAXS intensity curves and a known reference structure. The great-

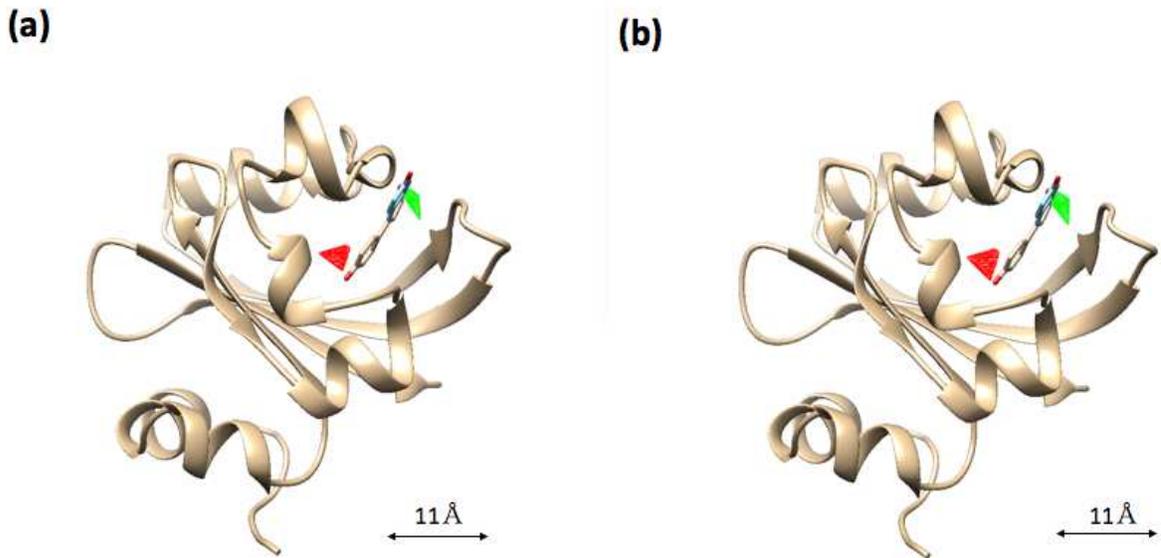
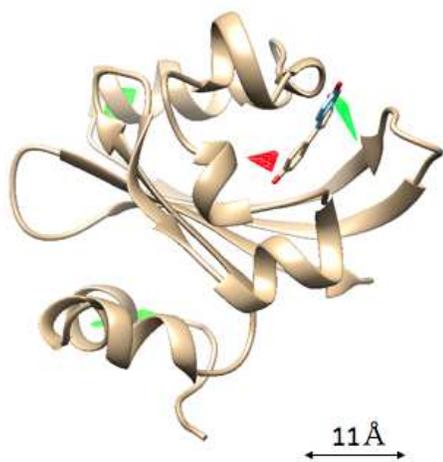


Figure 4.19: (a) For 13,000,000 photons per Shannon Pixel corresponding to billion particles and with 10,000 snapshots, (b) For 130,000 photons per Shannon Pixel corresponding to ten million particles and with 10,000 snapshots.

est negative and positive peaks are localized in the region of moving chromophore atoms as shown in Fig 4.19(a and b).

Then with the same 10,000 snapshots but one million molecules are taken. It is seen that the greatest negative peak is localized at the position of chromophore of the reference structure, however, the positive peak on the swung location of chromophore is the second strongest peak (Fig 4.20 (a)) when difference map is calculated. This is near the detection for this method. Similarly, with the same 10,000 snapshots but with hundred thousand molecules it is noticed again that the first strongest negative peak is at original position of chromophore and the positive peak on the swung position of chromophore is the second most positive (Fig 4.20(b)). This is also near the detection for this method.

**(a)**



**(b)**

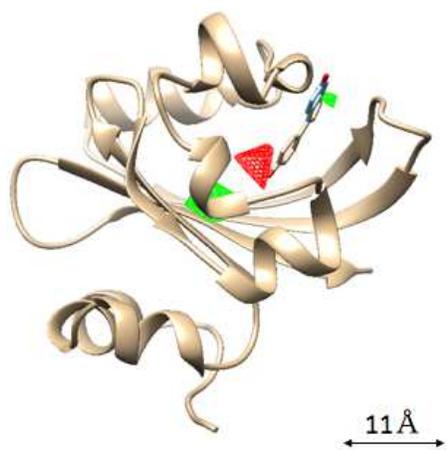
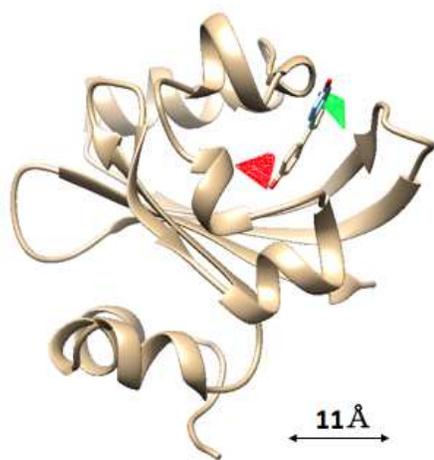


Figure 4.20: (a) For 13,000 photons per Shannon Pixel corresponding to million particles and with 10,000 snapshots, (b) For 1300 photons per Shannon Pixel corresponding to hundred thousand particles and with 10,000 snapshots.

**(a)**



**(b)**

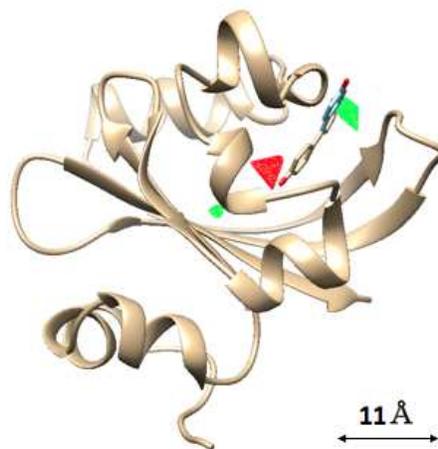


Figure 4.21: (a) For 13,000 photons per Shannon Pixel corresponding to million particles and with 1 million snapshots, (b) For 1300 photons per Shannon Pixel corresponding to hundred thousand particles and with 1 million snapshots.

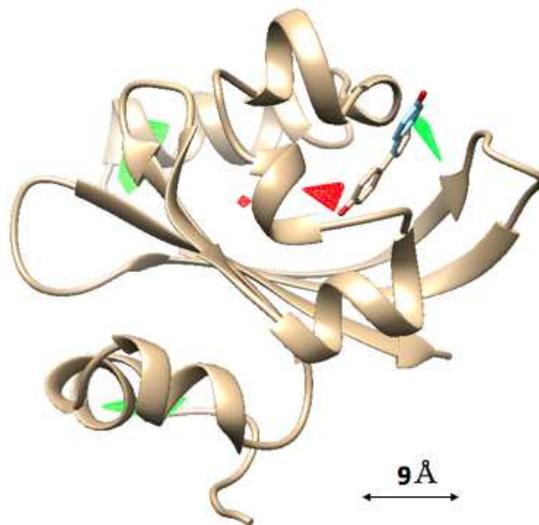


Figure 4.22: For 130 photons per Shannon Pixel corresponding to ten thousand particles and with 1 million snapshots.

With the hope that more number of snapshots would help, one million snapshots are taken into consideration. And it really does. For one million molecules and one million snapshots, the difference map calculated from the theory of this study recovers locations from noise implemented SAXS/WAXS intensity curves and a known reference structure. The greatest negative and positive peaks are localized in the region of moving chromophore atoms as shown in Fig 4.21 (a). However, with one million snapshots and one hundred thousand molecules interesting feature is seen. The negative peak at the chromophore position of the reference structure is the most negative feature in the map whereas the both positive peaks i.e. one at the swung position of chromophore and another a bit further away from the chromophore appears simultaneously and are the first peaks of the map. This is also near the detection for this method.

Furthermore with one million snapshots and ten thousand molecules, it is seen that

the negative peak at the chromophore location of the reference structure is the strongest negative peak whereas the positive peak at the swung position of chromophore is the third strongest peak. This may be also near the detection for this method.

## 4.5 Experimental SAXS/WAXS data

Experimental data was obtained from Prof. Ihee's group [67] to apply this theoretical approach. The best hope that it possibly work was the time-resolved difference SAXS/WAXS intensity for *2ms* structural changes shown in Fig 4.23 .

The results are not quite successful as the data are polychromatic and also difficult to identify right dark structure(3PHY.pdb) from Protein Data Bank [74]. Future study is to get monochromatic XFEL data from Levantino et al. or Neutze et al. to apply this theoretical approach and hope to explain structural changes involved in protein quake in their recent work (from XFELs SAXS/WAXS).

## 4.6 Limitations of this proposed method

The main limitation of this study is paucity of SAXS/WAXS data. After thorough study with simulation, it is safer to say that the theory described in Chapter 3 for rapid direct method is an approximate theory rather than an exact theory. In general, due to the difficulty of producing true direct methods for getting difference electron densities from the data one uses some approximation. The problem is linearized to find an approximate direct method. This is justified at some level as the Taylor series expansion enters at some stage. The whole approach with the direct method is only justified in this linear limit. A

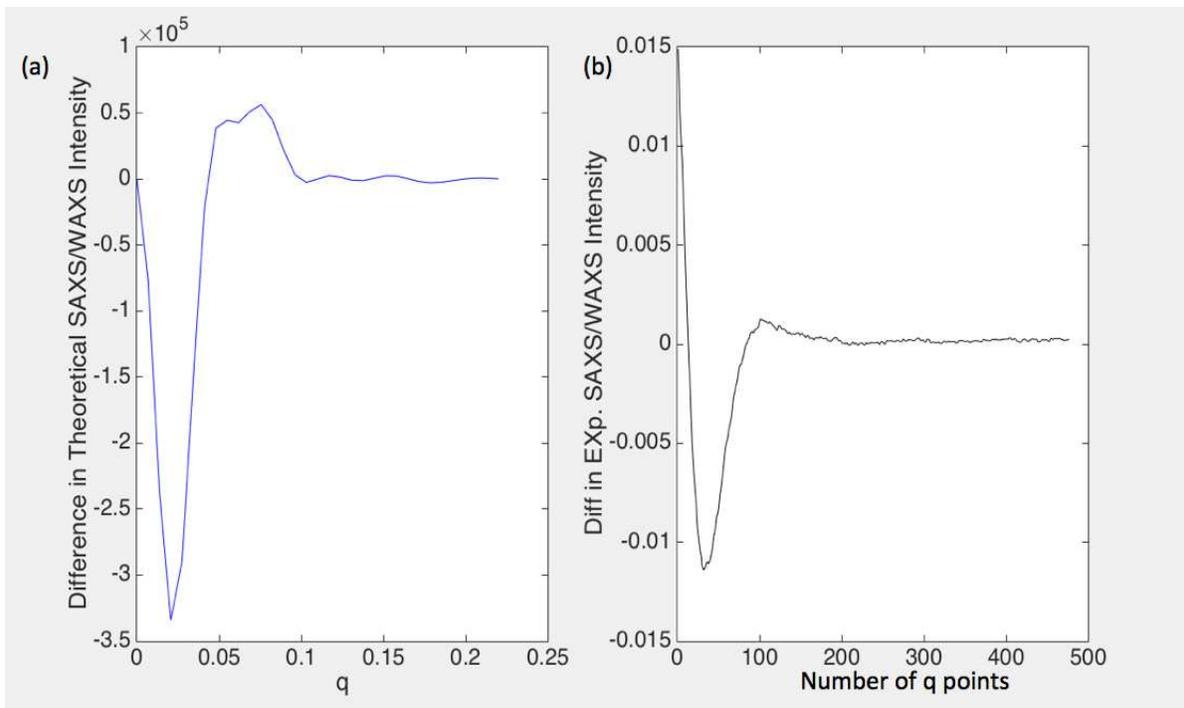


Figure 4.23: Comparison of Simulated SAXS/WAXS intensity (panel a) to experimentally measured SAXS/WAXS Intensity (panel b)

fuller justification of everything that needs to be taken into account for reproducing the data includes nonlinear terms. It is unclear that non-linear terms can be included in a direct method.

Using equation  $\delta\rho_u(\mathbf{r}_k) = M_{\mathbf{k},q}^{-1}\delta I_u(q)$  to find the difference electron density has also some limitations. The pseudo-inverse of a matrix allows directly to get the approximate electron density from simulated SAXS/WAXS difference data and known atomic coordinates of the “dark” structure on one hand but there is a loss of information on the other hand. This loss of information during psuedo-inversion may be one of the reasons that the peaks and dips of the electron density are not much more to cover the region around moved residues, although those peaks and dips correctly identifies the residues in which atomic displacement occur.

The step 2 ie Simulated Annealing method of this study is for the refinement purpose only once the rapid direct method identifies the perturbed residues. It enables significant reduction in the search space, provides a more reasonable starting configuration and hence cuts down on the time necessary to carry out the Simulated Annealing. Moreover, Simulated Annealing method here doesnot use chemical constraints/guidance. If used, the calculations probably speed up and there might be even close agreement in the atomic coordinates of perturbed state. It should be noted that Fig 4.3 and Fig 4.8 are the results from the rapid direct method only but Fig 4.12(a) and Fig 4.13(a) are the results from two-step approaches i.e. direct method for extraction of an approximate difference electron density and Simulated Annealing method. The other limitations are described in Conclusion and Outlook section.

# Chapter 5

## Conclusion and Outlook

The results in Chapter 4 are of extraordinary importance. Up to now it has been necessary to crystallize proteins to find the time variation of such molecules. This is a limitation for two reasons. Not all proteins are easily crystallized. It is a rarity that molecules are crystallized in physiological process of life. Indeed there are a number of important membrane proteins that are very difficult to crystallize due to the presence of hydrophobic surfaces on their exterior. Even one of the scientific justifications of development of XFEL sources was the possible structural determination of important proteins without the need for crystallization [75, 76]. Secondly it is questionable whether the time-resolved structures found by conventional time-resolved crystallography are not artificial due to the presence of steric constraints from neighboring unit cells.

The two-step process i.e. Rapid direct method for extraction of difference electron density and Simulated Annealing described in this study eventually gives a resolution of disordered molecules in SAXS/WAXS similar to that which is obtained in classical time-resolved crystallography on crystallized proteins. It is certainly arguable therefore that structures recovered from solution are more representative of proteins under physiological conditions.

This study successfully demonstrated the validity of algorithm but there are a number of assumptions and limitations. Firstly, it is assumed that the magnitudes of atomic displacements seen in time-resolved crystallography are also same in SAXS/WAXS experiment. Accordingly, the assumption is that there is an average behaviour in solution that leads to a narrow distribution within the each member of the pair of states. Also in the artificial data the effect of the solvent is neglected, which is a source of noise in difference spectra. Few cases of noise are assumed in results/demonstration and the detailed effect of solvent remains to be investigated.

Its possible that the development of a reliable and reasonably fast method for SAXS/WAXS could be of crucial importance for the study of time-resolved structures under physiological conditions. For example there has been a recent experiment by Arnlund et al. [12], which purports to explain the apparent promotion of chemical reactions that require only meV of energy though the absorption of a photon from the sun that has perhaps an energy of about an eV. The postulation is the formation of a “protein quake” in the material to dissipate most of the energy. There is some evidence for this from molecular dynamics, although since this is a theoretical technique that depends on a number of assumptions, this evidence is not as strong as if structural evidence of a protein quake were obtained directly from SAXS/WAXS experiments alone.

What has been lacking up to now is a method of recovering the structure of time-resolved changes directly from experiments. It is hoped that this study and results demonstrated will be able to work directly with the experimental data to clarify this important event in the earth’s history. The method has performed well in initial tests on simulated, artificial data. It remains to be seen how the theory here can be applied to experimental/real data and the next future plan is to request and get monochromatic

difference SAXS/WAXS data from different experimental groups and test the efficacy of this theory.

Of course an assumption is that the step 1 (rapid direct method) identifies the residues in which atoms move. These ideas extend the previous work on extracting information from difference in SAXS/WAXS spectra [40, 43] by providing direct and objective locations of differences in molecular structure. While this is true of time-resolved changes of PYP as a result of photoexcitation, it remains to be seen how general this is of typical time-resolved structure.

The power of this study is demonstrated in objectively recovering spatial information from simulated SAXS/WAXS data. The calculation from traditional crystallographic methods, the displacement of the chromophore in the photoactive yellow protein, PYP (PDB ID 2phy) [68]. The equation 3.15 is used to directly get an approximate electron density from simulated SAXS data and known atomic coordinates of the “dark” structure. Thus the validity of algorithm on artificial data is successfully demonstrated. Of course the method is not a high-resolution method, but nevertheless has the potential to reveal differences in a spatial coordinate system that identifies their locations in three dimensions at a residue level.

The significance of this achievement is that there is no need of crystals of perturbed state but only a “dark” state and two SAXS curves to identify directly, the nature of perturbation. Despite an imbalance between the number of data points from the experimental data and the number of grid points at which information on the difference density is sought, the additional knowledge of the atomic structure from other methods can thus be brought into play to improve the extraction of information using this method.

Even if one can only obtain a resolution to the level of identifying the residue in which atoms have changed their positions, since the average residue in a protein has only a relatively small number of atoms, it is shown through step 2 that it is possible with higher resolution WAXS data to “refine” the positions of these atoms with a global optimization algorithm such as simulated annealing [20] to get a reasonably accurate atomic solution of the perturbed state [21].

Lastly, showing the inversion of time-resolved SAXS/WAXS difference data to identify the nature of perturbation with the prior knowledge of unperturbed state, significant reduction in the search space allowing the extraction of structural change by optimization algorithm like simulated annealing enabling a huge reduction in computing time along with better refinement and possibility of obtaining time-resolved structures of vital proteins/biomolecules without the need of crystallization are three major findings of this study.

# References

- [1] "Small Angle X-ray Scattering," edited by O. Glatter O. Kratky, Academic Press (1982)
- [2] "The Small Angle X-ray Scattering Technique: An Overview," <http://people.unipmn.it/gcroce/download/theory.pdf>
- [3] P. Debye, "Zerstreuung von Rontgenstrahlen," *Annalen der Physik*, 351:80923 (1915).
- [4] A Guinier, "Ann. Phys. (Paris)," **12** 161237 (1939).
- [5] H. B. Stuhrmann, "New method for determination of surface form and internal structure of dissolved globular proteins from small angle x-ray measurements," *Z. Phys Chem Neue Fol* **72**, 177-184 (1970).
- [6] "SASHA: Small Angle Scattering Shape Determination," Biological small angle scattering group, EMBL Hamburg.
- [7] D. I. Svergun, V. V. Volkov, M.B. Kozin and H. B. Stuhrmann, "New developments in direct shape determination from small-angle scattering. 2. Uniqueness," *Acta Crystallogr.***A52** (6): 419426 (1996).
- [8] Wenjun Zheng, Sebastian Doniach, "Fold recognition aided by constraints from small angle X-ray scattering data," *Protein Engineering, Design Selection* vol. 18 no. 5 pp. 209219 (2005).

- [9] H. B. Stuhrmann, “Interpretation of small-angle scattering functions of dilute solutions and gases - a representation of structures related to a one-particle-scattering function,” *Acta Crystallogr A* **26**, 297-306 (1970).
- [10] D. I. Svergun and H. B. Stuhrmann, “New developments in direct shape determination from small-angle scattering .1. theory and model-calculations,” *Acta Crystallogr A* **47**, 736-744 (1991).
- [11] Richard Neutze and Keith Moffat, “Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges,” *Current Opinion in Structural Biology*, **22**:651659 (2012).
- [12] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. De Ponte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. Berndt van Driel, R. B. Doak, K. S. Kjaer, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosholeva, C. Kupitz, M. Liang, A. V. Martin, M. M. Nielson, M. Meese/Schmidt, M. M. Siebert, J. Sjömann, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. H. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman and R. Neutze, “Visualizing a protein quake with time-resolved X-ray scattering at a free electron laser”, *Nature Methods* **11**, 923-926 (2014).
- [13] K. Moffat, “Time-resolved crystallography,” *Acta Crystallogr A* **54**, 833-841 (1998).
- [14] M. Schmidt, “Structure based kinetics by time-resolved X-ray crystallography,” in *Ultra-short Laser Pulses in Biology and Medicine*, edited by M. Braun, P. Gilch and W. Zinth (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 201241.
- [15] K. Pande, C. D. M. Hutchison, G. Groenhof, A. Aquila, J. S. Robinson, J. Tenboer, S. Basu, S. Boutet, D. P. DePonte, M. N. Liang, T. A. White, N. A. Zatsepin, O.

- Yefanov, D. Morozov, D. Oberthuer, C. Gati, G. Subramanian, D. James, Y. Zhao, J. Koralek, J. Brayshaw, C. Kupitz, C. Conrad, S. Roy-Chowdhury, J.D.Coe, M. Metz, P. L. Xavier, T. D. Grant, J. E. Koglin, G. Ketawala, R. Fromme, V. Srajer, R. Henning, J. C. H. Spence, A. Ourmazd, P. Schwander, U. Weierstall, M. Frank, P. Fromme, A. Barty, H. N. Chapman, K. Moffat, J. J. van Thor, and M. Schmidt, “Femtosecond structural dynamics drives the tran/cis isomerization in photoactive yellow protein,” *Science* **352**, 725-729 (2016).
- [16] M. Schmidt, R. Pahl, V. Srajer, S. Anderson, Z. Ren, H. Ihee, S. Rajagopal, K. Moffat, “Protein kinetics: structures of intermediates and reaction mechanism from time-resolved x-ray data,” *Proc. Natl Acad. Sci. USA* **101**, 47994804 (2004).
- [17] YO Jung, JH Lee, J Kim, M Schmidt, K. Moffat, V. Srajer, H. Ihee, “Volume-conserving trans-cis isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography ”, *Nat. Chem.* **5**, 212 - 220 (2013).
- [18] Friedrich Schotte, Hyun Sun Cho, Ville R. I. Kaila, Hironari Kamikubo, Naranbaatar Dashdorj, Eric R. Henry, Timothy J. Graber, Robert Henning, Michael Wulff, Gerhard Hummer, Mikio Kataoka, and Philip A. Anfinrud, “Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography ”, *Proc. Natl Acad. Sci. USA* **109**, 19256 - 19261 (2012).
- [19] Hyotcherl Ihee, Sudarshan Rajagopal, Vukica rajer, Reinhard Pahl, Spencer Anderson, Marius Schmidt, Friedrich Schotte, Philip A. Anfinrud, Michael Wulff, and Keith Moffat, “Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds ”, *Proc. Natl Acad. Sci. USA* **102**, 7145 - 7150 (2005).
- [20] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, “Optimization by simulated annealing”, *Science* **220**. 671-680 (1983).

- [21] H.C. Poon, M. Schmidt, and D. K. Saldin, “Extraction of fast changes in the structure of a disordered ensemble of photoexcited biomolecules”, *Adv Cond Matter Phys*, 750371 (2013).
- [22] W. A. Barletta, J. Bisognano, J. N. Corlett, P. Emma, Z. Huang, K. J. Kim, R. Lindberg, J. B. Murphy, G. R. Neil, D. C. Nguyen, C. Pellegrini, R. A. Rimmer, F. Sannibale, G. Stupakov, R. P. Walker, and A. A. Zholents, “Free electron lasers: Present status and future challenges,” *Nucl. Instrum. Methods Phys. Res., Sect. A* **618**, 69 (2010).
- [23] P. Emma et al., “First lasing and operation of an angstrom-wavelength free-electron laser,” *Nat. Photonics* **4**, 641647 (2010).
- [24] Adrian Cho, “What Shall We Do With the X-ray Laser?” *Science* **330**, 470-471 (2010).
- [25] Henry N. Chapman, Petra Fromme, Anton Barty, Thomas A. White, Richard A. Kirian, Andrew Aquila, Mark S. Hunter, Joachim Schulz, Daniel P. DePonte, Uwe Weierstall, R. Bruce Doak, Filipe R. N. C. Maia, Andrew V. Martin, Ilme Schlichting, Lukas Lomb, Nicola Coppola, Robert L. Shoeman, Sascha W. Epp, Robert Hartmann, Daniel Rolles, Artem Rudenko, Lutz Foucar, Nils Kimmel, Georg Weidenspointner, Peter Holl, Mengning Liang, Miriam Barthelmess, Carl Caleman, Sbastien Boutet, Michael J. Bogan, Jacek Krzywinski, Christoph Bostedt, Saa Bajt, Lars Gumprecht, Benedikt Rudek, Benjamin Erk, Carlo Schmidt, Andr Hmke, Christian Reich, Daniel Pietschner, Lothar Strder, Gnter Hauser, Hubert Gorke, Joachim Ullrich, Sven Herrmann, Gerhard Schaller, Florian Schopper, Heike Soltau, Kai-Uwe Khnel, Marc Messerschmidt, John D. Bozek, Stefan P. Hau-Riege, Matthias Frank, Christina Y. Hampton, Raymond G. Sierra, Dmitri Starodub, Garth J. Williams, Janos Hajdu, Nicusor Timneanu, M. Marvin Seibert, Jakob Andreasson, Andrea

- Rocker, Olof Jnsson, Martin Svenda, Stephan Stern, Karol Nass, Robert Andritschke, Claus-Dieter Schrter, Faton Krasniqi, Mario Bott, Kevin E. Schmidt, Xiaoyu Wang, Ingo Grotjohann, James M. Holton, Thomas R. M. Barends, Richard Neutze, Stefano Marchesini, Raimund Fromme, Sebastian Schorb, Daniela Rupp, Marcus Adolph, Tais Gorkhover, Inger Andersson, Helmut Hirsemann, Guillaume Potdevin, Heinz Graafsma, Bjrn Nilsson John C. H. Spence, “Femtosecond X-ray protein nanocrystallography”, *Nature* **470**, 73-81 (2010).
- [26] Sbastien Boutet, Lukas Lomb, Garth J. Williams, Thomas R. M. Barends, Andrew Aquila, R. Bruce Doak, Uwe Weierstall, Daniel P. DePonte, Jan Steinbrener, Robert L. Shoeman, Marc Messerschmidt, Anton Barty, Thomas A. White, Stephan Kasse-  
meyer, Richard A. Kirian, M. Marvin Seibert, Paul A. Montanez, Chris Kenney, Ryan Herbst, Philip Hart, Jack Pines, Gunther Haller, Sol M. Gruner, Hugh T. Philipp, Mark W. Tate, Marianne Hromalik, Lucas J. Koerner, Niels van Bakel, John Morse, Wilfred Ghonsalves, David Arnlund, Michael J. Bogan, Carl Caleman, Raimund Fromme, Christina Y. Hampton, Mark S. Hunter, Linda C. Johansson, Gergely Katona, Christopher Kupitz, Mengning Liang, Andrew V. Martin, Karol Nass, Lars Redecke, Francesco Stellato, Nicusor Timneanu, Dingjie Wang, Nadia A. Zatsepin, Donald Schafer, James Defever, Richard Neutze, Petra Fromme, John C. H. Spence, Henry N. Chapman, Ilme Schlichting, “High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography,” *Science* **337**, 362-364 (2012).
- [27] LCLS fact sheet 2014, [portal.slac.stanford.edu](http://portal.slac.stanford.edu)
- [28] H. N. Chapman et al., *Phil. Trans. Roy. Soc Bi* **369**, 20131313 (2014).
- [29] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, *Nature* **406**, 752757 (2000).

- [30] J. C. H. Spence, U. Weierstall, and H. N. Chapman, “X-ray lasers for structural and dynamic biology,” *Rep. Prog. Phys.* **75**, 102601 (2012).
- [31] R. Neutze and K. Moffat, “ Time-resolved structural studies at synchrotrons and x-ray free electron lasers: Opportunities and challenges,” *Curr. Opin. Struct. Biol.* **22**, 651659 (2012).
- [32] Matteo Levantino, Giorgio Schiro, Henrik Till Lemke, Grazia Cottone, James Michael Glowacki, Diling Zhu, Mathieu Chollet, Hyotcherl Ihee, Antonio Cupane Marco Cammarata, “Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser”, *Nature Communications* **6**,6772 (2015).
- [33] K. Moffat, “Time-resolved macromolecular crystallography,” *Annual Review of Biophysics and Biophysical Chemistry*, vol. 18, pp. 309332 (1989).
- [34] M. Schmidt, H. Ihee, R. Pahl, and V. Srajer, “Protein-ligand interaction probed by time-resolved crystallography,” *Methods in Molecular Biology*, vol. 305, pp. 115154 (2005).
- [35] K. Moffat, “Time-resolved biochemical crystallography: a mechanistic perspective,” *Chemical Reviews*, vol. 101, no. 6, pp. 15691581 (2001).
- [36] Marten H. Vos, Fabrice Rappaport, Jean-Christophe Lambry, Jacques Breton Jean-Louis Martin, “ Visualization of coherent nuclear motion in a membrane protein by femtosecond spectroscopy ”, *Nature* **363**, 320-325 (1993).
- [37] Pradeep L. Ramachandran, Janet E. Lovett, Patrick J. Carl, Marco Cammarata, Jae Hyuk Lee, Yang Ouk Jung, Hyotcherl Ihee, Christiane R. Timmel, Jasper J. van Thor, “ The Short-Lived Signaling State of the Photoactive Yellow Protein Photoreceptor Revealed by Combined Structural Probes”, *Journal of the American Chemical Society*, **2011**, 133 (24), pp 9395-9404.

- [38] Heikki Takala, Alexander Bjrling, Oskar Berntsson, Heli Lehtivuori, Stephan Niebling, Maria Hoernke, Irina Kosheleva, Robert Henning, Andreas Menzel, Janne A. Ihalainen Sebastian Westenhoff, “Signal amplification and transduction in phytochrome photosensors ”, *Nature* **509**, 245-248 (2014).
- [39] Magnus Andersson, Erik Malmerberg, Sebastian Westenhoff, Gergely Katona, Marco Cammarata, Annemarie B. Whri, Linda C. Johansson, Friederike Ewald, Mattias Eklund, Michael Wulff, Jan Davidsson, Richard Neutze, “ Structural Dynamics of Light-Driven Proton Pumps ”, *Structure* **17**, 1265 - 1275 (2009).
- [40] Erik Malmerberg, Ziad Omran, Jochen S. Hub, Xuewen Li, Gergely Katona, Sebastian Westenhoff, Linda C. Johansson, Magnus Andersson, Marco Cammarata, Michael Wulff, David van der Spoel, Jan Davidsson, Alexandre Specht, and Richard Neutze, “ Time-Resolved WAXS Reveals Accelerated Conformational Changes in Iodoretinal-Substituted Proteorhodopsin ”, *Biophys. J.* **101**, 1345 - 1353 (2011).
- [41] Marco Cammarata, Matteo Levantino, Friedrich Schotte, Philip A Anfinrud, Friederike Ewald, Jungkweon Choi, Antonio Cupane, Michael Wulff Hyotcherl Ihee, “Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering ”, *Nature Methods* **5**, 881 - 886 (2008).
- [42] Kyung Hwan Kim, Srinivasan Muniyappan, Key Young Oang, Jong Goo Kim, Shunsuke Nozawa, Tokushi Sato, Shin-ya Koshihara, Robert Henning, Irina Kosheleva, Hosung Ki, Youngmin Kim, Tae Wu Kim, Jeongho Kim, Shin-ichi Adachi, and Hyotcherl Ihee, “ Direct Observation of Cooperative Protein Structural Dynamics of Homodimeric Hemoglobin from 100 ps to 10 ms with Pump-Probe X-ray Solution Scattering ”, *J. Am. Chem. Soc.* **134**, 70017008 (2012).

- [43] Sena Ahn, Kyung Hwan Kim, Youngmin Kim, Jeongho Kim and Hyotcherl Ihee, “ Protein Tertiary Structural Changes Visualized by Time-Resolved X-ray Solution Scattering ”, *J. Phys. Chem. B* **113**, 1313113133 (2009).
- [44] Hyun Sun Cho, Naranbaatar Dashdorj, Friedrich Schotte, Timothy Graber, Robert Henning, and Philip Anfinrud, “ Protein structural dynamics in solution unveiled via 100-ps time-resolved x-ray scattering ”, *Proc. Natl. Acad. Sci. USA* **107**, 72817286 (2010).
- [45] D. T. Cromer, J. B. Mann, X-ray scattering factors computed from numerical Hartree-Fock wave functions, *Acta Cryst.* (1968). A24, 321-324
- [46] INTERNATIONAL TABLES FOR X-RAY CRYSTALLOGRAPHY, MacGillavry, Kluwer Academic Pub
- [47] Berman, Helen M. “The protein data bank: a historical perspective.” *Acta Crystallographica Section A* 64.1 (2007): 88-95
- [48] R. M. Housley and F. Hess “Analysis of Debye-Waller-Factor and Mossbauer-Thermal-Shift Measurements. I. General Theory,” *Phys. Rev.* **146**,517526 (1966).
- [49] Bernhard Rupp *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, 1st edn. Garland Science, New York
- [50] M. Schmidt, “Mix and inject: Reaction initiation by diffusion for time-resolved macromolecular crystallography,” *Adv. Cond Matter Phys* , 167276 (2013).
- [51] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “ The protein data bank,” *Nucleic Acids Res* **28**, 235-242 (2010).
- [52] J. B. Pendry, “ Low Energy Electron Diffraction,” Academic Press, London (1974).

- [53] D.A. Varshalovich, A.N. Moskalev, V.K. Khersonskii, “Quantum Theory of Angular Momentum,” P72-P163, World Scientific Publishing Co. Pte. Ltd.(2008).
- [54] J Key, V Srajer, R Pahl, K. Moffat, “ Time-resolved crystallographic studies of the heme domain of the oxygen sensor FixL: structural dynamics of ligand rebinding and their relation to signal transduction,” *Biochemistry* **46** 4706-4715 (2007).
- [55] W. Cochran, “Some properties of the (Fo-Fc)-synthesis,” *Acta Cryst.* **4** 408-411 (1951).
- [56] P. C. Hansen, “The truncated svd as a method for regularization, *BIT Numer. Math.* **27**, 534553 (1987).
- [57] V. Granville, M. Krivanek, J. -P. Rasson, “ Simulated annealing: A proof of convergence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(6): 652656 (1994).
- [58] P. J. Rous, “A global approach to the search problem in surface crystallography by low energy electron diffraction”, *Surf. Sci.* **296**, 358 (1993).
- [59] T. E. Meyer, “Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*,” *Biochim. Biophys. Acta* **806**, 175183 (1985).
- [60] Jozef J. Van Beeumen, Bart V. Devreese, Stefaan M. Van Bun, Wouter D. Hoff, Klaas J. Hellingwerf, Terrance E. Meyer, Michael A. Cusanovich, Duncan E. Mcree, “ Primary structure of a photoactive yellow protein from the phototrophic bacterium *Ectothiorhodospira halophila*, with evidence for the mass and the binding site of the chromophore”, *Protein Sci* 2:11141125 (1993).

- [61] R. Kort, W.D. Hoff, M. van West, A.R. Kroon, S.M. Hoffer, K.H. Vlieg, W. Crielaard, J.J. van Beeumen, K.J. Hellingwerf (1996). The Xanthopsins: a new family of eubacterial blue-light photoreceptors. *EMBO J.*, **15**, 3209-3218.
- [62] Shailesh Tripathi, Vukica rajer, Namrta Purwar, Robert Henning, and Marius Schmidt, “pH Dependence of the Photoactive Yellow Protein Photocycle Investigated by Time-Resolved Crystallography,” *Biophys J* 102:325332 (2012).
- [63] K.J. Hellingwerf, J. Hendriks, Th. Gensch, “On the Configurational and Conformational Changes in Photoactive Yellow Protein that Leads to Signal Generation in *Ectothiorhodospira halophila*,” *J Biol Phys* 28:395412(2002).
- [64] C. N. Lincoln, A. E. Fitzpatrick, and J. J. van Thor, “Photoisomerisation quantum yield and non-linear cross-sections with femtosecond excitation of the photoactive yellow protein,” *Phys. Chem. Chem. Phys.* **14**, 1575215764 (2012).
- [65] I. H. van Stokkum, D. S. Larsen, and R. van Grondelle, “Global and target analysis of timeresolved spectra,” *Biochim. Biophys. Acta* 1657, 82104 (2004).
- [66] Marius Schmidt, “ A short history of structure based research on the photocycle of photoactive yellow protein”, *Structural Dynamics*, **4**, 023302 (2017).
- [67] T.W. Kim, J. H. Lee, J. Choi, K. H. Kim, L. J. van Wilderen, L. Guerin, Y. Kim, Y. O. Jung, C. Yang, J. Kim, M. Wulff, J. J. van Thor, and H. Ihee, “Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe x-ray solution scattering”, *J. Am. Chem. Soc.* **134**, pp 3145-3153 (2012).
- [68] G. E. O. Borgstahl, D. R. Williams, and E. D. Getzoff, “ 1.4 angstrom structure of photoactive yellow protein, a cytosolic photoreceptor - unusual fold, active-site, and chromophore,” *Biochemistry* **34**, 6278-6287 (1995).

- [69] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF chimera - a visualization system for exploratory research and analysis," *J Comput Chem* **25**, 1605-1612(2004).
- [70] Perspective of Imaging of Single Protein Molecules with Present Design of European XFEL, NOTKESTRASSE 85 22607 HAMBURG: ISSN 0418-9833 (2014).
- [71] K. Pande, P. Schwander, M. Schmidt and D. K. Saldin, "Deducing fast electron density changes in randomly orientated uncrystallized biomolecules in a pumpprobe-experiment," *Phil. Trans. R. Soc. B*, **369**, 20130332(2014).
- [72] Frank A. Haight, "Handbook of the Poisson Distribution," New York: John Wiley Sons (1967)
- [73] Chimera Team, University of California San Diego, "https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/pdbintro.html"
- [74] <http://www.pdb.org>
- [75] J. Hajdu, K. Hodgson, J. Miao, D. van der Spoel, R. Neutze, C.V. Robinson, G. Faigel, C. Jacobsen, J. Kirz, D. Sayre, E. Weckert, G. Materlik, A. Szke, "Structural Studies of Single Particles and Biomolecules, LCLS: The First Experiments" pp 35-62(2000) [http://www-ssrl.slac.stanford.edu/lcls/papers/lcls\\_experiments\\_2.pdf](http://www-ssrl.slac.stanford.edu/lcls/papers/lcls_experiments_2.pdf).
- [76] Abela R et al, "The European x-ray free-electron laser Technican Design Report" ed M Altarelli et al pp 401420(2007) [http://xfel.desy.de/tdr/index\\_eng.html](http://xfel.desy.de/tdr/index_eng.html)

# Appendices

# Appendix A

## Protein Data Bank Format

The purpose of this appendix is to show the details of the pdb file.

Table A.1: Explanation of the format of pdb file [73]

Protein Data Bank Format: Coordinate Section				
Record Type	Columns	Data	Justification	Data Type
ATOM	1-4	ATOM		character
	7-11	Atom serial number	right	integer
	13-16	Atom name	left*	character
	17	Alternate location indicator		character
	18-20	Residue name	right	character
	22	Chain identifier		character
	23-26	Residue sequence number	right	integer
	27	Code for insertions of residues		character
	31-38	X orthogonal coordinate	right	real (8.3)
	39-46	Y orthogonal coordinate	right	real (8.3)
	47-54	Z orthogonal coordinate	right	real (8.3)
	55-60	Occupancy	right	real (6.2)
	61-66	Temperature factor	right	real (6.2)
	73-76	Segment identifier	left	character
	77-78	Element symbol	right	character
79-80	Charge		character	
HETATM	1-6	HETATM		character
	7-80	same as ATOM records		
TER	1-3	TER		character
	7-11	Serial number	right	integer
	18-20	Residue name	right	character
	22	Chain identifier		character
	23-26	Residue sequence number	right	integer
	27	Code for insertions of residues		character

# Appendix B

## Pseudocode

The purpose of this appendix is to show the major steps taken to develop the algorithm for data analysis.

### For Rapid Direct Method

1. Input as Simulated SAXS/WAXS Intensity for Unperturbed( $u$ ) and Perturbed Structure( $p$ ) i.e.  $I_{u,p}(q)$ ,  $q = 0, 1, 2, \dots, 33$   $q$  points.
2. Difference SAXS/WAXS Intensity i.e.  $I_p(q) - I_u(q)$ ,  $q = 0, 1, 2, \dots, 33$   $q$  points.
3. Spherical Bessel function  $j_l$  with argument  $(qr_k)$ ,  $l = 0, 1, 2, \dots, 30$ ,  $q = 0, 1, 2, \dots, 33$ ,  $k = 1, 2, \dots, 1331$ .
4. Structure factor components  $F_{lm}(q)$  in terms of spherical harmonic basis and its complex conjugate,  $q = 0, 1, 2, \dots, 33$ ,  $l = 0, 1, 2, \dots, 30$ ,  $m = -30$  to  $30$ .
5. Spherical harmonics  $Y_{lm}(\hat{\mathbf{r}}_k)$  and its complex conjugate,  $l = 0, 1, 2, \dots, 30$ ,  $m = -30$  to  $30$ .
6. Formation of Matrix mainly with 3., 4., and 5.
7. Pseudo-inverse of Matrix with Singular Value Decomposition(SVD) method.

8. Premultiplication of the difference SAXS/WAXS Intensity by the elements of the inverse matrix.
9. Output as difference electron density.

### **For Simulated Annealing(SA) Method**

1. Input as perturbed(experimental) SAXS/WAXS Intensity.
2. Input as unperturbed SAXS/WAXS intensity(model intensity).
3. Minimization of Cost Function  $[I_{expt} - I_{the}(R_j)]^2$  by varying coordinates of the atoms  $R_j$  of the dark structure residue identified by rapid direct method.
4. Output coordinates from Simulated Annealing.
5. Electron density for Perturbed Structure with the coordinates recovered from SA for perturbed residue.
6. Output as difference electron density.

### **For Poisson Noise Consideration**

1. Conversion of unperturbed and perturbed SAXS/WAXS Intensity to experimentally realizable intensity.
2. Introduction of shot noise.
3. Difference in SAXS/WAXS intensity.
4. Then same as the steps to obtain difference electron density from rapid direct method.

# CURRICULUM VITAE

## **Prakash Nepal**

**Place of birth:** Putalibazar Municipality, Syangja, Nepal

**Birthday:** April 18

### **EDUCATION**

1. **PhD in Physics:** University of Wisconsin Milwaukee (2011-2017)

Dissertation Title : Difference Structures from Time-resolved SAXS / WAXS

Dissertation Defense Date : August 02, 2017

2. **MSc in Physics:** Tribhuvan University, Nepal (2003-2006)

Research Project Title : A study of the effective potentials for time-like circular orbits in Schwarzschild de-Sitter space-time

### **RESEARCH INTERESTS**

Computational and Theoretical Physics, Data Analysis Techniques, Any Aspects of Simulation Related to Physics, Experimental Data Interpretation

### **RESEARCH EXPERIENCE**

**Graduate Research Assistant,** Department of Physics, UWM, Milwaukee, Wisconsin (2012 - 2017)

### **HONORS AND AWARDS**

**BioXFEL Mac Gyver Award** for Winning Proposal Writing Competition at BioXFEL STC 2<sup>nd</sup> Annual Site Visit, June 3-5, 2015

**Best Poster Prize Award** for Poster Competition at 4<sup>th</sup> BioXFEL International Conference, Las Vegas, NV, January 13 - 15, 2017

**Chancellor's Graduate Student Award(CGSA):** Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013, Spring 2014, Fall 2015, Spring 2016, Fall 2016, Spring 2017

**Physics Graduate Student (DIN) Award :** Fall 2011, Spring 2012

**Research Excellence Award:** Fall 2012, Spring 2013, Fall 2013, Spring 2014, Fall 2015, Spring 2016, Spring 2017

## **PRESENTATION AND INVITED TALKS**

**BioXFEL Journal Club Seminar,** (Article Presentation, Link: <https://www.youtube.com/watch?v=y69lthls36o>), 20 May, 2015

**BioXFEL NSF STC 2<sup>nd</sup> Annual Site Visit,** “ Research in Saldin Group ” (Brief INVITATION TALK in Front of NSF Distinguished Delegates),Buffalo, NY, June 3-5, 2015

**UWM Biophysics Seminar,** “ Extraction of Difference Electron Densities from Time-Resolved Pump-Probe XFEL SAXS/WAXS ” (TALK), November 5, 2015

**UWM Research Foundation Lunch & Learn**, “ Difference Structures from Time-Resolved SAXS/WAXS ” (INVITED TALK), September 28, 2016

**UWM Biophysics Seminar**, “ Protein Structural Dynamics Observed with Time-Resolved SAXS/WAXS ” (TALK), April 13, 2017

## **PUBLICATIONS**

R. O’Shaughnessy , **P. Nepal** , A. Lundgren “ A semi analytic Fisher Matrix for precessing binaries for single significant spin ” (arxiv: 1509.06581)

**P. Nepal** , D.K. Saldin “Difference structures from time-resolved small angle x-ray scattering(SAXS)/wide-angle x-ray scattering(WAXS) ” recommended for publication by referees in Physical Review B

## **ORGANIZATIONS**

Life member of Nepal Physical Society(NPS)

BioXFEL Scholar

## **TEACHING EXPERIENCE**

**Physics Lecturer**: Capital College and Research Center, Kathmandu, Nepal (2007-2011)

**Physics 123 (General Physics Lab II, Non-Calculus Treatment), Lab Instructor:** Department of Physics, University of Wisconsin Milwaukee (Fall 2011 and Spring 2012)

**Physics 121 (General Physics Lab I, Non-Calculus Treatment), Lab Instructor:** Department of Physics, University of Wisconsin Milwaukee (Fall 2012)

**Physics 122 (General Physics II, Non-Calculus Treatment), Discussion TA:** Department of Physics, University of Wisconsin Milwaukee (Fall 2013, Spring 2014, Fall 2014 (**Highest Discussion Evaluation**), Spring 2015, Spring 2016)

**Physics 210 (Physics II, Calculus Treatment), Discussion TA:** Department of Physics, University of Wisconsin Milwaukee (Summer 2014)

**Physics 120 (General Physics I, Non-Calculus Treatment), Discussion TA:** Department of Physics, University of Wisconsin Milwaukee (Fall 2015)

**Physics 209 (Physics I, Calculus Treatment), Discussion TA:** Department of Physics, University of Wisconsin Milwaukee (Spring 2017)

**Physics 531 (Graduate Quantum Mechanics I), Grader:** Department of Physics, University of Wisconsin Milwaukee (Fall 2014)

**Physics 532 (Graduate Quantum Mechanics II), Grader:** Department of Physics, University of Wisconsin Milwaukee (Spring 2015)

**Physics 711 (Graduate Theoretical Physics - Dynamics), Grader :** Department of Physics, University of Wisconsin Milwaukee ( Spring 2015)

**CONFERENCES, SEMINARS, WORKSHOPS, LCLS BEAM TIME ATTENDED**

**23rd Midwest Relativity Meeting in Milwaukee,** 25-27 October 2013

**BioXFEL STC 2<sup>nd</sup> Annual International Conference** Puerto Rico, 13-15 January 2015 (Presented Poster)

**Crystal Growth Workshop** Buffalo, NY, 2-3 June 2015

**BioXFEL STC 2<sup>nd</sup> Annual Site Visit** Buffalo, NY, 3-5 June 2015 (Presented Poster)

**UWM BioXFEL Workshop** Milwaukee, WI, 29 September 2015

**LCLS BEAM TIME FOR SPI EXPERIMENT** Menlo Park, CA, 4-5 December 2015 (Participated for Data Analysis and Algorithm Development)

**BioXFEL STC 3<sup>rd</sup> Annual International Conference** Puerto Rico, 13-15 January 2016 ( Presented Poster)

**LCLS BEAM TIME FOR SPI EXPERIMENT** Menlo Park, CA, 31 March - 02 April 2016 (Participated for Data Analysis and Algorithm Development)

**XFEL Lightsource Applications Seminar** Buffalo, NY, 20 June 2016

**BioXFEL STC 3<sup>rd</sup> Annual Site Visit** Buffalo, NY, June 20-23, 2016

**SSRL/LCLS Annual Users' Meeting, Workshops and BioXFEL Session**  
Menlo Park, CA, 5-8 October 2016 (Presented Poster)

**Single Particle Workshop** Las Vegas, NV, 9 January, 2017

**BioXFEL STC 4<sup>th</sup> Annual International Conference** Las Vegas, NV, 10-12  
January, 2017 (Presented Poster)

**BioXFEL STC 4<sup>th</sup> Annual Site Visit** Buffalo, NY, 22-24 May, 2017

**BioXFEL Workshop on Graduate Research Fellowship Proposal Writing**  
Buffalo, NY, 25 May, 2017