Femtosecond X-ray Nanocrystallography of Membrane Proteins

by

Mark Hunter

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

Approved June 2011 by the
Graduate Supervisory Committee:

Petra Fromme, Chair
Marcia Levitus
George Wolf

ARIZONA STATE UNIVERSITY

August 2011

ABSTRACT

Membrane proteins are very important for all living cells, being involved in respiration, photosynthesis, cellular uptake and signal transduction, amongst other vital functions. However, less than 300 unique membrane protein structures have been determined to date, often due to difficulties associated with the growth of sufficiently large and well-ordered crystals. This work has been focused on showing the first proof of concept for using membrane protein nanocrystals and microcrystals for high-resolution structure determination. Upon determining that crystals of the membrane protein Photosystem I, which is the largest and most complex membrane protein crystallized to date, exist with only a hundred unit cells with sizes of less than 200 nm on an edge, work was done to develop a technique that could exploit the growth of the Photosystem I nanocrystals and microcrystals. Femtosecond X-ray protein nanocrystallography was developed for use at the first high-energy X-ray free electron laser, the LCLS at SLAC National Accelerator Laboratory, in which a liquid jet would bring fully-hydrated Photosystem I nanocrystals into the interaction region of the pulsed X-ray source. Diffraction patterns were recorded from millions of individual PSI nanocrystals and data from thousands of different, randomly oriented crystallites were integrated using Monte Carlo integration of the peak intensities. The short pulses ($\sim$ 70 fs) provided by the LCLS allowed the possibility to collect the diffraction data before the onset of radiation damage, exploiting the diffract-before-destroy principle. At the initial experiments at the AMO beamline using 6.9-Å wavelength, Bragg peaks were recorded to 8.5-Å resolution, and an electron-density map was determined that did not show any effects of X-ray-induced radiation damage. Recently, femtosecond X-ray protein nanocrystallography experiments were done at the CXI beamline of the LCLS using 1.3-Å wavelength, and Bragg reflections were recorded to 3-Å resolution; the data are currently being processed. Many additional techniques still need to be developed to explore the femtosecond nanocrystallography technique for experimental phasing and

time-resolved X-ray crystallography experiments. The first proof-of-principle results for the femtosecond nanocrystallography technique indicate the incredible potential of the technique to offer a new route to the structure determination of membrane proteins.

DEDICATION

I dedicate this dissertation to my parents, Charles J. Hunter and Elaine S. Hunter

ACKNOWLEDGEMENTS

I am very grateful to my supervisor, Dr. Petra Fromme, for the years of excellent insight into a complicated subject matter. The enthusiasm which Petra exhibited on a daily basis was contagious and led to a greater appreciation of the biophysical sciences. She has shown me what it means to be a true scientist.

I am grateful for the support and advice of my committee members, Dr. George Wolf and Dr. Marcia Levitus. Dr. Wolf and Dr. Levitus taught me physical chemistry and biophysics courses that helped spur my interest to pursue an advanced degree in chemistry, and I would like to thank them for instilling general intellectual curiosity about the physical sciences in me.

I would like to thank Ingo Grotjohann for help with the protein preparations, the cell growth, and many thoughtful discussions regarding biochemistry lab techniques and protein crystallization. I would like to thank Raimund Fromme for collecting data sets, analyzing the diffraction data for crystals I grew, and offering valuable discussions on X-ray crystallography. I would like to thank Chris Kupitz for his assistance in the higher-energy LCLS experiments and PSI re-crystallization experiments and for being a dependable friend. It was a pleasure to teach Chris the intricacies of protein crystallization and femtosecond nanocrystallography. I would like to thank Jim Zook, also, for discussions of liquid chromatography and light scattering, as well as for being a good friend.

The femtosecond X-ray protein nanocrystallography experiments required the monumental effort of many members from the Physics Department at Arizona State University. John C.H. Spence has been a pleasure to work with, and he has really impressed upon me what it means to have enthusiasm for a topic that is challenging. The tireless work of John has left an invaluable impression on me, and it was an honor to work with him, so I would like to thank John for teaching me so much about so many different

areas of science. Uwe Weierstall taught me a great deal about professionalism and attention to detail, so I would like to thank him for passing these attributes to me. I would like to thank Bruce Doak for always offering thoughtful insight into my ideas and my work. It was always a pleasure to work with Rick Kirian, as I was able to have many discussion with Rick about theoretical aspects of our project. I would also like to thank Rick for his excellent analysis of the femtosecond nanocrystallography data, as well as his creative approaches to solving difficult problems in our research, in addition to being a good friend. I would like to thank XiaoYu Wang for providing theoretical discussions of X-ray diffraction with me and for passing on valuable information. Dingjie Wang is relatively new to the project, but I would like to thank Dingjie for his help on the higher-energy LCLS work. I would like to thank Nadia Zatsepin for helping me with my dissertation and being a good friend.

A large international collaboration was formed to work on the femtosecond X-ray protein nanocrystallography. I would like to thank all authors on the 2011 Nature letter for their help. Specifically, I would like to thank Henry Chapman, Mike Bogan, Ilme Schlichting, Robert Shoeman, Richard Neutze, Anton Barty, Tom White, Andy Aquila, Andrew Martin, Francesco Stellato, Karol Nass, Dan DePonte, and Janos Hajdu. I would like to thank Mike Bogan for always allowing us unobstructed use of his laboratory at the Pulse Institute at Stanford, located at SLAC national accelerator laboratory, as well as for the excellent discussions regarding my work. I would like to thank Robert Shoeman for his help in sample loading and delivery, and for his innovative solutions to our problems, many of which he formed under intense circumstances and without much time. I would like to thank Dan DePonte for making the gas-dynamic virtual nozzles, and for lots of discussions about the project and physics in general, as well as preparing me for the real world with his no-nonsense attitude. I would like to thank Fransceco Stellato and Karol Nass for providing spirited discussions regarding sample preparation and characterization and for suggesting ways of doing data analysis

on the femtosecond nanocrystallography data. I would like to thank Tom White, as he wrote the programs that will become the standard for the analysis of femtosecond nanocrystallography data, and for his patience in discussing crystallography and various aspects of the project with me. I would like to thank Andy Aquila for his expertise in "piloting" the experiments; only upon trying to replace him for a few shifts of the experiments did I realize how thankless of a job it is. Additionally, I would like to thank Andy for general discussions about optics in relation to our project, and for discussions about life after graduate school and the postdoctoral process. I would like to thank Andrew Martin for his work on the crystal reconstructions from the shape transforms in the femtosecond nanocrystallography data as well as for discussions pertaining to coherent diffractive imaging. I would like to thank Janos Hajdu for his enthusiasm and hard work in making the XFELs available to the biological community; in addition, it was always a pleasure to have discussions about the future of biological imaging with Janos. I would like to thank Ilme Schlichting for providing many of the necessary resources and innovative thoughts to allow the project to come to fruition, and I would like to thank her for discussions regarding the experiments and subject matter that were matter-of-fact; it was always very enjoyable to have discussions with Ilme. I would like to thank Henry Chapman for bringing together groups of such varied backgrounds and for providing the leadership and insight in order to allow this project to be successful. Henry was a pleasure to work with, and I greatly enjoyed theoretical and practical discussions with him regarding the project. I could always count on Henry for creative ideas and insight at a moment's notice.

I want to thank David Shapiro for his dedicated work with the powder diffraction experiments at beamline 9.0.1 of the Advanced Light Source and for his quick analysis of our data. I thank the staff of the LCLS for their support in carrying out these experiments, and David van der Spoel for providing computational resources. Specifically, I would like to thank John Bozek and Christoph Bostedt for helping us get the maximum

value from our AMO experiments, as well as Sebastien Boutet, Garth Williams, and M. Marvin Seibert, for working with us through the commissioning of the CXI beamline at the LCLS and helping us with appropriate strategies for data collection. I also want to thank James Holton for his help with data analysis and electron-density map generation as well as his willingness to discuss X-ray crystallography with me at a moment's notice.

Experiments were carried out at the Linac Coherent Light Source national user facilities operated by Stanford University, as well as the Advanced Light Source user facilities operated by the University of California Berkeley, on behalf of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences.

This work was supported by NSF award IDBR 0555845, the Center for Biophotonics Science and Technology (University of California at Davis), the Lawrence Berkeley National Laboratory Seaborg Fellowship award, by the U.S. Department of Energy through the PULSE Institute at the SLAC National Accelerator Laboratory, and by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, and the Joachim Herz Stiftung.

I acknowledge support from the Helmholtz Association; the Max Planck Society, for funding the development and operation of the CAMP instrument within the ASG at CFEL; DOE, through the PULSE Institute at the SLAC National Accelerator Laboratory, and by the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344; the US National Science Foundation (awards 0417142 and MCB-1021557); the US National Institutes of Health (awards 1R01GM095583-01 (ROADMAP) and 1U54GM094625-01 (PSI:Biology)); the Joachim Herz Stiftung, the Swedish Research Council; the Swedish Foundation for International Cooperation in Research and Higher Education, Stiftelsen Olle Engkvist Byggmastare.

The Advanced Light Source is supported by the Director, Office of Science, Office of

Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

I would like to thank my family and friends, especially my parents, for their patience as I matured and I would like to thank my sister, Sara, and brother, Justin, for making my life more interesting and fruitful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Structural biology and the quest to solve life's mysteries

The structure and function of molecules are intimately related at the atomic and molecular levels. The desire to determine the structure of matter in order to possibly infer the mechanisms and pathways has been a large motivational force in the disciplines of physics, chemistry, and biology. For instance, the structures of inorganic and organic molecules offer insight into catalysis and reaction pathways, and the information can be used to elucidate the action of drugs, or the potential environmental impact of an organometallic molecule, as two examples.

The biomacromolecules are the facilitators of life. Thereby the structure determination of biomacromolecules is one important clue to help understand the complexity observed in life. Whether the desire is to understand cellular respiration, nutrient uptake and transport by a cell, or various other cellular functions, one focus is to determine high-resolution structures of the molecules involved. The structure forms the basis to elucidate the reaction mechanisms and understand how the structure relates to the function and the dynamics of the molecules. To date, over 60,000 protein structures have been solved by X-ray crystallography, electron microscopy, and nuclear magnetic resonance (NMR), with the majority of the structures determined by protein X-ray crystallography.

The impact of structural biology on the biological sciences becomes tangible when noting the examples of success for structure determination. One area of study that has seen considerable success through the use of structural biology—and is of primary focus in the presented research—is oxygenic photosynthesis.

1.2 Oxygenic photosynthesis

Overview of oxygenic photosynthesis

Oxygenic photosynthesis, as performed by plants, algae, and cyanobacteria, is the most important biological process on earth. Oxygenic photosynthesis converts light energy into chemical energy and involves multiple electron-transfer reactions and the function of many pigment-protein complexes with various cofactors. The major protein complexes involved in oxygenic photosynthesis as well as the orientation of the protein complexes within the thylakoid membrane is represented in Fig. 1.1A. The protein complexes are represented by models derived from the high-resolution electron-density maps as determined by X-ray crystallography. Fig. 1.1B shows the cofactor arrangement involved in the electron-transfer reactions and will be discussed below and in subsequent subsections.

The electron-transfer reactions of oxygenic photosynthesis initiate at Photosystem II (PSII). PSII captures light with an internal chlorophyll system, and the light energy is used to induce charge separation at the center of the complex. Four electrons are extracted in four subsequent charge-separation steps, with concurrent oxidation of two molecules of water to form molecular oxygen, four protons, and four electrons at the oxygen-evolving complex (OEC) of PSII. The electrons ultimately reduce a mobile plastoquinone (PQ), which after being reduced twice will bind two protons to form plastoquinole (PQH_2) and is released into the thylakoid membrane. The PQH_2 serves as a mobile electron and proton carrier and is constantly exchanged with PQ in the membrane (Loll et al., 2005).

The PQH_2 will diffuse within the membrane plane and dock to the binding pocket of the cytochrome b_6f complex. The PQH_2 transfers the two electrons and two protons to the cytochrome b_6f complex, which subsequently releases two protons to the interior of the thylakoid (the lumen) and subsequently reduces two molecules of

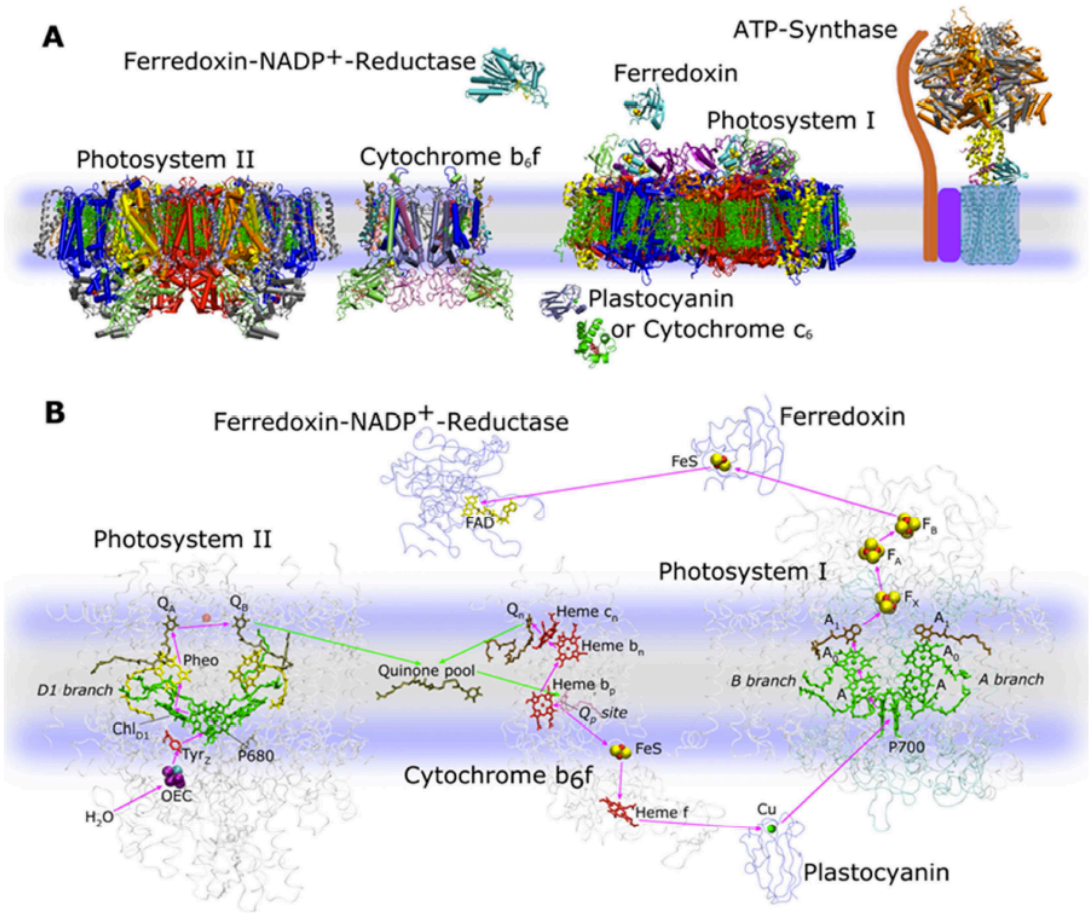

Figure 1.1: **Overview of oxygenic photosynthesis** Image showing the major proteins involved in oxygenic photosynthesis with cartoon models of the structures (A) with emphasis on the cofactors that take part in the electron transfer chain from Photosystem II through ferredoxin-NADP$^+$-reductase (B). The proteins are contained within the thylakoid membranes of photosynthetic organisms with the lumen below the membrane and the stroma above the membrane in the images. Images taken from (Fromme, 2008).

plastocyanin (PC) or two molecules of cytochrome c_6. In addition, the cytochrome b_6f complex will transfer another proton to the interior of the thylakoid membrane in a process known as the Q-cycle, furthering the development of an electrochemical gradient across the thylakoid membrane (Kurisu et al., 2003).

The availability of cytochrome c_6 or PC depends on the organism and the iron content of the growth media. Plants solely use the copper-containing plastocyanin, whereas some cyanobacteria have the genes for both soluble electron carriers and will produce plastocyanin in iron-deficient conditions (Fromme et al., 2003).

Both cytochrome c_6 and PC are one-electron soluble electron carriers that transfer electrons from the cytochrome b_6f complex to Photosystem I (PSI) . PSI uses light energy to perform the charge-separation reaction that transfers the electrons from the luminal side to the stromal side of the thylakoid membrane. Upon reaching the stromal side, the electron is used to reduce the one-electron carrier ferredoxin (Fd). The reduced ferredoxin undocks from Photosystem I and diffuses in the cytoplasm until binding to ferredoxin: $NADP^+$-reductase (FNR). Two electrons are required to reduce $NADP^+$ to NADPH; consequently two reduced Fd must sequentially bind to FNR (Setif and Bottin, 1995).

The electrochemical gradient generated across the thylakoid membrane is used in the production of ATP from ADP and inorganic phosphate by the ATP-synthase protein. The ATP and NADPH are used in the subsequent dark reactions of photosynthesis (the reactions that do not directly depend on light energy) in the production of carbohydrates and other biomacromolecules. The dark reactions are outside the scope of the work presented here and will therefore not be discussed in detail.

The main proteins from oxygenic photosynthesis used in the work presented here are Photosystem I, Photosystem II, and ferredoxin. Much work has been devoted to the structure and dynamics of Photosystem I and Photosystem II.

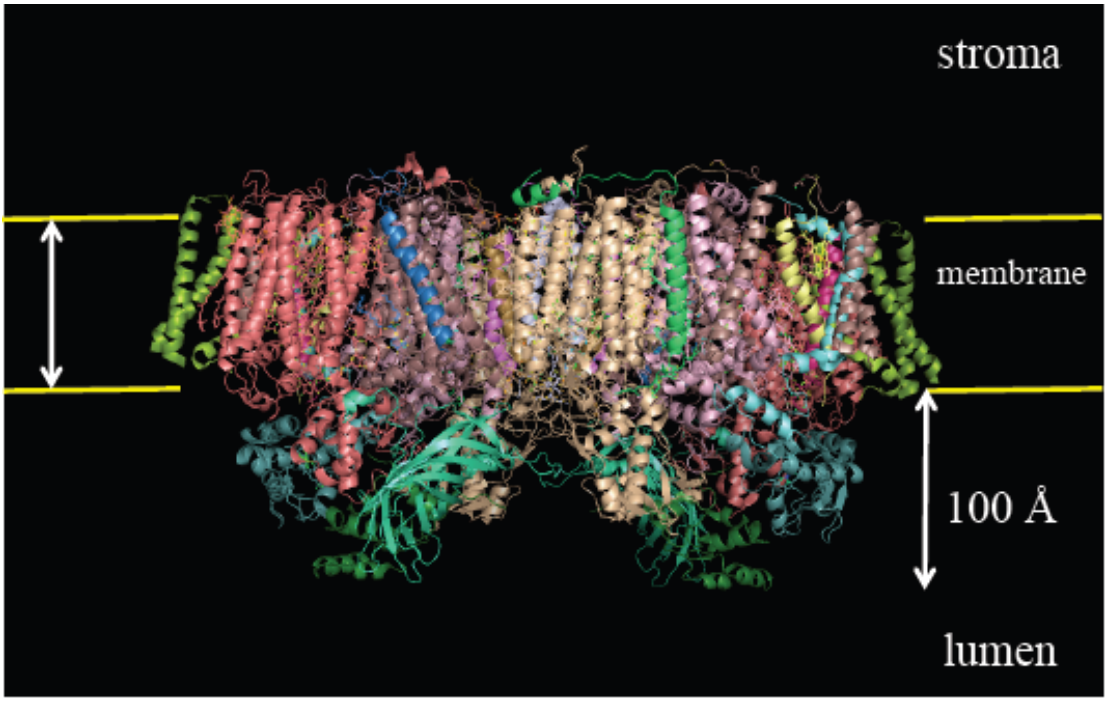

Figure 1.2: **Structure of Photosystem II** Image of Photosystem II (pdb code: 2AXT) along the membrane showing the dimeric structure (Loll et al., 2005). The luminal hump, extending approximately 100 Å into the lumen, contains the oxygen-evolving complex.

Structure of Photosystem II

The function of the large PSII complex is to capture the light from the sun and to subsequently use the light energy to generate charge-separation across the membrane. Photosystem II subsequently oxidizes water, with two molecules of water producing four protons, four electrons, and one molecule of molecular oxygen. Photosystem II is the only system in nature capable of forming molecular oxygen from water and sunlight, and it provides all the oxygen in the atmosphere, sustaining the biosphere and remaining responsible for the large complexity and diversity of the organisms on earth that depend on the O_2 for energy supply in the form of respiration.

PSII is a large membrane protein complex, consisting of 20 protein subunits and more than 50 cofactors. PSII exists in solution and in the crystals as a dimer, and the

structure was solved to 3.8-Å resolution in 2001 (Zouni et al., 2001). A 3.0- structure of PSII with more complete cofactor assignment was determined in 2005 (Loll et al., 2005) and a model based upon the 3.0-Å structure (pdb code: 2-AXT) is shown in Fig. 1.2. Recently, a 1.9-Å resolution structure of PSII was published (Umena et al., 2011). Each monomer has four large, membrane-intrinsic subunits that contain the electron transfer chain and most of the antenna system: D1, D2, CP43, and CP47; the D1 and CP43 subunits provide ligands for the oxygen-evolving complex. Additionally, 13 membrane-intrinsic, low-molecular-mass protein subunits surround the core subunits.

Each monomer has 36 transmembrane helices, with 10 transmembrane helices being part of the D1 and D2 subunits. An additional 12 transmembrane helices are contained within the CP43 and CP47 subunits. A large, membrane-extrinsic region of the protein is comprised of extra-membrane loops regions of D1, D2, CP43, and CP47 as well as three luminal membrane-extrinsic subunits, PsbO, PsbU, and PsbV and houses the oxygen-evolving complex (Umena et al., 2011).

The electron-transport chain and S-cycle of PSII

The electron transfer chain of PSII is, in principle, comprised of four chlorophyll *a* molecules, two pheophytin molecules, and two plastoquinones, of which one plastoquinone is immobile and the other is a mobile electron and proton carrier, as shown schematically in Fig. 1.3. However, only one branch of the electron transfer chain is active, the D1 branch. Charge separation occurs at a quartet of chlorophylls in the D1 and D2 proteins, after which the electron is transferred to the pheophytin on the D1 side of the molecule, with subsequent transfer to PQ_A, the immobile plastoquinone, and finally transfer to PQ_B, the mobile plastoquinone. Once PQ_B is doubly reduced and additionally binds two protons forming PQ_BH_2, the plastoquinol is released into the membrane. The initially oxidized $P_{680}{}^+$ formed during charge separation is reduced by an electron donated from a tyrosine residue located between P_{680} and the catalytic

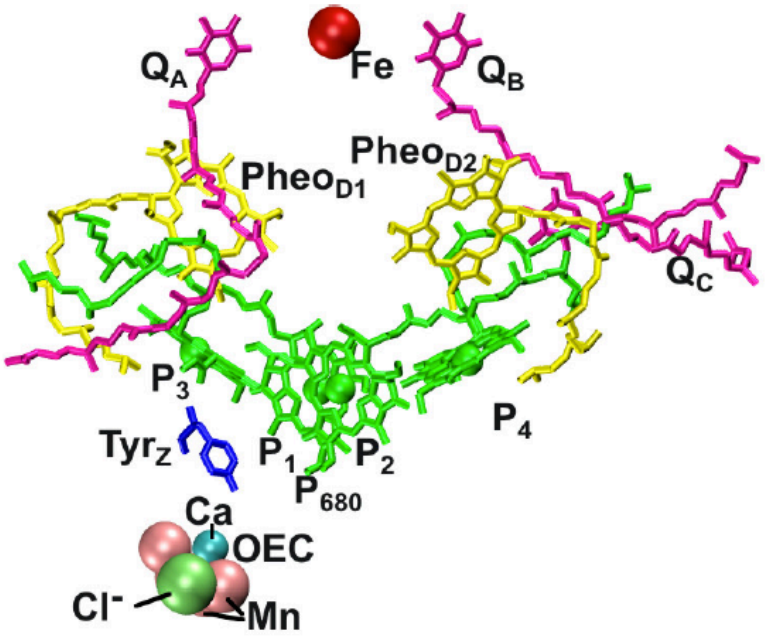

Figure 1.3: **Electron transport chain of Photosystem II** Schematic showing the cofactors involved in the electron transport chain of Photosystem II. P_{680} and $Chl_{D1,2}$ are chlorophyll *a* molecules, $Pheo_{D1,2}$ are pheophytin molecules, and PQ_A and PQ_B are plastoquinones. The Mn_4Ca cluster oxidizes water. The D1 branch is the only active branch.

center of the oxygen-evolving complex (Zouni et al., 2001).

The oxidation of water that occurs in Photosystem II is catalyzed by the oxygen-evolving complex (OEC). The catalytic center of the OEC consists of a Mn_4Ca-Cl metal center that couples the four-electron chemistry of water oxidation to the one-electron photochemistry of the reaction center. The coupling is accomplished through the sequential storing of oxidization states by the manganese in the series of the S-states, S_0 to S_4. Each light-driven charge-separation event in Photosystem II (P_{680}) extracts one electron from the OEC, leading to the release of one oxygen molecule after four charge-separation events, which returns the system from the S4 state to the S0 state, whereby two substrate water molecules are bound to the metal cluster (Yano and Yachandra, 2008).

The mechanism of proton release from the OEC during the S cycle is still de-

bated. The H-atom abstraction model proposed by Babcock supports a mechanism where one electron and one proton are extracted from the OEC in each step of the S cycle (Hoganson and Babcock, 1997; Tommos and Babcock, 1998). However, another mechanism proposes that the proton release pattern is 1:0:2:1 (S_0 S_1 S_2 S_3 S_4) and is consistent with electrochromic measurements (Messinger, 2004).

Structure of Photosystem I

Photosystem I (PSI) is a large membrane-intrinsic protein complex that is responsible for light-induced charge-separation that transfers an electron from the luminal side to the stromal side of the thylakoid membrane. The electron-transfer catalyzed by PSI provides the electrons utilized in reducing $NADP^+$ to NADPH in oxygenic photosynthesis.

PSI exists in a number of oligomeric forms in nature; in plants, PSI exists as a monomer surrounded by four antenna proteins LHC-(I-IV), whereas in cyanobacteria, PSI exists predominantly as a trimer, but the monomer can be found when cells are grown under high-light conditions, and the trimer-to-monomer ratio depends on the light conditions. Cyanobacterial PSI is the largest and most complex membrane-protein complex to be crystallized (Fromme and Witt, 1998) and to have its structure solved to high resolution (Jordan et al., 2001).

The trimeric PSI, as shown in Fig. 1.4, has a mass of 1,056,000 Da, with each monomer consisting of 12 protein subunits and 127 non-covalently bound cofactors. Trimeric PSI has a diameter of approximately 220 Å with a stromal hump that jettisons 40 Å into the cytoplasm. The cofactor composition in the monomer is 96 chlorophyll molecules, 22 carotenoid molecules, three 4Fe-4S clusters, three lipids, two phylloquinone molecules, and one Ca^{2+} ion (Jordan et al., 2001).

The PsaA and PsaB subunits form a large heterodimer at the center of the PSI molecule, which acts as the joint reaction center and core antenna complex. Each

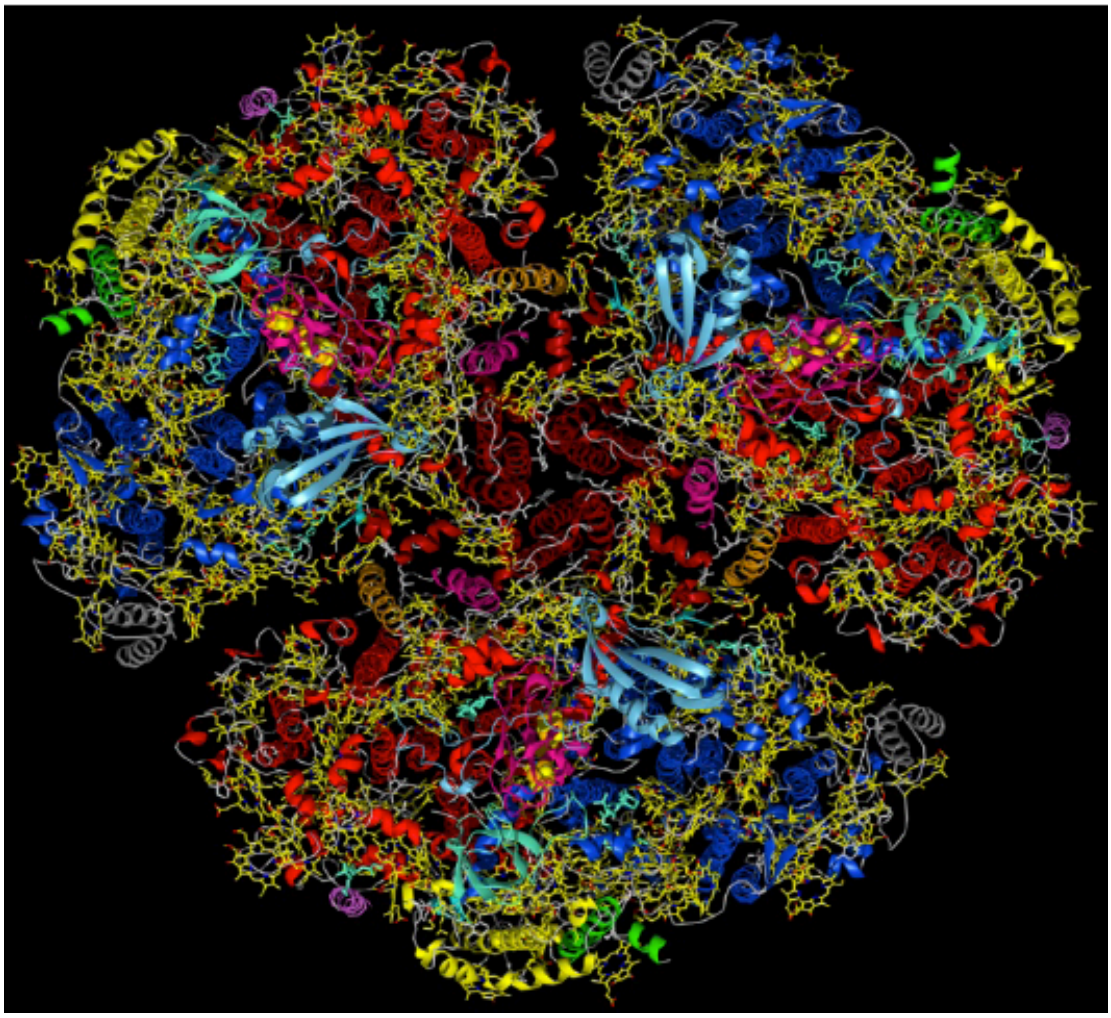

Figure 1.4: **Structure of cyanobacterial Photosystem I** Membrane normal view of the trimeric Photosystem I from cyanobacteria (pdb code: 1JB0)at 2.5-/AA resolution (Jordan et al., 2001). Cyanobacterial Photosystem I is the largest membrane protein complex solved to molecular resolution.

subunit has 11 transmembrane helices and in total PsaA and PsaB coordinate 79 of the 96 antenna chlorophylls, and most of the carotenoids show hydrophobic interaction with PsaA or PsaB (Jordan et al., 2001). PsaA and PsaB also coordinate the majority of the cofactors of the electron transport chain, containing P_{700}, the electron acceptors A, A_0, and A_1, as well as the first iron-sulfur cluster F_X.

The PsaA and PsaB subunits are surrounded by seven smaller protein subunits in the membrane: PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM, and PsaX. PsaI, PsaL, and

PsaM are involved in the trimerization of PSI, whereas PsaF, PsaK, PsaM, and PsaX cluster in the membrane exposed portion of each monomer. While PsaF is involved in the docking of plastocyanin in plants, PsaF is not involved in plastocyanin docking in cyanobacteria (Fischer et al., 1998).

The stromal hump of PSI is comprised of three protein subunits: PsaC, PsaD, and PsaE. PsaC contains the iron-sulfur clusters F_A and F_B of the electron transfer chain, and PsaD has been shown to be necessary for the stable assembly of PsaC, and may be the location of ferredoxin binding (Fischer et al., 1998).

The electron transfer chain of PSI

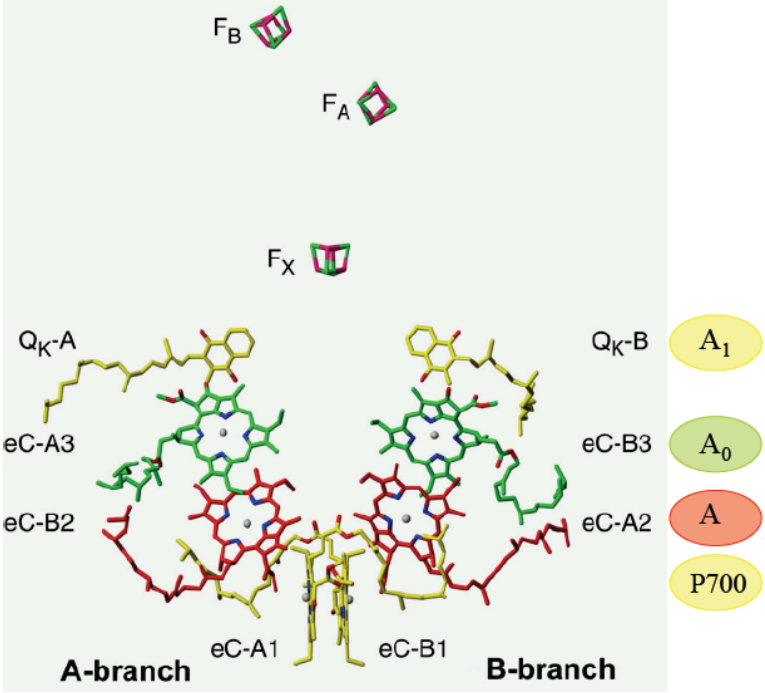

Figure 1.5: **Overview of the electron transport chain of Photosystem I** Schematic showing the general positions and cofactors involved in the electron transport chain of Photosystem I (Jordan et al., 2001). P_{700} is a heterodimer of one chlorophyll *a* molecule and one chlorophyll *a*' molecule, A and A_0 are chlorophyll *a* molecules, A_1 are phylloquinone molecules, and F_x, F_A, F_B are 4Fe-4S clusters.

The electron transfer chain of PSI is shown in Fig. 1.5 and consists of six molecules of chlorophyll, 2 phylloquinones, and three Fe_4-S_4 clusters. Light energy is

used to induce charge separation at a pair of chlorophyll molecules (P_{700}), called the special pair, at the center of PSI. From the electronically excited state of the special pair, $P_{700}{}^{*}$, the electron is transferred to A, a chlorophyll *a* molecule, to produce $P_{700}{}^{+}$ and A^{-}. The electron is then transferred to A_0, another chlorophyll *a* molecule, followed by A_1, a phylloquinone, and subsequently to the iron-sulfur clusters F_x, F_A, and finally to the terminal iron-sulfur cluster of PSI, F_B. From F_B, the electron is transferred to ferredoxin (Fromme et al., 2003). However, recent results indicate that that charge separation may initiate at the chlorophyll *a* molecule A, and not P_{700}, with the positive charge residing on the P_{700} after quick reduction of A (Muller et al., 2010).

The reduction of ferredoxin by Photosystem I

Ultimately, ferredoxin (Fd) is the terminal electron acceptor for PSI in oxygenic photosynthesis, except in iron-deplete conditions, in which case flavodix will replace ferredoxin. Upon determining the first low-resolution structure of PSI at 6 Å binding pockets for ferredoxin were proposed based upon the charge-density maps of PSI. Based upon the 6-Å structure of PSI (Krauss et al., 1993), a binding pocket was proposed that had the ferredoxin close to the terminal 4Fe-4S cluster of PSI (Fromme et al., 1994).

Mutagenesis studies were done in order to ascertain which subunits were involved in the binding of ferredoxin to PSI, with results implicating all subunits in the stromal hump of PSI; PsaC, PsaD, and PsaE might be involved in ferredoxin docking (Fischer et al., 1998). The mutagenesis studies in combination with the low-resolution model of PSI indicate that PsaD and PsaE may facilitate the binding of ferredoxin while the 4Fe-4S of PsaC provides the electrons. However, PsaE can be deleted without strong impairment of the PSI-Fd electron transfer reaction, which contradicts other results showing that a single point mutation in PsaE docks ferredoxin permanently to PSI (Rousseau et al., 1993). The contradictory results may indicate that the PSI-Fd interaction is dynamic.

In addition to structural models and studies on the PSI-ferredoxin interaction, experiments were done to study the kinetics of the interaction. The kinetics of the reduction of ferredoxin by PSI have been studies using flash spectroscopy (Setif and Bottin, 1994, 1995) . Analysis of the absorption transients indicates three first-order phases with $t_{1/2} < 1$ μs, 13-20 μs, and 103-123 μs. The spectroscopic data is interpreted such that the fast first-order phase is the reduction of ferredoxin docked in close proximity to F_A^- or F_B^-, whereas the intermediate first-order phase may correspond to a different, more distal, binding site for ferredoxin. The slowest phase is dependent on the concentration of ferredoxin and PSI, and is thereby attributed to the collision of undocked ferredoxin with PSI (Setif and Bottin, 1995).

Unfortunately, there is no direct way of comparing the spectroscopy results with a structural model without determining a structure with ferredoxin bound to PSI. Although crystals of PSI-Fd have been grown and X-ray crystallography data has been collected, only limited resolution data have been recorded and therefore no molecular has so far been derived from the experimental data (Fromme et al., 2003). Improved crystals, producing diffraction to 3.8-Å resolution have been grown recently by H.Q. Yu and R. Fromme, and a structure has been determined, however the data has not yet been published (Fromme and Fromme., 2011).

Structure determination of the photosynthetic proteins

The structures of the major proteins involved in oxygenic photosynthesis have been solved using X-ray crystallography. X-ray crystallography does not have size restrictions on the protein of interest, unlike NMR and single-particle electron microscopy, which is one of the reasons for its ubiquitous role in structural biology (Mueller et al., 2007). Additionally, X-ray crystallography has proven to be an invaluable tool for the determination of the structure of many small molecules, and its extension to the large macromolecules of the cell was inevitable.

1.3 X-ray protein crystallography

Conventional X-ray protein crystallography

The fundamentals of X-ray crystallography have been well established for close to a century, and the reader is directed to books on the subject for a thorough treatment of the subject (Woolfson, 1997; Drenth, 2007; Rupp, 2010). The introduction presented here will focus on the specifics of crystallography related to proteins and specifically the problems associated with X-ray protein crystallography, with references, as appropriate, to comprehensive sources.

X-ray crystallography is the workhorse of structural biology, having accounted for the majority of the biomacromolecular structures determined to date. The power of X-ray crystallography resides in the ability of reproducibly ordered molecules in a lattice to scatter electromagnetic radiation coherently. Coherent scattering implies that the scattered waves have a defined phase relationship, which allows for the addition of the amplitudes of the scattered waves, creating a pronounced effect for scattering to high resolution (Woolfson, 1997).

The scattering of X-rays by molecules depends on the atomic number of the atoms incorporated, which corresponds to the number of electrons present. In the forward direction, the limit of the atomic scattering factor as the scattering approaches zero is Z (Woolfson, 1997). However, biological macromolecules are composed mostly of H, C, N, O and minor amounts of S and P, with relatively low atomic numbers, causing the scattered intensity to be small. Additionally, unlike in the case of small inorganic or organic molecules, crystals of biomacromolecules contain a large percentage of water, commonly between 30 and 70% of the unit cell volume (Matthews, 1968), but as much as 90% solvent content has been found, creating large unit cells with few crystal contact sites. Consequently, in order to record X-ray diffraction to large angles and high resolution, large crystals of proteins are necessary for protein crystallography experi-

ments, due to the weak scattering of the constituent elements as well as the large unit cell solvent contents.

As an example of the power, and perils, of X-ray crystallography, consider an example from photosynthesis: PSI from the cyanobacterium *Thermosynechococcus elongatus*. The first micron-sized crystals of PSI from the thermophylic cyanobacterium *T. elongatus* were reported in 1988 (Witt et al., 1988). The first structural model of PSI, based on crystal diffraction to 6 Å resolution, was determined in 1993 (Krauss et al., 1993; Witt et al., 1994), followed by a 4 Å structure in 1996 (Krauss et al., 1996) and an improved structure at 4 Å in 1999 (Klukas et al., 1999b,a). In 2001, the structure of Photosystem I was unraveled at 2.5 Å resolution (Jordan et al., 2001) from cryogenically cooled crystals that were incubated in sucrose before freezing. Therefore, it took 13 years from the growth of the first microcrystals to the determination of the first near-atomic-resolution structure of Photosystem I based on large well-ordered single crystals under cryogenic conditions (Fromme and Mathis, 2004).

Photosystem I remains the largest membrane protein complex solved to molecular resolution, and the use of X-ray crystallography to determine the structure of such a complicated molecule is a *tour de force* for the technique. However, the work on the project took the tireless effort of many dedicated researchers for over a decade. A question that needs to be addressed pertains to the causes of the structure determination of a protein taking 13 years. In order to address this question, the difficulties with X-ray protein crystallography will need to be discussed from the general point of view. Besides the problems associated with expression and purification of proteins, the major difficulties of X-ray crystallography are: the growth of large and simultaneously well-ordered crystals of proteins, X-ray-induced radiation damage, the cryogenic cooling of protein crystals, and the handling of sensitive proteins.

Protein crystallization and the necessity for large crystals

Obtaining a suitable crystal for X-ray protein crystallography is the least understood step of the structure determination of a protein (Drenth, 2007). Although crystallization of small molecules is well understood, various properties of proteins could allow for more complicated mechanisms of crystal growth. Protein molecules are much more complicated in regard to composition, structure, degrees of freedom, surface properties, as well as surface potential, than small organic molecules. Additionally, due to the much larger sizes, proteins have far different transport properties in solution, causing slower crystal growth relative to the small organic crystals (Malkin et al., 1995). Membrane proteins are a notoriously difficult class of proteins to crystallize, and due to the importance of membrane proteins, an entire section—Section 1.4—is devoted to the description of the specific difficulties associated with X-ray membrane protein crystallography.

Protein crystal nucleation and growth are driven by supersaturation of the protein in solution. One of the difficulties associated with protein crystallization is the need for a much larger supersaturation—the supersaturation is defined as $\ln(c/s)$, where c is the concentration and s is the concentration at saturation—than what is typically needed in the case of small organic molecules (Malkin et al., 1995). The high supersaturation is necessary for the formation of the critical nuclei of the new phase, but high supersaturation is not desirable for the growth of large, well-ordered single crystals. Additionally, multiple protein-replete phases are possible in the case of proteins, such as amorphous precipitates, that are not ordered. High supersaturation will favor the formation of amorphous precipitates as opposed to crystals because while the protein crystals are thermodynamically favored, the amorphous precipitates are kinetically favored under high supersaturation conditions (McPherson et al., 1995).

An additional difficulty with protein crystal growth is that the growth of protein

microcrystals is much more common than the growth of macroscopic protein crystals (Mueller et al., 2007). Ultimately, crystal growth is determined by the aforementioned supersaturation as well as a kinetic factor—the kinetic factor dictates the rate of crystal growth and not *whether* it will grow—that takes into account the adsorption, lateral diffusion, and uptake of the protein molecules on the surface of the crystal. An unexpected difficulty is that even though protein crystal growth is determined by the supersaturation of the protein in solution, protein crystals will stop growing even in the presence of supersaturation (Kam et al., 1978). The difficulties associated with the growth of protein crystals are major issues for protein crystallography because of the weak scattering of the constituents; therefore, much work and extensive optimization is often necessary to grow large protein crystals (if conditions are found at all).

Even when a large protein crystal grows, the crystal will often suffer from large internal disorder, as measured by the mosaicity. In the mosaic model of crystals, a crystal is considered to be composed of smaller mosaic domains, and mosaicity, or mosaic spread, is a term used to describe the degree of angular misalignment of mosaic domains within a crystal (Woolfson, 1997). Working from the mosaic domain model, Bragg reflections can be broadened by the finite size of mosaic blocks, the angular misalignment of mosaic blocks, and additionally by variations in the unit cell parameters in different mosaic blocks (Nave, 1998). However, the mosaic domain model is an incomplete picture in the case of membrane proteins, and if PSI crystals are considered, it is seen that only four salt bridges make up the crystal contacts, which would be a primary cause of misalignment. However, misalignment of the unit cells could also occur due to excess surface area with respect to volume, because the molecules at the interface will have a reduced number of crystal contacts. Therefore, two effects can be seen as influencing the mosaicity of the crystal: for large crystals, it is the propagation of slight misalignments in different domains, whereas for small crystals, the increased surface area with respect to volume causes misalignment of a large number of molecules.

The broadening of the reflections by mosaic spread causes the scattered intensity to be subtended over a larger solid angle as shown in Fig. 1.6. The broadening can result in severe consequences, such as being unable to resolve the high-resolution diffraction spots, making it impossible to evaluate the structure to the maximum resolution recorded in the diffraction pattern.

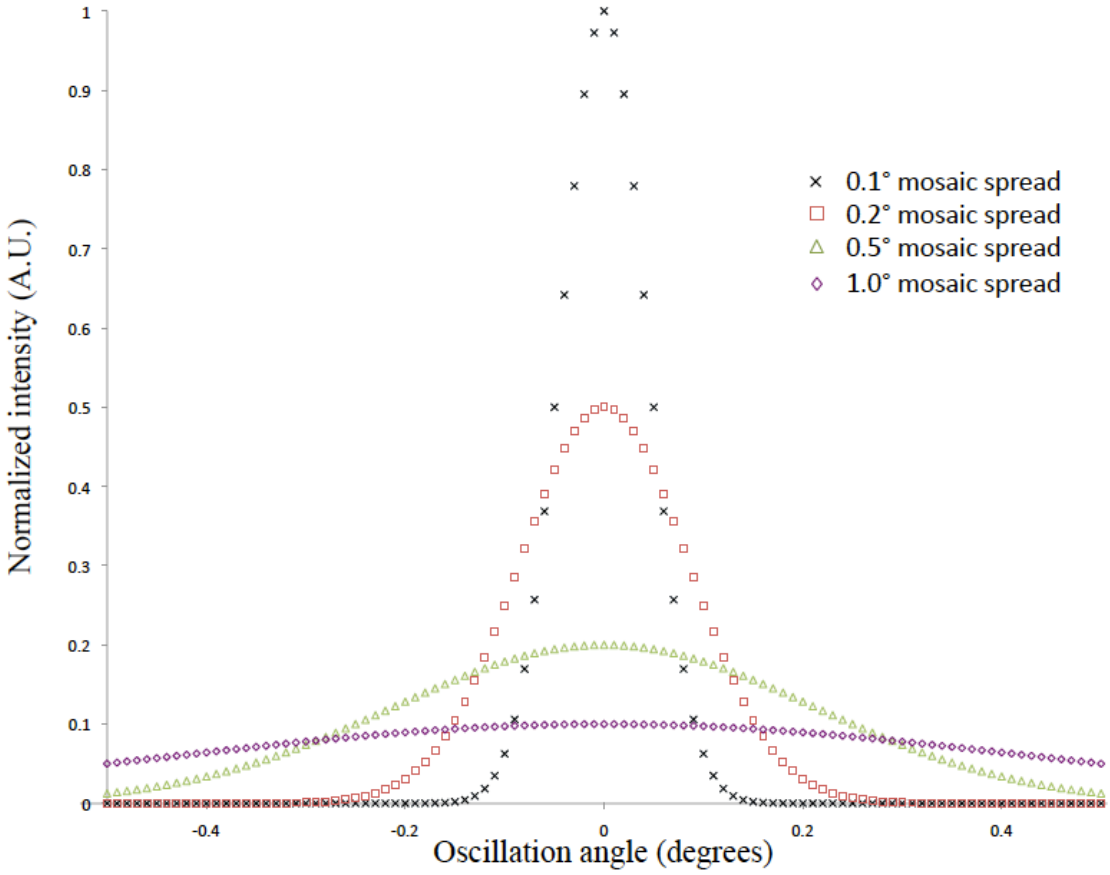

Figure 1.6: **Effects of mosaicity** Plot of Gaussian peak profiles showing the effect of increase mosaicity. The mosaicity is defined as the full-width at half-maximum of the peak and is related to the variance by FWHM = $2.3548(\sigma^2)^{1/2}$. The peak height has a linear dependence on the standard deviation of the Gaussian peak. The higher the mosaicity, the lower the peak maximum and the broader the peak width, which can have major implications for high-resolution data.

To give an example of the effect of mosaicity, a conventional diffraction pattern of Photosystem I is shown in Fig. 1.7a, with a resolution of 2.3 Å. The pattern was collected from a PSI crystal that was 2 mm long with a diameter of 0.5 mm at beamline

8.2.2 of the Advanced Light Source of Lawrence Berkeley National Laboratory. The pattern was collected by oscillating the crystal by only 0.2°, but as can be seen in Fig 1.7b, the individual spots become less distinguishable and eventually overlap, making the effective resolution of the data set less than the limit of the measured spots. The mosaicity of the PSI crystal was calculated to be 0.6° for this crystal.

The theoretical dependence of the scattered intensity on the size of a crystal can be readily understood though the Darwin equation. The Darwin equation can be used to calculate the integrated scattered intensity in an X-ray crystallography experiment, when using the oscillation method, and shows the scattered intensity with the following form (Holton and Frankel, 2010):

$$I = I_0 r_e^2 \frac{(1-\cos^2(2\theta))}{2\sin(2\theta)} \frac{\lambda^3}{\omega} \frac{V_x}{V^2} |F_{hkl}|^2 \tag{1.1}$$

in which I_0 is the incident intensity, $r_e{}^2$ is the classical electron radius, (1-$\cos^2(2\theta)$) is the polarization factor and assumes unpolarized radiation, 2sin2θ is the Lorentz factor, λ is the wavelength, ω is the angular speed of rotation, V_x is the crystal volume, V is the unit cell volume, and F_{hkl} is the structure factor associated with the particular reflection. From Eq. 1.1, it becomes clear that the integrated intensity is proportional to the crystal volume relative to the unit cell volume. Consequently, the larger the crystal, without regarding the internal order, the stronger the scattered intensity.

As mentioned already, protein crystallization experiments may result in the production of a shower of microcrystals, without subsequent optimization experiments producing large, well-ordered protein crystals necessary for data collection at conventional protein crystallography beamlines (Cusack et al., 1998). However, certain benefits can be found to smaller protein crystals, when issues such as high mosaicity or twinning plague the data (Perrakis et al., 1999), but the tradeoff is weaker scattering. In addition to the weaker scattering provided by a smaller protein crystal, if the crys-

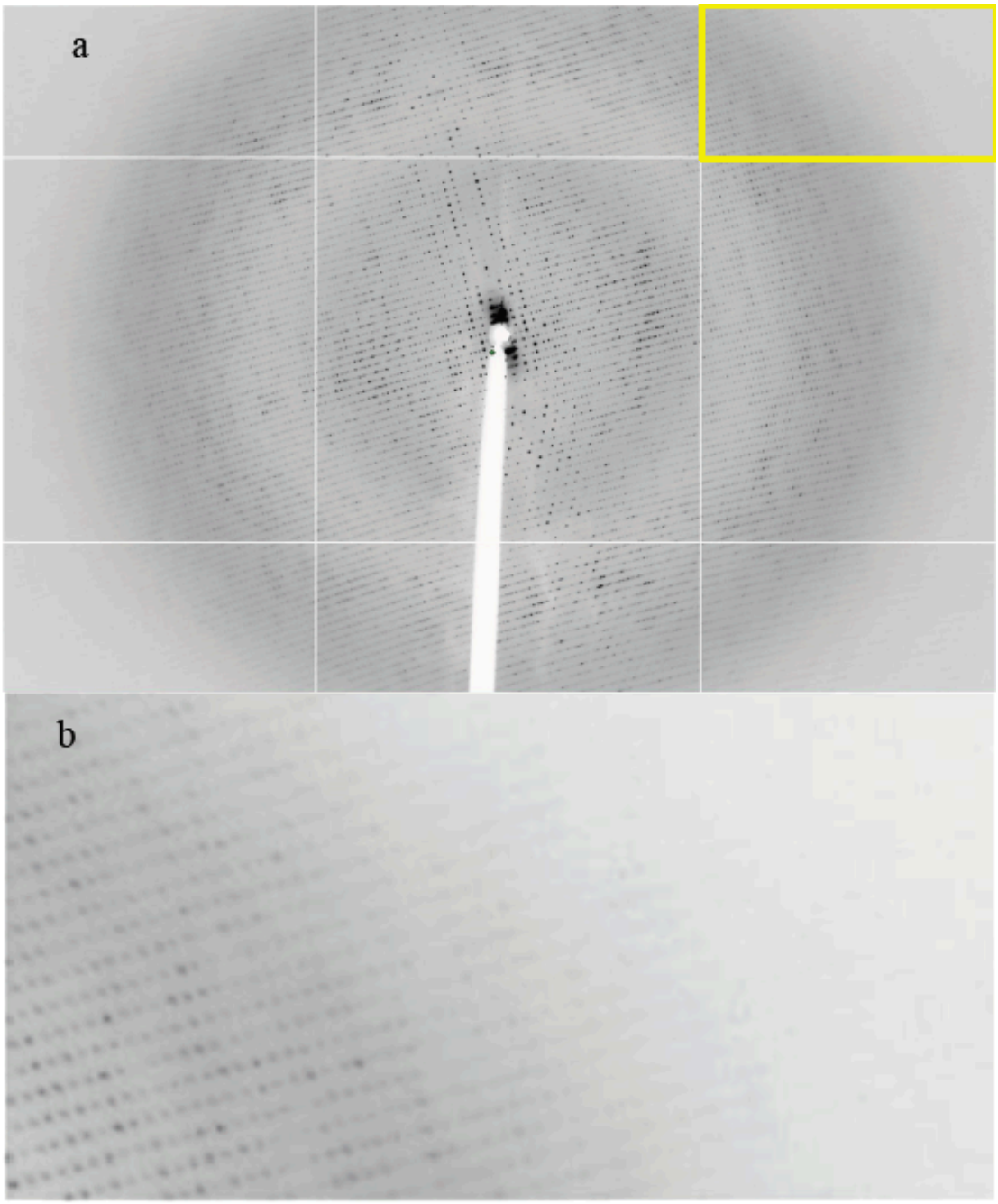

Figure 1.7: **Conventional X-ray diffraction pattern for Photosystem I** (a) Diffraction pattern of Photosystem I collected to 2.3-Å resolution at beamline 8.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory. The divergence of the beam was set to 2 mrad and the measured mosaicity was 0.6°. Above 3-Å resolution, the spots become less distinguishable and start to become unresolved, decreasing the effective resolution of the diffraction pattern. (b) The magnified upper-right detector panel from (a) showing the streaking of the spots due to the high mosaicity.

tal is smaller than the X-ray beam focus, then an increased background relative to the signal will also be recorded. One technology development that has increased the usefulness of protein microcrystals for structure determination is the microfocussed beamline (Bilderback et al., 2010).

The microfocussed beamline compensates for the weaker scattering of a smaller crystal by increasing the flux density at the sample, I_0 on the RHS of Eq. 1.1. A (20-μm)^3 crystal will scatter 1000x weaker than a (200-μm)^3 crystal, which could be compensated for through a 1000x increase in the flux density on the sample by decreasing the focus spot radius by a factor of $1000^{1/2} \approx 31.6$. The reduced beam focus will also result in a smaller background scattering from the mother liquor. However, radiation damage becomes a limiting factor in data collection at microfocused beamlines, as discussed in the following section.

Another way to interpret the Darwin equation (Eq. 1.1) is that instead of increasing the size of a protein crystal, or increasing the flux density incident on the sample, the exposure time can be increased by an equivalent amount. Consequently, one route to consider in an attempt to increase the integrated scattered intensity from a protein crystal is to simply irradiate the crystal longer in the X-ray beam by reducing ω during the collection, or by re-measuring the same angular rotation multiple times. Fourier-transform infrared spectroscopy (FTIR spectroscopy) and NMR utilize similar techniques in which multiple measurements are made in order to increase the signal-to-noise ratio of the data. However, this is not feasible with X-ray diffraction, as will be discussed in the subsequent section, as radiation damage ultimately becomes the biggest limitation to X-ray protein crystallography.

Radiation damage and the unending quest to mitigate it

X-ray diffraction is inherently a destructive imaging technique because X-rays are a form of ionizing radiation. Unfortunately, the ionizing nature of the X-rays leads to

radiation damage, which along with sample heterogeneity ultimately limits the resolution of a biomacromolecular structure determined using X-ray crystallography (Meents et al., 2010). The problem that ultimately arises with the use of conventional X-ray sources to determine a structure is that the smaller the object one wishes to determine the structure of, the smaller the scattering cross-sectional area. Consequently, to obtain enough scattered intensity to high resolution, the total fluence incident on the sample must be increased, as shown in Fig 1.8, increasing the risk of X-ray-induced radiation damage and ultimately decreasing the effective resolution of the reconstructed object (Chapman, 2009).

There are two types of X-ray-induced radiation damage during protein crystallography: global damage and local damage (Holton, 2009). Global radiation damage appears quantitatively in the data set and is not associated with any one particular element of the electron density, and can become manifest as a reduced total scattered intensity, increased unit cell parameters, and increased mosaicity of the crystal (or some combination), amongst other effects, but has a dependence on the experimental conditions (Meents et al., 2010). Local damage is damage affecting specific sites such as amino acids within the protein structure itself, and is often ascertained when looking at the electron density maps of the structure of interest (Holton, 2009).

When protein crystals are irradiated by X-rays, an X-ray can either interact with the sample, or it can pass through without being affected by the presence of the sample. In the case of 12.4 keV X-rays (λ = 1 Å), the wavelength typically used for macromolecular crystallography, only approximately 2% of incident X-rays will interact with a 100-μm thick crystal. Of the interacting 2%, 84% will interact through the photoelectric effect, causing the ejection of photoelectrons. Only 8% of the interacting X-rays will produce elastic scattering events, whereas the remaining 8% will produce inelastic scattering (Paithankar et al., 2009). The relatively low ratio of the scattering-cross-sectional area of the light elements for X-rays with respect to photo-ionization

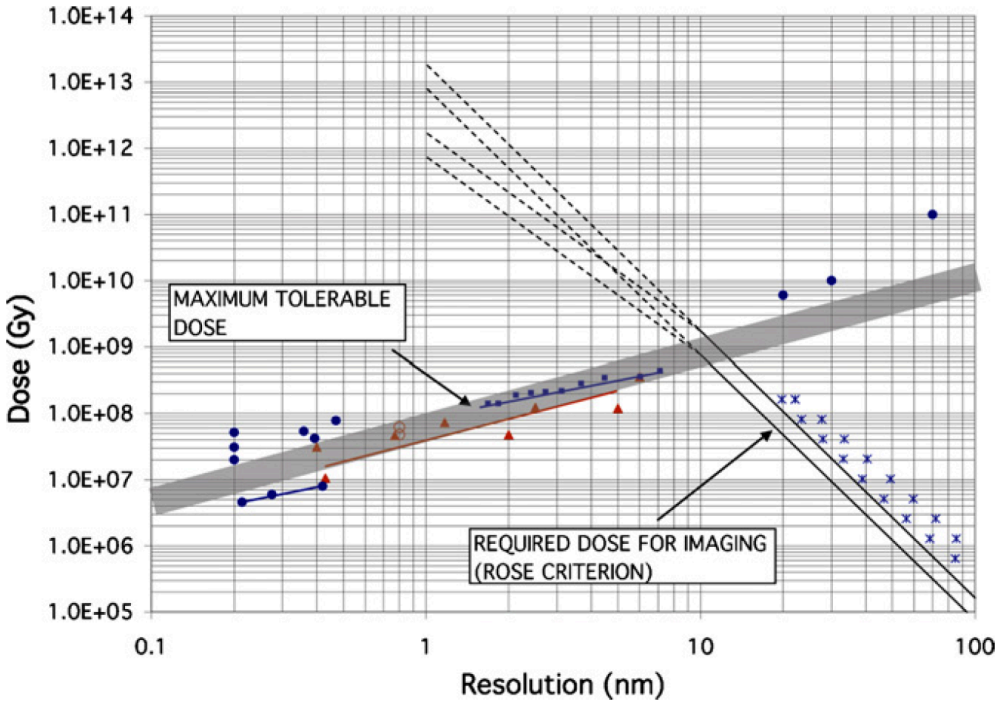

Figure 1.8: **Dose versus resolution using X-rays** Graph summarizing information on the *required dose for imaging]* and the *maximum tolerable dose*. The required dose for imaging is calculated for protein of empirical formula H50C30N9O10S1 and density 1.35 g/cm^3 against a background of water for X-ray energies of 1 keV (lower continuous line) and 10 keV (upper continuous line). The dashed continuations of these lines refer to the transition region from coherent (d-4 scaling) to incoherent (d-3 scaling) behavior, both of which are shown down to 1 nm resolution. Some experimental measurements of the required dose for imaging are plotted as crosses. The maximum tolerable dose is obtained from a variety of experiments and from the literature. The types of data from the literature are identified by the symbols as follows: filled circles: X-ray crystallography; filled triangles: electron crystallography; open circles: single-particle reconstruction; open triangles: electron tomography; diamonds: soft X-ray microscopy (including XDM); filled squares: ribosome experiment. Figure and caption taken from (Howells et al., 2009).

and inelastic scattering is one of the major weaknesses of X-ray crystallography with respect to related electron microscopy techniques (Henderson, 1990).

The primary radiation damage occurs due to the inelastic events, *i.e.* the photoelectric effect and Compton scattering. The photoelectric effect will cause the ejection of energetic (tens of keV) electrons from the K-shells of many of the light elements (and of other shells in heavier metals). The energy deposited by the primary events will cause a cascade of additional photoelectrons, with energies of a few to tens of eV. Possible effects of the primary and secondary damage are the breakage of chemical bonds, oxidation-reduction reactions, the generation of free radicals, and the production of gaseous species. The accumulation of all of the damage at the microscopic level can lead to strain in the crystal and cause distortion of the long-range order of the molecules in a crystal, leading to the tertiary damage, or global damage (Meents et al., 2010).

The specific damage to amino acid residues/sites within the protein depends on where the amino acid resides within the structure, but generally speaking, the more solvent-exposed amino-acid residues/sites are more radiation-damage susceptible. Studies have indicated that the order of damage to residues/sites generally proceeds with a reduction of disulfide bonds, decarboxylation of aspartates and glutamates, loss of hydroxyl groups on tyrosines, followed by a loss of methylthio groups from methionines (Burmeister, 2000). The damage also depends on the cofactor content, with metals or cofactors that are more prone to forming radicals upon X-ray exposure incurring a proportionally higher damaging effect from the X-rays.

Although the radiation chemistry was not fully understood at the outset of protein X-ray crystallography, very quickly after the initial development of the technique, at which point data collection was done at room temperature, it was realized that damage was occurring due to the X-ray exposure. Mitigation of radiation damage became a driving force for new developments in techniques for X-ray crystallography.

Any amount of cooling during X-ray exposure was found to help mitigate radiation damage (Haas and Rossmann, 1970) and ultimately cryogenically cooling of protein crystals was developed (Hope, 1988). The vitreous-solvent matrix that is formed does not reduce the number of radicals, but limits the diffusion of the destructive radicals throughout the unit cell and crystal (Henderson, 1990). However, once the use of 3^{rd} generation synchrotrons became more widespread for structural biology, the high X-ray doses lead to significant damage even under cryogenic conditions; therefore, data sets were still being recorded that were produced from damaged species. That the damage still occurred at cryogenic temperatures could be attributed to the high amount of radicals produced; ultimately there is an absolute limit to the dose a protein sample can tolerate, after which it requisitely deteriorates due to the detrimental effect of the large amount of radicals. The effects of the damage would ultimately be manifest in the degradation of the diffraction pattern, which is the global damage (Murray et al., 2004).

Empirically, for cryogenically cooled protein crystals, the global damage does not depend on the dose rate but is only related to the accumulated dose (Holton, 2009), possibly due to the low diffusion rates of the radicals at ˜100 K in a vitreous solid matrix. The conventional cutoff for X-ray crystallography data collection, defined as the dose required to reduce the total scattering to half of its initial value, using cryogenically cooled protein crystals was estimated to be 20 MGy based upon estimations from cryo-EM (Henderson, 1995). More recently, using synchrotron radiation, the number was empirically determined to be 42 MGy (Owen et al., 2006). However, the authors recommended that data collection should cease once the scattered intensity decreased by a factor of ln(2) = 0.693 from the initial scattered intensity, which corresponded quite well with a 30 MGy limit. The damage limits as discussed are only in relation to the global damage. However, specific local damage, situated on individual amino acid residues, or at specific sites, can start well before the global damage thresholds

are reached (Meents et al., 2010; Owen et al., 2006). In the case of PSII, individual structure elements that are prone to oxidation-reduction chemistry could be damaged with much lower doses; thereby, the allowed dose will depend very specifically on the biomacromolecule of interest; metal containing proteins; proteins with more radiation-damage susceptible amino acids in the catalytic site, solvent exposed, or at crystal contacts; as well as proteins with solvent exposed disulfide bridges, should have the data collection strategy carefully considered, and reduced dose limits (Meents et al., 2010). A (very general) rule of thumb for proteins that can be used is that for every 1 Å of resolution in the structure, 10 MGy can be absorbed, such that for a 3-Å structure, 30 MGy can be absorbed (Howells et al., 2009). It should be noted, however, that photoreduction of metals can occur from a dose that is one to two orders of magnitude lower than the 30 MGy limit (Yano et al., 2005), and consequently, the rule of thumb from above should be applied cautiously and only for proteins that do not contain significant amounts of metals or cofactors that would increase the damage.

Of course, cryogenically cooling a crystal will have a greater effect than simply mitigating X-ray-induced radiation damage. Cryogenic cooling of a protein crystal allows for a reduction of the Debye-Waller factor, and this could lead to more intense scattering to high resolution for a cryogenically cooled crystal (Hope, 1988).

A few major limitations arise in the use of protein X-ray crystallography at cryogenic temperatures. A major difficulty with cryogenic cooling of protein crystals is that most protein crystals cannot simply be harvested from their crystallization growth medium and cooled directly. The protein crystals must be treated prior to cooling in order to ensure preservation of the crystallographic integrity of the specimen. Some protein crystals can be directly cryogenically cooled using a cold stream, but these are not to be considered the typical case (Parkin and Hope, 1998).

A commonly encountered problem is that of cubic ice or Type I ice formation during the cooling process, which is produced due to the cooling rate being insufficient

to trap the solvent in a super-cooled meta-stable phase. Early attempts to rectify the situation involved procedures to modify the composition of the water within solvent channels (Petsko, 1975), possibly due to the belief that the cubic ice formation was occurring in an internal portion of the crystal. However, the protein environment inside of the protein crystals likely inhibits nucleation of cubic ice (Warkentin and Thorne, 2009), and, consequently the ice will nucleate in the solvent surrounding the crystal, then grow into and within the crystal, through solvent channels, damaging the crystalline integrity of the sample due to the lower density of cubic ice with respect to water. Additionally, the formation of cubic ice will produce strong diffraction rings within the diffraction pattern that must be systematically removed in order to properly evaluate the data (Parkin and Hope, 1998).

Ultimately, the use of cryo-protectants to avoid cubic ice formation can prove problematic. The addition of the cryo-protectant will change the properties of the solution and could lead to dissolution of the protein crystal, or in some cases, lead to the cracking or breaking of a protein crystal. Consequently, different cryo-protectants must be tested for compatibility with a protein crystal, possibly by studying the effects of the cryo-protectant on smaller crystals of the same protein. However, such a task may be a daunting undertaking.

An additional difficulty arising from cryogenically cooling protein crystals is an increase in the mosaic spread of the crystal (Gonzalez et al., 1992). The reasons for the decrease in the crystalline order upon cooling are still not completely understood. The increase in mosaicity upon cryogenic cooling depends on the size of the crystal and cooling rate; for this reason, crystals of PSI greater than 1 mm in size must be cryogenically cooled with propane and not liquid nitrogen, as liquid propane has a higher heat capacity and has a 30X higher thermal conductivity than liquid nitrogen. Interestingly, slow-cooling a protein crystal at 0.1 K/s, which is 10^3-10^4 times slower than flash cooling, has been shown to produce cryogenically cooled protein crystals

with low mosaic spread. In order to slow cool the protein crystals, however, the crystal and mother liquor must be transferred to an oil drop, and mother liquor surrounding the protein crystal needs to be removed (Warkentin and Thorne, 2009). The final layers must be mechanically removed through the use of a chisel or other instrument, which may limit the technique in applicability, as soft membrane-protein crystals need to be handled with caution, as discussed in Section 1.4.

As can be seen, much effort is devoted to the minimization of radiation damage. Cryogenic cooling of protein crystals and the use of microfocused beams can allow for less of the crystal to be inundated with damaging X-rays which allow for the frequent shifting of the crystal during data collection to unexposed regions. In sum, the main difficulty with X-ray protein crystallography emanates from the fact that X-ray diffraction is inherently a destructive structural determination technique. Unfortunately, the difficulties associated with X-ray protein crystallography appear particularly acute for one class of proteins, the membrane intrinsic proteins.

1.4 Membrane proteins and the difficulties with crystallization

The success of structure determination of the membrane protein complexes associated with oxygenic photosynthesis is the exception, and not the rule. In fact, membrane proteins are strongly under represented in the protein data bank (www.pdb.org), when compared to the soluble proteins (White, 2004).

About 30% of all proteins in cells are membrane proteins, and they are also of great importance for human health, with more than 60% of all current drugs targeting membrane proteins. However, structural information is rare for membrane proteins, especially membrane protein complexes, with less than 300 unique membrane protein structures presently identified. Membrane protein structure determination is of extreme importance for understanding fundamental principles in biology as they represent the key players in the most important processes of all living cells, such as respiration,

photosynthesis, ion-, nutrient and hormone transport, cell communication, signal transduction, and vision and nerve function. Despite their extremely high impact, only four medically relevant human membrane protein structures have been determined to date, that of a G protein-coupled receptor, the β_2-andronergic receptor (Cherezov and Caffrey, 2007; Rosenbaum et al., 2007; Rasmussen et al., 2007), human Aquaporin-5 (Horsefield et al., 2008), and human leukotriene C_4 synthase (Ago et al., 2007), and the human dopamine receptor (Chien et al., 2010).

There are two main challenges associated with solving the structure of a membrane protein. The problem that is ultimately technique independent is the limited natural abundance of many membrane proteins, and methods to produce sufficient quantities of membrane proteins for structural studies is an ongoing challenge (Gabrielsen et al., 2009). However, a discussion of the production of large amounts of membrane protein is beyond the scope of this work.

The major challenge of utilizing X-ray protein crystallography for membrane proteins, once sufficient quantities of the protein are purified, is the difficulty of producing membrane protein crystals. In principle, the crystallization of membrane proteins is similar to that of soluble proteins, but the major difference is that membrane proteins are amphiphilic, having both hydrophobic and hydrophilic domains (Gabrielsen et al., 2010). In order to solubilize the membrane protein, the hydrophobic regions must be shielded from the hydrophilic solvent, and this is mainly accomplished through the use of detergents.

The interaction of detergents with membrane proteins is complicated, but in general, in order to be used for crystallization, a detergent must have the ability to surround the entire hydrophobic section of the protein while still allowing enough surface area for crystal contacts to be made via the hydrophilic portion of the protein. Often, even when a suitable detergent has been found, extra-membrane loops that are flexible may be present and inhibit crystallization (Prive, 2007).

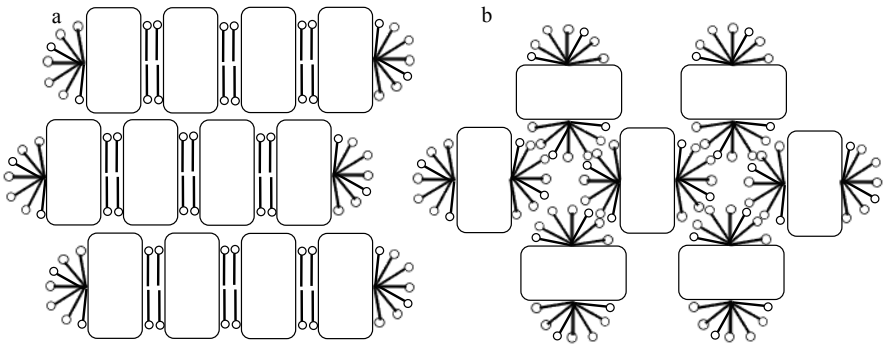

Figure 1.9: **Type I and II membrane protein crystals** (a) Type I membrane protein crystal in which a lamellar layer of detergent forms with the imbedded proteins forming a stack of two-dimensional crystal layers. Additional layers add to the direction normal to the membrane, forming a three-dimensional crystal with crystal contacts between the hydrophilic regions of the protein molecules. (b) Type II membrane protein crystal in which crystal contacts are made between the hydrophilic regions of the protein molecules and the detergent remains as a micelle around the hydrophobic portion of the protein.

Three different types of membrane protein crystals have been found to date and are called Type I, II (Ostermeier and Michel, 1997), and Type III (Liu et al., 2004), with a schematic drawing found in Fig. 1.9. Type I membrane protein crystals, as shown in Fig. 1.9a, have lamellar layers of lipid form in two dimensions and subsequently stack to form a three-dimensional crystal, with crystal contacts between the hydrophilic regions of the protein of different layers; Photosystem I is an example of a membrane protein that forms Type I crystals. Type II membrane protein crystals, as shown in Fig. 1.9b, are the most common type of membrane protein crystal, and have the proteins crystallize with micelles surrounding the hydrophobic region while the crystal contacts form at the hydrophilic regions; Photosystem II is an example of a membrane protein that forms Type II crystals. Only one example of a Type III membrane protein crystal has been reported to date, but the crystal contacts are from the lipid vesicles surrounding the hydrophobic regions; the light-harvesting complex II is the membrane protein that forms Type III crystals (Liu et al., 2004). Many different chemicals have been tested for use in membrane protein crystallization, such as replacing surfactants with amphipols (Popot et al., 2003), or utilizing *in meso* crystallization

matrices (Cherezov et al., 2006), to name only two. However, showers of microcrystals of membrane proteins are more common than macroscopic crystals of membrane proteins in most crystallization screens (Cherezov, 2011). With a higher probability of producing microcrystals than their macrocrystal counterparts, finding a method tin which to use the microcrystals while avoiding radiation damage would help to improve the structure determination process.

Chapter 2

MOTIVATION AND OBJECTIVES

2.1 General Motivation

The role of X-ray protein crystallography in structural biology cannot be overstated. However, as discussed in Section 1.3, many difficulties present themselves during X-ray protein crystallography, especially of membrane proteins. The two major problems associated with the technique are the requisite growth of large, well-ordered protein crystals and the accrual of radiation damage in the sample. A method that would allow the use of the showers of membrane protein crystallites that are more commonly produced during crystallization screens would greatly increase the effectiveness and efficiency of X-ray crystallography for protein structure determination. However, the invention of a new type of X-ray source may be the necessary breakthrough for X-ray protein crystallography on protein microcrystals.

2.2 VUV and X-ray free electron lasers

Third-generation X-ray sources, which refer to storage rings that utilize insertion devices such as undulators and wigglers, have revolutionized X-ray experiments in many fields of science, with one prominent example being X-ray protein crystallography. X-ray sources are usually described in terms of the brilliance (photons s^{-1} mm^{-2} $mrad^{-2}$ /0.1% bandwidth) of the source, which is a measure that takes into account many properties of the X-ray beam, such as photon flux, beam divergence, and bandwidth, and can succinctly be interpreted as a measure of the quality of the X-ray beam. Third-generation X-ray sources offer much higher brilliance than any previous X-ray source, as shown in Fig. 2.1, with APS, Spring-8, ESRF, and PETRA III being examples of third-generation X-ray sources (Altarelli, 2010). However, as discussed in 1.3, serious limitations are still restricting the X-ray protein crystallography community even with the microfocused beamlines of the third-generation X-ray sources, due to X-ray-

induced radiation damage.

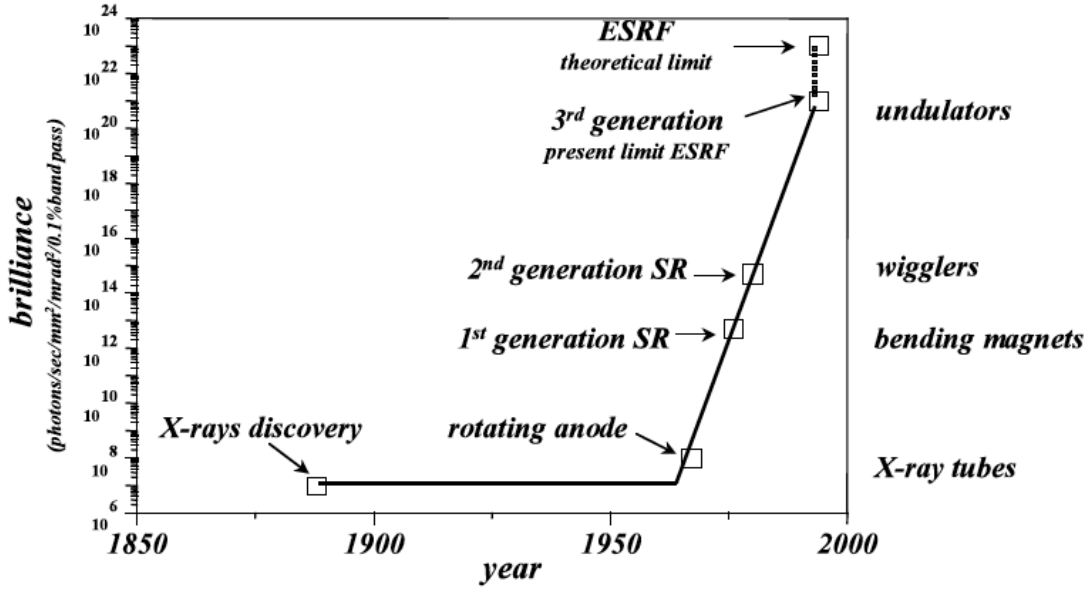

Figure 2.1: **Brilliance of third-generation X-ray sources.** Development of average x-ray source brilliance and introduction of specific radiation sources. SR refers to synchrotron radiation. The European Synchrotron Radiation Facility (ESRF), the Advanced Photon Source (APS) at Argonne National Laboratory, and PETRA III at Deutsches Elektronen-Synchrotron are three examples of current third-generation synchrotron sources. Figure taken and caption adapted from (Altarelli, 2010).

Recently, fourth-generation VUV and X-ray sources have been designed and built, with noticeable examples being FLASH at Deutsches Elektronen-Synchrotron (DESY) (Ayvazyan et al., 2002), the LCLS at SLAC National Accelerator Laboratory (Arthur et al., 1995), the SPring-8 Compact SASE Source at SPring-8 (Shintake et al., 2003), and the upcoming European XFEL at DESY (Geloni et al., 2010). The peak brilliance of these fourth-generation X-ray sources, known as the X-ray free electron lasers or XFELs, is up to 10 orders of magnitude higher than the peak brilliance currently available at even the most brilliant third-generation X-ray sources (Arthur et al., 1995) as can be seen in Fig. 2.2. Due to the incredible peak brilliance offered by the XFELs, many new areas of physics and chemistry could be examined with unprecedented spatial resolution.

XFELs involve relativistic electron bunches from a linear accelerator that are

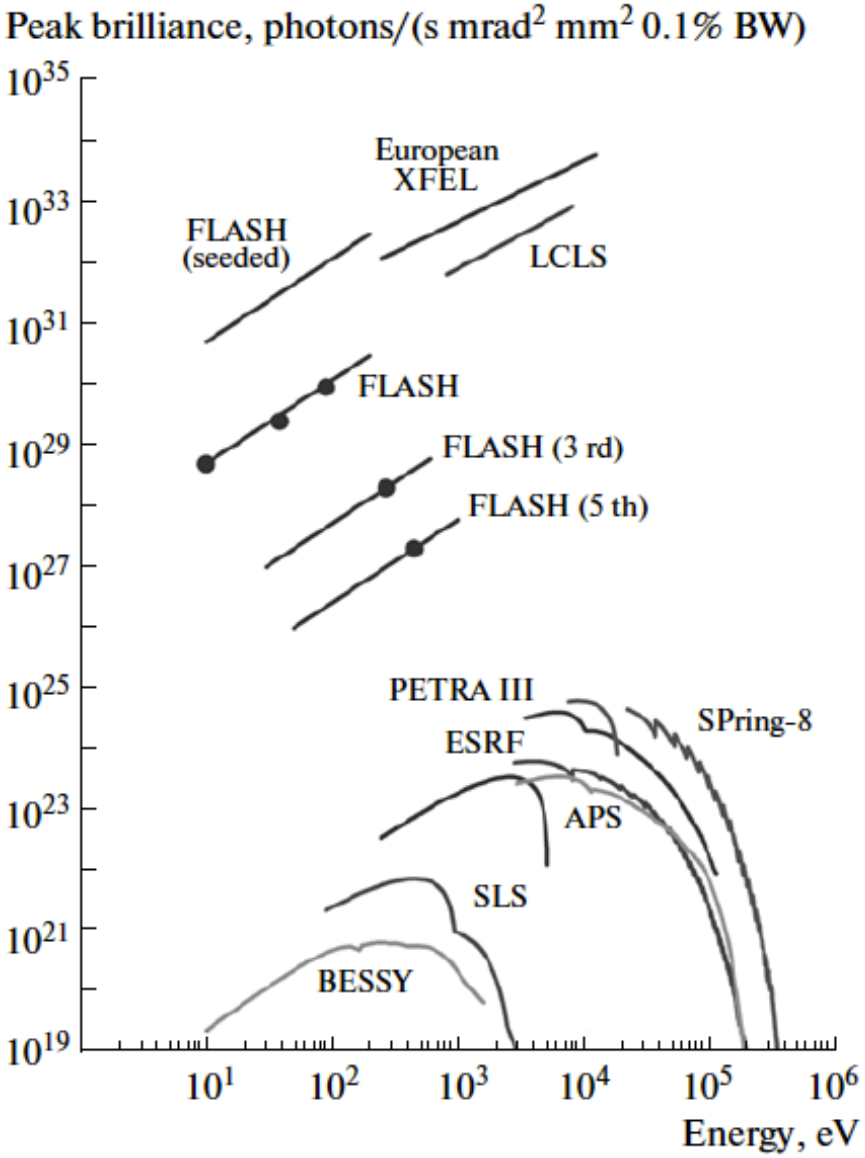

Figure 2.2: **Comparison of X-ray sources** Peak brilliance as a function of photon energy of FLASH (including an upgraded version of FLASH based on "seeding") and hard X-ray FELs (the European XFEL in Hamburg, and LCLS in Stanford, compared with some 3rd-generation synchrotron radiation facilities. Blue dots denote measured values. The third and fifth harmonics of the FLASH undulator, on which lasing was observed, but not saturation, are also shown. Figure taken and caption adapted from (Altarelli, 2010).

compressed in a bunch compressor and fed into a long undulator. In the undulator, X-rays with extremely high brilliance are created by self-amplified spontanteous emission (SASE). To achieve SASE, the XFEL is tuned such that when an electron bunch traverses one period of the undulator, the radiation emitted by the adjacent upstream electron bunch is in phase with that emitted by the downstream electron bunch. The ponderomotive force (the force felt by a charged particle in an oscillating electromagnetic field) accelerates those electrons (in the downstream bunch) that are out of phase with the radiation, while decelerating those that are in phase, creating a fine structure of microbunches within the electron bunch. As a result, the spontaneously emitted radiation from the microbunches further amplifies the coherence of the X-ray wavefield, which in turn forces the microbunches downstream in a more well-defined bunch, thus emitting even more coherently. An exponential increase in the spontaneous emission through this resonant process leads to saturation of the XFEL and to the arrival of X-ray bunches of extremely high brilliance at the interaction region (Bonifacio et al., 1984).

The peak intensity of an XFEL is much higher than that of any third-generation X-ray source and could be of use for X-ray protein crystallography. However, as discussed in Section 1.3, the increased intensity will not solve the major problem of X-ray protein crystallography, which is the X-ray-induced radiation damage. Fortunately, the XFELs offer one important characteristic that may lend itself to mitigation or removal of X-ray-induced radiation damage altogether, the ability to produce ultra-short X-ray pulse durations.

2.3 Diffract before destroy

The X-ray beam produced at an XFEL is not a continuous X-ray beam (and neither is the X-ray beam produced at a third-generation X-ray source, for that matter), but the beam arrives in pulses with pulse durations related to the electron bunch length in the undulator (Young et al., 2010). The XFELs are able to operate with X-ray pulse

durations of several femtoseconds to several hundreds of femtoseconds (Arthur et al., 1995; Doniach, 1996).

The time scale of the pulse duration offered by XFELs is on the same (or shorter) time scale as many of the important physical and chemical processes that are involved in X-ray-induced radiation damage. Initial ionization and excitation by X-rays occurs on the 100 attosecond to tens of femtoseconds scale. Relaxation, which is the X-ray fluorescence and Auger recombination events, occurs at 10 fs and longer time scales. As the relaxation events are the main mechanism for the specific damage in X-ray protein crystallography, the short time scale of the X-ray pulses may allow diffraction data to be collected before significant damage degrades the crystal and the diffraction pattern (Hau-Riege et al., 2004).

In the year 2000, a paper was published showing simulations of the damage pathway of a lysozyme molecule in the gas phase that interacts with an intense X-ray pulse from an XFEL (Neutze et al., 2000). Ultimately, the large positive charge harbored by the lysozyme molecule led to a Coulomb explosion, but the simulations showed that the explosion occurred on the time scale of 5-10 femtoseconds, as shown in Fig. 2.3. The simulation results can be interpreted such that if an X-ray pulse terminates before the onset of the Coulomb explosion, all of the diffraction data collected will be without the degraded signal from the damaged sample, which could add significant background to the diffraction patterns. The results led the authors to propose the diffract-before-destroy principle, in which X-ray-induced radiation damage is avoided by the termination of the X-rays used before the onset of the secondary damage processes. However, the simulation results were published before any VUV or X-ray free electron laser became opperational, and so it was not known whether the simulations would be commensurate with experimental data obtained using an XFEL.

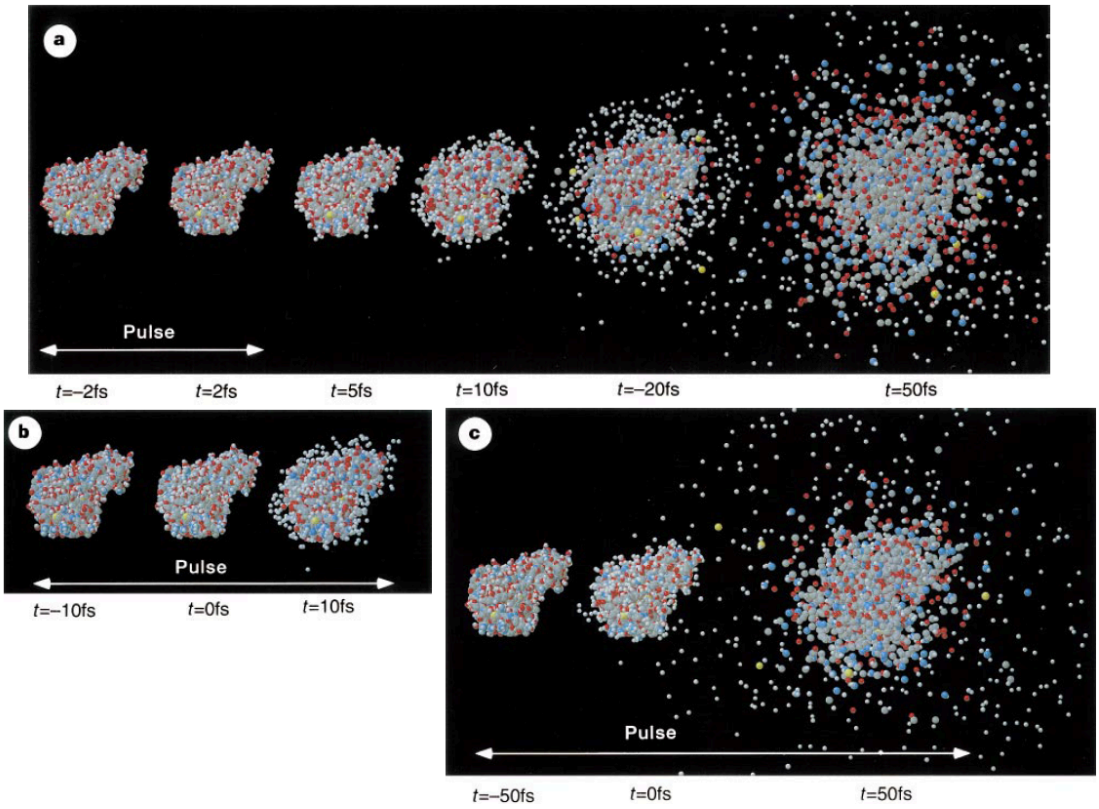

Figure 2.3: **Simulations of the Coulomb explosion** Explosion of T4 lysozyme (white, H; grey, C; blue, N; red, O; yellow, S) induced by radiation damage. The integrated X-ray intensity was $3x10^{12}$ (12 keV) photons per 100-nm diameter spot ($3.8x10^{6}$ photons per Å^2) in all cases. (a) A protein exposed to an X-ray pulse with an FWHM of 2 fs, and disintegration followed in time. Atomic positions in the first two structures (before and after the pulse) are practically identical at this pulse length because of an inertial delay in the explosion. R_{nucl} = 3%, R_{elec} = 11% (b) Lysozyme exposed to the same number of photons as in a, but the FWHM of the pulse was 10 fs. Images show the structure at the beginning, in the middle and near the end of the X-ray pulse. R_{nucl} = 7%, R_{elec} = 12% (c) Behavior of the protein during an X-ray pulse with an FWHM of 50 fs. R_{nucl} = 26%, R_{elec} = 30%. Figure taken and caption adapted from (Neutze et al., 2000).

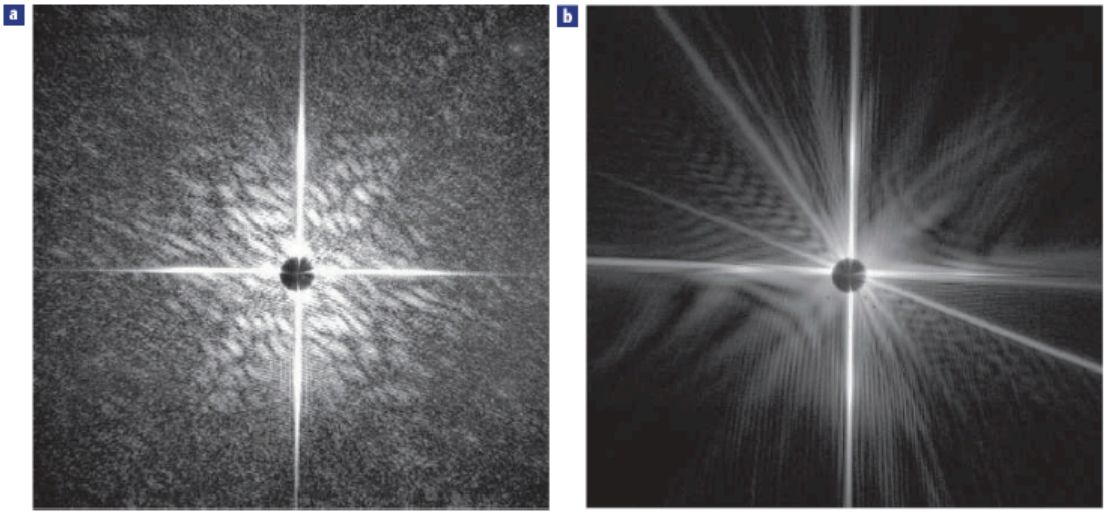

Figure 2.4: **Flash X-ray coherent diffraction patterns** a,b, Coherent diffraction patterns recorded for a single $(4\pm2)x10^{14}$ Wcm^{-2}, 25$\pm$5 fs pulse (a) and for the subsequent pulse of similar intensity and duration, 20 s later (b), showing diffraction from the damage caused by the pulse that formed a. The intensity is shown on a logarithmic greyscale, with black denoting one photon per pixel and white denoting 2,000 photons per pixel for a and 50,000 photons per pixel for b. The entire patterns are shown as detected by the CCD, and extend to a diffraction angle of 15° at the midpoint of the edges (corresponding to a momentum transfer of 8.1 μm^{-1}. Figure and caption adapted from (Chapman et al., 2006).

2.4 First results of VUV and X-ray free electron lasers

The Free electron LASer in Hamburg Germany, or FLASH, started operation in 2005 and was the first VUV free-electron laser. FLASH offered the short pulses expected from the upcoming XFELs, 10-50 fs in duration – with current abilities of 10-100fs – with high peak intensities of up to 10^{16} W/cm^2. FLASH offers wavelengths between 47 and 470 Å (Chapman et al., 2010). Although FLASH did not offer the necessary wavelengths to generate high-resolution structures, FLASH offered an important opportunity to probe the interaction of matter with the high-intensity free electron laser pulses. As such, many experiments were designed and carried out on inorganic objects, cells, and viruses, amongst other samples, in order to determine whether the simulation results were indeed indicative of the damage-free potential of XFELs.

One of the most exciting experiments involved with the new fourth-generation X-ray sources tested the diffract-before-destroy principle experimentally, albeit at low resolution for the initial FLASH work. Using objects etched into silicon nitride windows, experiments were carried out to determine whether the diffraction patterns collected from FLASH could be used to produce an undamaged image of the object (Chapman et al., 2006). Fig. 2.4 shows two diffraction patterns recorded for an object etched into a silicon nitride membrane. The diffraction pattern in (a) shows the recorded image for an initial pulse of FLASH, while (b) shows the diffraction pattern from a subsequent pulse, indicating that the object had been destroyed by the first FLASH pulse. One of the ubiquitous images associated with XFELs was reconstructed from Fig. 2.4a, and is depicted in Fig. 2.5, showing that the reconstructed density did not show damage, to 32-nm resolution. Due to the limited resolution offered by FLASH, it was unknown whether the undamaged reconstruction would be available at higher resolutions or whether damage would be manifest in the data.

Another experiment of great interest was to determine the explosion dynamics of samples irradiated by intense VUV and X-rays. Results from FLASH showed that the presence of a layer of solvation around the sample of interest will cause the Coulomb explosion to be retarded (Hau-Riege et al., 2010). The retarded explosion provided evidence that the pulse durations of the XFELs could be longer than the simulation of the gas-phase biomolecule, due to a sacrificial tampering by the solvent layer. A mechanism to explain the sacrificial tamper was suggested in which the photoelectrons ejected by the solvation layer could partially, and temporarily, neutralize the large positive charge generated at the sample by the intense radiation, allowing more time before the Coulomb explosion. The sacrificial tamper effect could have a major effect in X-ray crystallography at XFELs, especially since the protein crystals used require solvent layers, and the solvent layers could provide the sacrificial tamper and minimize damage to the sample during the X-ray exposure at an XFEL.

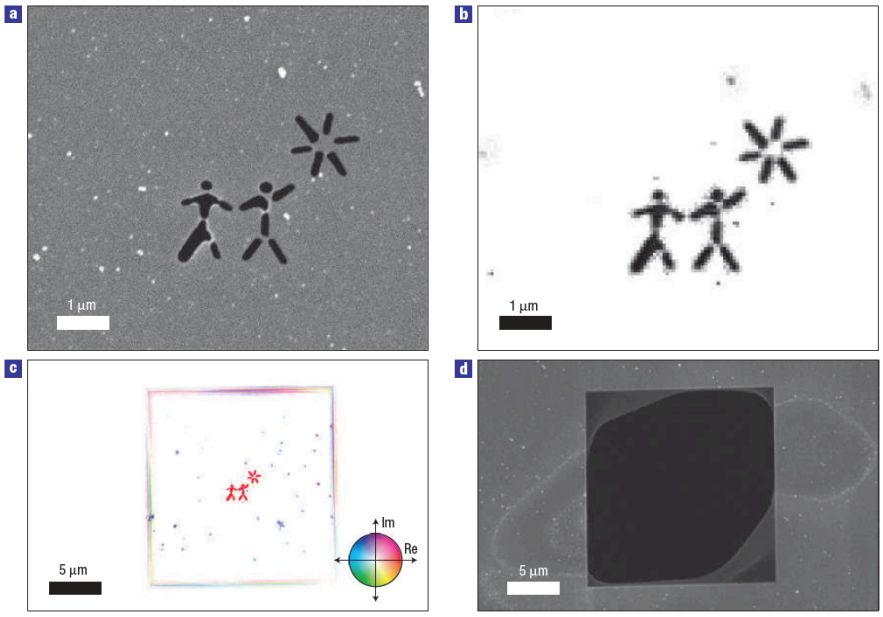

Figure 2.5: **Single particle reconstruction without damage** a, SEM image of the sample before exposure to the FEL beam. The 20-nm-thick sample was held in a square supporting window that is 20 μm wide. b,c, Image reconstructed, from the ultrafast coherent diffraction pattern of Fig. 2a, by phase retrieval and inversion using the Shrinkwrap algorithm (Marchesini et al., 2003). In b, the squared modulus of the retrieved complex image is shown on a linear greyscale from zero scattered photons per pixel (white) to $1.5x10^6$ scattered photons per pixel (black), and c shows the complex image, with complex values on the Argand plane mapping to color by saturation (absolute value) and hue (phase) according to the color wheel shown. Pixel size in the reconstruction equals 62 nm, corresponding to the half-period of the finest spatial frequency that can be recorded on our camera at 32 nm wavelength. The retrieved image clearly shows the silicon-window edge (in c), the FIB pattern and putative dirt particles, many of which correspond to the SEM image and which have an 80-100° phase shift relative to the FIB pattern. d, SEM image of the test sample after the exposures to the FEL beam, showing the square 20 μm window and some remaining silicon nitride, as well as visible damage to the silicon support caused by the non-circular beam. Figure and caption taken from (Chapman et al., 2006).

Many other experiments utilized the newly developed abilities of FLASH, such as imaging single cells and viruses (Seibert et al., 2010) as well as aerosol experiments involving soot (Bogan et al., 2008). Although these experiments are beyond the scope of this work, the work allowed for two-dimensional cross-sections of cells or soot to be reconstructed from individual diffraction patterns collected from one pulse of FLASH, showing the potential of the new sources for biological applications.

Although FLASH allowed for some interesting experiments, the wavelengths offered were too long to be of direct interest to the crystallography community. However, higher-energy free electron lasers, such as the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory, offered the potential to operate at crystallography-relevant wavelengths. The first lasing of the LCLS occurred in April 2009, and the Atomic and Molecular Optics (AMO) beamline became available to users at the end of 2009 (Emma et al., 2010). The initial operation of the LCLS generated X-rays with up to 1800 eV energy (6.9-Å wavelength) with a repetition rate of 30 Hz and a flux of 10^{12}-10^{13} photons/pulse, before taking the optics in the account.

First experiments on material science and physics were carried out at the AMO beamline of the LCLS when it became operational (Hau-Riege et al., 2010; Young et al., 2010). However, the energies available at the AMO beamline were still not high enough to provide high-resolution structures for structural biology.

The beamline that would be of interest for X-ray protein crystallographers is the Coherent X-ray Imaging (CXI) beamline, which became available for users in February 2011 (Boutet and Williams, 2010). The CXI instrument will provide X-ray energies between 2 and 30 keV, with repetition rates up to 120 Hz and as much as 10^{12} photons/pulse. The CXI instrument could provide X-ray wavelengths as short as 1.3 Å using the first harmonic, which should allow atomic-resolution diffraction patterns to be recorded.

The size of the protein microcrystals and the intense pulses of the LCLS would destroy a protein microcrystal in one pulse, making one crystal usable for only one pattern. Consequently, the conventional method of X-ray protein crystallography, and the imaging of biological specimens in general, would require complete revision in order to accommodate the new sample types and instrumentation. The initial goals were to show the first proof of concept for X-ray protein crystallography at the new XFEL sources.

2.5 Objective and Hypotheses

The success of prior experiments utilizing XFELs leads to the possibility that XFELs could play a role in X-ray protein crystallography of membrane proteins. The short time scales of the X-ray pulses, and our first femtosecond nanocrystallography experiments, which show the diffract-before-destroy principle, leads to the notion that XFELs may offer a new avenue of structure determination of membrane proteins in which the intense X-ray pulses of XFELs are used to obtain high resolution diffraction patterns from microcrystals of membrane proteins. The use of XFELS may allow the diffraction patterns to be collected without the X-ray-induced radiation damage that normally accompanies the larger exposures needed to obtain the high resolution diffraction patterns.

The predominant hypothesis of the project was that the membrane protein microcrystals, which are more commonly found in protein crystallization screens than protein macrocrystals, could be used to determine high-resolution structures of membrane proteins. However, the membrane protein microcrystals could not be used with the currently available third-generation X-ray sources, as X-ray-induced radiation damage becomes the major problem. Consequently, the main objective of the project was to determine whether XFELs could be utilized for X-ray protein crystallography experiments in which the microcrystals of membrane proteins are used to obtain molecular,

or even atomic, resolution diffraction patterns. We had to show that the single-snapshot diffraction patterns could be used to reconstruct an electron-density map that would be without any ill-effects of X-ray-induced radiation damage. For the work, the intrinsic membrane protein Photosystem I was chosen as a model system.

Questions and Challenges Associated with the Objective

X-ray protein crystallography is a very mature technique, both theoretically and experimentally, which has had improvements to the technique developed over many decades. However, the novel instrumentation of the XFELs, and the use of protein microcrystals for structural studies, leads to interesting challenges in extending the X-ray protein crystallography to XFELs. Questions and challenges that immediately arise from consideration of the objective involve protein crystallization, sample introduction, data collection, data analysis, as well as radiation physics and chemistry.

The unique properties of XFEL sources could allow the use of much smaller protein crystals than conventionally allowed. It is well documented that protein microcrystals are more commonly encountered in protein crystallization screens than protein macrocrystals. However, the decades of work on protein crystallization were done to take the crystallization conditions that produced the microcrystals and to generate macrocrystals from adjacent areas of the crystallization phase diagram. Consequently, one major challenge lies in determining the reproducibility and quality of protein microcrystals and to understand more about the production of protein microcrystals. Since microcrystals of proteins were always considered a stepping stone toward the generation of macrocrystals, the prevalence of formation was never thoroughly evaluated. An additional question to address is how small the membrane protein crystals can be, while still exhibiting enough order to be useful in a crystallographic experiment, and how could the nanocrystals be observed and measured.

The properties of the unattenuated XFEL beam makes sample introduction and

data collection a novel process if X-ray protein crystallography is to be attempted at an XFEL source. The intense (unattenuated) X-ray beam of an XFEL destroys the sample, and therefore only one diffraction pattern could be collected from one protein microcrystal. New methods of bringing the protein microcrystals into the interaction region would be needed, as mounting individual microcrystals in a crystal loop, as is the case in conventional X-ray protein crystallography, would be impractical, since a newly mounted microcrystal would be needed for each pulse. Additionally, because no precession or oscillation method could be used to collect the data, novel methods of the intensity merging and structure factor extraction would be needed as well. The individual diffraction patterns would be "partial" reflections, and not integrate the intensities of individual reflections, as is the case in conventional X-ray protein crystallography.

As such, a large international collaboration of experimental and theoretical groups from around the world was established. Only the hard work of all people involved could allow for such an ambitious endeavor.

Chapter 3

MATERIALS AND METHODS

3.1 Isolation and Purification of Photosystem I for crystallization

Materials for cell lysing

MCM buffer: (20 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.4, 10 mM $CaCl_2$, 10 mM $MgCl_2$)

MMCM buffer: (20 mM MES pH 6.4, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 500 mM D-mannitol)

PMSF solution: 500 mM PMSF in dimethylsulfoxide (DMSO)

Acetone:

Materials for solubilization and anion-exchange chromatography

Lyophilized n-dodecyl-β-maltopyranoside (β-DDM), Glycon

A_{100} buffer: (A_X: 20 mM MES pH 6.4, 0.02% (m/v) β-DDM, X mM $MgSO_4$)

A_{150} buffer:

Materials for initial crystallization of PSI

G_0 buffer: (G_X: 5 mM MES pH 6.4, 0.02% β-DDM, X mM $MgSO_4$)

G_3 buffer:

G_{100} buffer:

Thylakoid preparation

Photosystem I was purified as described previously (Fromme and Witt, 1998) with modifications. *Thermosynechococcus elongatus* cells were grown in TE media (see

Appendix A) at 56° C using the 122-L bioreactor shown in Fig. 3.1, with 2% CO_2 added to air bubbled in at 10L/min, so as to support a higher cell density and faster growth rate. Low-light conditions ($\sim$ 50 μmol s^{-1} m^{-2}) were used, and the physiological status of the cells was monitored by measuring a spectrum from 400-800 nm on a daily basis. The specific growing conditions of the cells can be monitored by taking the 1^{st} and 2^{nd} numerical time-derivatives of the growth curves at wavelengths of 440 nm, 630 nm, 680 nm, and 730 nm; the 1^{st} derivative allows for determination of the growth rate – a decline in the growth rate is indicative of cells leaving the exponential growth phase – whereas the 2^{nd} derivative allows the cell vitality to be determined. In practice, taking the derivatives requires the sample to be quite homogeneous, and the measurements should be made in triplicate to reduce the uncertainty in the measurement. Phosphate and nitrate additions that were approximately 25% of the nominal value of the media were supplied on a weekly basis, or as needed, judging from the second derivative; ideally, the cells would be harvested minimally once per week. The cell density was monitored using the absorbance at 730 nm, and the cells were grown until a maximum density of approximately 0.8 g cells/ L. The cells were harvested by filtration through a tangential filtration system, frozen and stored in a dense pellet without cryo-protectants in a -80° C freezer.

A fresh volume of PMSF solution was made prior to starting the preparation. A 25-35 g quantity of frozen *T. elongatus* cells was defrosted and resuspended in a 250-mL centrifuge tube using <50 mL of MCM that had been warmed to 56° C using a water bath. The cells were re-suspended by vigorously shaking the cells in 150 mL of warm MCM in a closed centrifuge tube, and the base of the centrifuge tube was immersed in the 56° C water to maintain proper temperature. The cells should not be allowed to reach 4° C for any length of time because this would cause the cells to go through apoptotic pathways and produce many proteases that would degrade the cellular proteins. The cell suspension was centrifuged at 7,000 rpm (7,400 *x g*) for

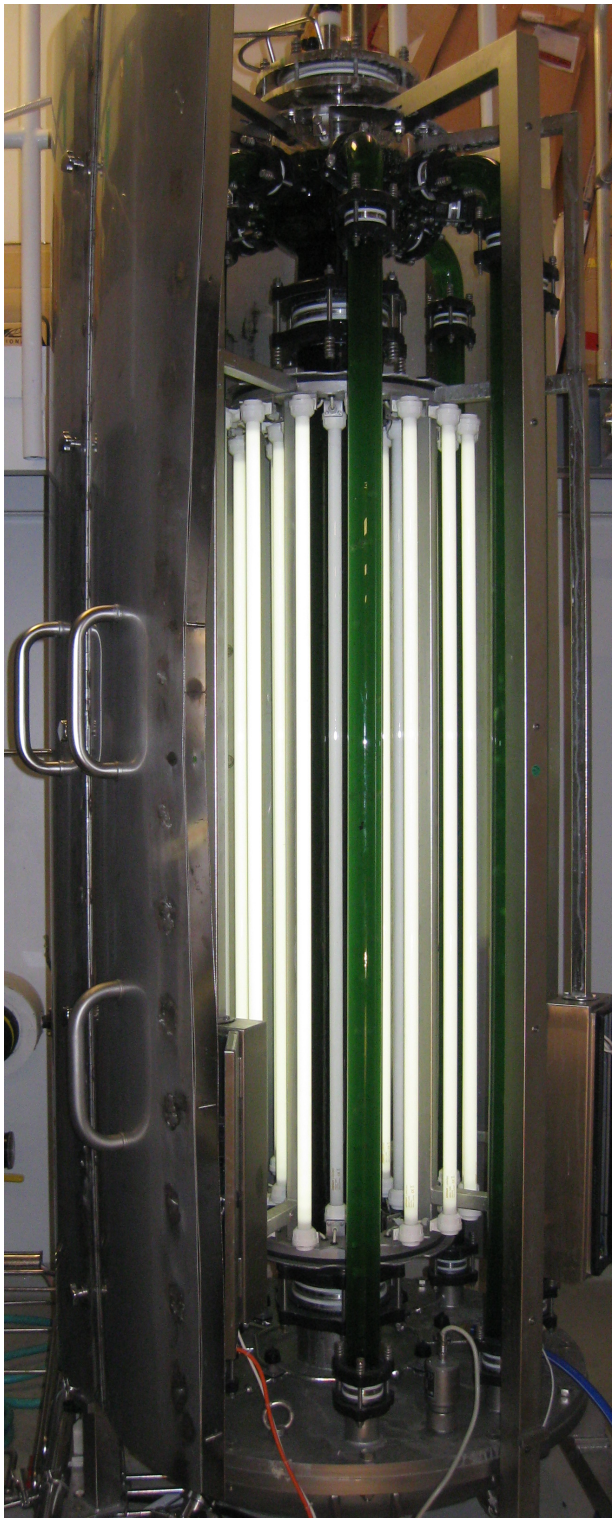

Figure 3.1: **Bioreactor used to grow thermophilic cyanobacteria.** The bioreactor used to grow the *Thermosynechoccus elongatus*. The bioreactor supports a volume of 122 L of cells and can be used to maintain a constant temperature of 56° while monitoring the airflow, light intensity, and pH of the cell culture.

10 min – all centrifugation steps in the thylakoid preparation were carried out using a fixed-angle rotor, SLA-1500 (Beckman Coulter), and a Sorvall RC-3C Plus centrifuge (Beckman Coulter) – at 24° C. During the centrifugation run, the microfluidizer (Microfluidics Model M-110L, Newton, MA) was prepared for cell disruption by passing water (2x) and MMCM (3x) through the system. The supernatant of the centrifugation was discarded and a brush was used to lightly remove the orange layer (consisting of *spirillum* cells, a symbiont that lives with the *T. elongatus* culture) from the top of the pellet. The pellet was resuspended in the smallest quantity of MMCM possible, typically 5-10 mL, and the solution was diluted to approximately 150 mL using MMCM, with 100 μL of PMSF stock added. The cell suspension was passed through a 16-32 mesh wire sieve before addition to the microfluidizer (Microfluidics Model M-110L, Newton, MA), to prevent large aggregates of cells from blocking the channel. The cell suspension was passed through the microfluidizer, with a standing pressure of approximately 10.5 kpsi, and a working pressure of approximately 9.5 kpsi. The microfluidizer lyses cells by forcing the cell through a narrow channel, which causes a large shearing force, rupturing the cell membrane plus cell walls. The solution was passed through the microfluidizer a second time, using the same pressure settings. The lysed cells were immediately placed on ice, divided into two centrifuge tubes, diluted to 250 mL each using MCM plus 200 μL of PMSF stock, total. The lysate was quickly centrifuged at 14,000 rpm (29,774 *x g*) for 10 min at 4° C. The supernatant was removed, except a small portion which was stored at -20° C for future characterization. Each pellet was resuspended in approximately 25 mL of ice-chilled MCM with 40 μL of PMSF stock solution. The solutions were diluted to 250 mL with ice-cold MCM and 160 μL of PMSF stock solution. The diluted solutions were centrifuged at 11,000 rpm (18,270 *x* g) for 10 min at 4° C. The pellets were resuspended again as described above and centrifuged at 11,000 rpm (18,270 *x* g) for 10 min at 4° C. The previous step was repeated one additional time. The supernatant was discarded, and the pellets, containing

the thylakoids, were resuspended in 10 mL of MCM, which is ultimately diluted to approximately 40-50 mL using MCM.

Determining the chlorophyll concentration in the sample

Chlorophyll assays are done in order to determine the concentration of chlorophyll in the sample, allowing a rough estimate of the amount of the photosynthetic proteins in the sample. Chlorophyll is extracted from the sample using an 80% (v/v) acetone solution. A mass of 635 mg of acetone is added to an eppendorf tube using an analytical balance. A volume of sample and water will subsequently be added, but the total volume of the sample plus water must equal 200 μL, *e.g.* if 1 μL of sample is added then 199 μL of water must be added prior to the addition of the sample. *A (635 mg/ 195 μL/ 5 μL) chlorophyll assay should be read as: 635 mg of acetone with 195 μL of water and 5 μL of sample*. Chlorophyll assays are usually done in triplicate, so three eppendorf tubes with the appropriate amount of acetone and water are prepared. A blank is made in which 635 mg of acetone and 200 μL of water are added to an eppendorf tube.

For chlorophyll assays using 0.5-1.0 μL of sample and especially highly viscous samples, 0.5 μL or 1.0 μL glass capillaries are used. The capillaries should not be handled without gloves and the end that is handled should not be placed into the sample, thereby avoiding contamination of the sample. One end of the capillary is placed below the sample surface in order to draw the liquid into the capillary, being cautious to place only the minimal amount of the capillary into the sample as is necessary. Once enough sample volume is drawn into the capillary such that the entire capillary is full, the capillary is withdrawn from the solution. The end of the capillary that was placed into the sample is wiped clean of sample, and the capillary and sample are placed into an eppendorf tube containing the acetone and water. The eppendorf tube is tightly closed and vigorously shaken until the color inside the glass capillary is indistinguishable from the bulk solution. The above procedure is repeated for three different trials.

The three chlorophyll samples and a blank are mixed with a vortex and centrifuged at 14,100 *x g* for one minute. For larger volumes and less viscous mixtures, pipettes can be used to add the sample to the acetone-containing eppendorf tube.

A UV-Vis spectrophotometer is used to record a spectrum from 400 to 800 nm, and the absorbance at wavelengths of 664 and 710 nm are determined to calculate the chlorophyll concentration using the following formula:

$$C_{\text{Chl}} = \frac{A_{664} - A_{710}}{76780\ (M^{-1}cm^{-1}) \cdot l} \cdot \frac{1000(\mu L)}{V_{\text{sample}}} \tag{3.1}$$

where C_{Chl} is the chlorophyll concentration of the sample in molar, A_{664} and A_{710} are the absorbance measurements at 664 nm and 710 nm, respectively, l is the path length in centimeters, 76780 $\text{M}^{-1}\text{cm}^{-1}$ (Porra et al., 1989) is the molar absorption coefficient at 664 nm wavelength of chlorophyll in acetone, and V_{sample} is the volume of the sample, in microliters, used in the chlorophyll assay. The chlorophyll concentration is taken as the mean value of the three measurements and the standard deviation was used to evaluate whether the chlorophyll assay needed to be repeated. A value of $\sigma_{Chl} < 3\%\ \mu_{Chl}$ (standard deviation, σ_{Chl}, and mean values of the chlorophyll assay, μ_{Chl}) was found to be a suitable cutoff. Most of the steps in the protein crystallization and nanocrystallography experiments are predicated on precisely knowing the chlorophyll concentration, so stricter limits to the standard deviation are applied ($\sigma_{Chl} < 1\%\ \mu_{Chl}$).

Solubilization and extraction of PSI from thylakoids

A (635 mg/ 199.5 μL/ 0.5 μL) chlorophyll assay was always done in triplicate on the thoroughly re-suspended sample. The sample was diluted to 1 mM chlorophyll concentration using MCM and the β-DDM concentration was adjusted to 0.6% (m/v) final β-DDM concentration. The solution containing the β-DDM was added slowly, while

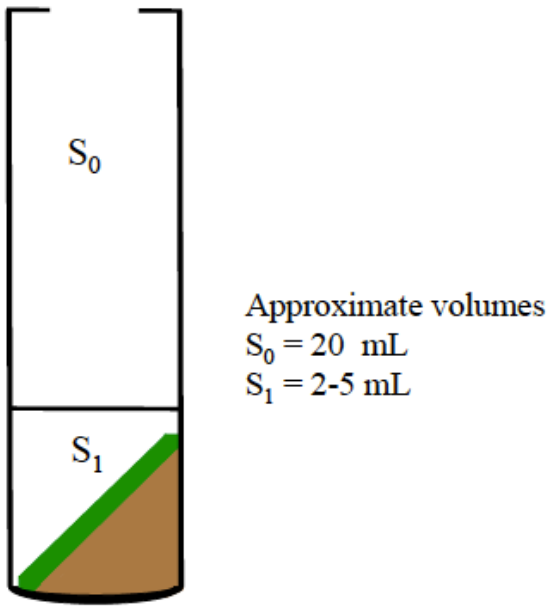

Figure 3.2: **Schematic of the result of the ultracentrifugation step for PSI** Schematic view of the pellet and supernatant produced during the ultracentrifugation step of the PSI preparation protocol. The PSI_T has a large enough sedimentation coefficient to sediment into the pellet during the ultracentrifugation step, whereas the other solubilized membrane proteins and remaining phycobiliproteins remain in the supernatant. However, some of the PSI_T remains in a 2-5 mL portion of the supernatant—referred to as S_1 —directly above the pellet, and S_1 is enriched in PSI_T with respect to the other soluble membrane proteins. The top-most portion of the supernatant contains mostly smaller solubilized membrane proteins, and consequently is discarded.

gently stirring and avoiding the formation of bubbles, so as to avoid the denaturing of the protein. After dilution, the solubilization was allowed to proceed for 45 min at room temperature.

Upon completion of the solubilization step, the solution was centrifuged at 50,000 rpm (184,000 *x g*) using a Ti 70 rotor (Beckman Coulter, Part No. 337922) and an Optima-100K centrifuge (Beckman Coulter, Part No. 393253) for 105 min at 4° C. The PSI trimer is such a large molecule that the ultracentrifugation spin will cause the protein to sediment during the centrifugation step. A schematic of the result of the ultracentrifugation step is shown in Fig. 3.2. The top portion of the supernatant, termed S_0, was discarded, as it had the lowest relative amount of PSI_T with respect to other solubilized membrane proteins and core phycobiliproteins. The bottom potion of the

supernatant, termed S_1, would contain a much higher percentage of PSI_T than S_1, but the pellet itself contained the highest amount of PSI_T relative to the other solubilized membrane proteins, as only PSI_T is large enough to sediment at the force applied in the given amount of time.

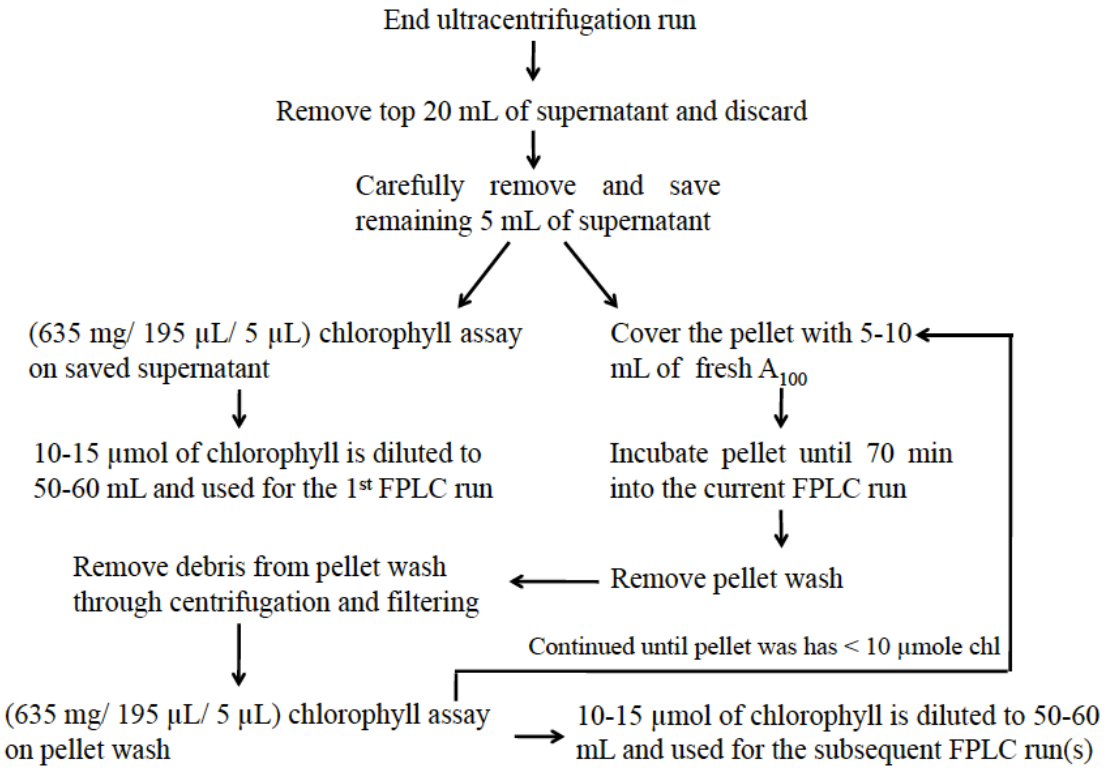

Figure 3.3: **General flow diagram explaining the pellet wash steps.** The incubation steps were done for the first 70 min of the current FPLC run(s). In steps with multiple exit arrows, both routes are taken in parallel.

Since most of the PSI_T was found in the pellet or S_1 portion of the supernatant, the first 20 mL of the supernatant of each centrifuge tube was discarded, and the remaining 5 mL of supernatant (referred to as the supernatant) was poured into to a graduated cylinder; ideally, a pipetting procedure would be done so that the PSI_T was not disturbed from the pellet, but the process takes much more time and for the majority of experiments (especially those done with only one individual), the supernatant was poured off. The pellet was washed using A_{100} and allowed to incubate at 4° C for approximately 2 hr in order to extract the PSI trimer from the pellet. The lengthy

washing and incubation time for the pellet using A_{100} was done during the FPLC run of the retained supernatant. A general overview of the pellet-washing process is shown in Fig. 3.3.

Preparing the anion-exchange column

Anion-exchange chromatography is used to separate the PSI trimers from other solubilized protein complexes present in the supernatant and pellet wash of the ultracentrifugation step. Q-Sepharose HP (GE Healthcare, 17-1014-01) is the anion-exchange resin used for the PSI trimer separations and has a quaternary ammonium as the functional unit with an agarose matrix. The column used is 20 cm in length and 50-mm inner diameter (XK 50/20, GE Healthcare, Part No. 18-1000-71) and has a total volume of 400 mL. The length of the resin bed inside the column should be between 10-11 cm in length, allowing the resin bed to have an approximate total volume of 200 mL. All washing and packing steps must be completed by the day before a PSI preparation to ensure the FPLC and column are fully functional.

The resin material is regenerated and cleaned after each PSI preparation. Initially, the regeneration and cleaning were done by washing the entire system with 200 mL of each of the following solutions, in the order presented: 0.1-1 M NaOH, filtered de-ionized water, 0.1 M HCl, filtered de-ionized water, 2 M NaCl, filtered de-ionized water, filtered de-ionized water, 50% 2-propanol in water, 100% 2-propanol, 50% 2-propanol in water, and filtered de-ionized water. The aqueous solutions were passed through the resin at 10 mL/min, the 50% 3-propanol solution was run at 5 mL/min, and the 100% 3-propanol solution was flowed at 2 mL/min.

Unfortunately, the hydrophobic washing steps could produce channels that would occlude portions of the resin from being thoroughly cleaned, when the resin was cleaned in the mantle. An alternate way to clean the resin was developed: before the organic solvent washes, the resin was removed from the mantle, and the resin was added in 35-

mL aliquots to 50-mL falcon tubes. The resin was centrifuged at 1100 rpm (257 *x g*) for 5 min at 22-24° C using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435) and the supernatant was discarded. The falcon tubes were filled to 45-50 mL with 100% 2-propanol and the centrifugation is repeated. The supernatant is examined for color—the color would likely be slightly orange or brown and is due to carotenoids—and the washing steps are continued with 100% 2-propanol until the color is completely removed. Once the color is removed, the resin is washed in an analogous manner until the supernatant becomes odorless—at this point the majority of the 2-propanol will have been removed—and the resin is re-suspended and added back to the mantle. The slurry containing the resin is slowly poured into the mantle using a glass rod as a guide in order to prevent the slurry from being poured unevenly. When the entire volume of resin has been added to the mantle, the glass rod is used to mix the entire slurry to homogenize the resin. The column is re-assembled and filtered, de-ionized water is passed through the media at 10 mL/min until the column is packed, at which point the plunger is dropped down until being approximately 1 mm below the top of the column bed. The filtered de-ionized water is passed through again at 10 mL/min, with the process repeated until the resin now longer packs into a smaller bed. The length of the resin bed should be at 10-11 cm in length. The separation program used to separate the PSI trimer is based on the resin bed being in this length range: A shorter bed will not lead to the desired separation, whereas a longer bed would cause the PSI trimer to elute at a later time, and the fraction collection program would need to be updated.

Cleaning the resin by centrifugation and re-suspension was very time consuming (approximately 5-8 hr, depending on the preparation) and produced similar results to the cleaning protocol in which the resin remained within the mantle. A few alterations to the protocol were made to arrive at the current cleaning protocl. The resin is briefly removed from the mantle before the 2-propanol washing step, and the walls of

the mantle are cleaned with a sponge. Additionally, the resin was re-suspended during the 2-propanol washing steps multiple times, and during the final wash with water, as well. This final procedure proved the most effective way of the three protocols for cleaning the column.

PSI purification through anion-exchange chromatography

During the ultracentrifugation step, the cold-water bath is connected to the outer jacket of the column, and the water temperature is set to 4° C. The column is equilibrated by washing the resin with 200 mL of A_{150} followed by washing the resin with 200 mL of A_{100}.

Time(min)	Percentage A_{150}	Flow rate (mL/min)
0	0	10
30	0	10
100	100	10
100.5	0	10
121	0	10

Table 3.1: **Elution gradient used for PSI.** The linear gradient used to separate Photosystem I from cell lysate using a strong anion-exchange media (Q-sepaharose HP).

A (635 mg acetone/ 195 μL water/ 5 μL sample) chlorophyll assay was performed on the detergent extract (*i.e.* the retained supernatant of the ultracentrifugation step). An aliquot containing 10-15 μmole of chlorophyll, from the detergent extract, was diluted to 60 mL and loaded onto the anion-exchange column using a 10 mL/min flow rate through the same pump used for A_{100} (pump A). After the sample is applied, the program with a constant flow rate of 10 mL/min for all steps, shown in tabular form in Table 3.1, was started by washing the column for 30 min with A_{100}. Using a linear gradient, the elution buffer proceeded from A_{100} to A_{150}, with an increase of 1 mM $MgSO_4$ per minute for 50 min, followed by a 20-min washing step at A_{150}. The resin is subsequently washed with 200 mL of A_{100} to re-equilibrate the resin for any subsequent runs. Fraction collection is programmed to begin at 30 min (the start of the

linear gradient) and end at 100 min (the end of the A_{150} wash). Fig. 3.3 shows the PSI trimer band as it appears on the resin once the other proteins of the detergent extract have eluted from the resin.

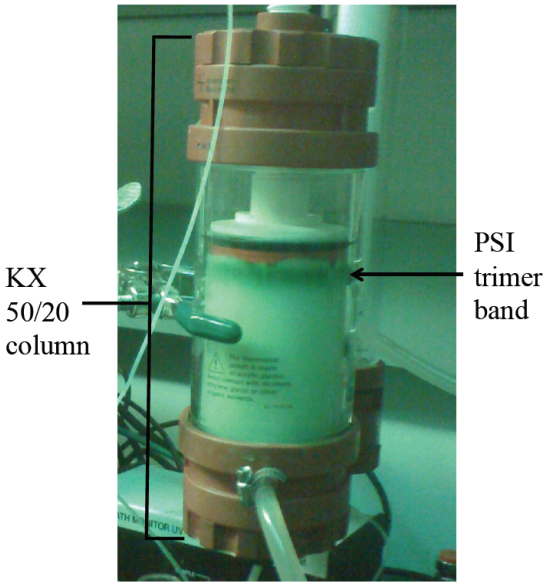

Figure 3.4: **Photosystem I trimer band.** An image taken of the protein mixture separation once most protein components (other than the Photosystem I trimer) have eluted. The dark green band is the Photosystem I trimer, which elutes at a [$MgSO_4$] concentration of approximately 135 mM.

The AEC purification yielded three peaks, as discussed in Section 4.1.3. The third peak as shown in Fig 4.1, which eluted at an approximate $MgSO_4$ concentration of 135 mM, contains the PSI trimer. The fractions containing the third peak were collected in a graduated cylinder – a total volume of approximately 200 mL if 13 μmole of chlorophyll were loaded – and the sample was homogenized by covering the graduated cylinder in plastic wrap and slowly inverting the cylinder multiple times. The volume of the fractions was recorded and a 200-μL aliquot was saved for a (635 mg acetone/ 150 μL water/ 50 μL sample) chlorophyll assay and SDS-PAGE characterization. However, care was taken to avoid collecting any fractions that would contain significant amounts of the second peak, which contains the PSI monomer and PSII monomers and dimers.

Another FPLC run would be done using the retained supernatant if 10-15 μmole

of chlorophyll was still retained after the first FPLC run. The remaining solution would be loaded onto the resin and an FPLC run would be initiated, as described above. Once the retained supernatant of the ultracentrifugation step was exhausted, the wash of the ultracentrifugation pellet was removed and prepared for a FPLC run by centrifuging at 5100 *x g* for 5 minutes using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435) with subsequent filtration through a 0.22 μm syringe filter. A (635 mg acetone/ 195 μL water/ 5 μL sample) chlorophyll assay was performed on the filtered pellet-wash sample. An aliquot containing 10-15 μmole of chlorophyll, from the first wash of the pellet, was loaded onto the column, and a FPLC run was done using the same parameters as the first run(s). The ultracentrifugation pellet was washed again using A_{100} and allowed to incubate in A_{100} at 4° C for approximately 2 hr, until the current FPLC run finished. Additional aliquots containing 10-12 μmole of chlorophyll from the first pellet wash were diluted and used for subsequent FPLC runs, as necessary.

The ultracentrifugation pellet was washed until the wash no longer contained a considerable chlorophyll concentration, typically by the third wash. A total of 4 FPLC runs are typical for a preparation starting with about 35 g of *T. elongatus* cells.

Concentrating the protein

Fractions containing the third peak from all FPLC runs were collected and concentrated using a 400-mL stirred ultrafiltration unit concentrator (Stirred cell model 8400, Millipore, Part No. 5124) using a 100 kDa cutoff membrane (Ultrafiltration Membrane Disc Filters, Pall Life Sciences, Part No. OM100076). The membrane used was first incubated in water, followed by two incubations in A_{100}, with each incubation for 20 min at 4° C. The ultrafiltration unit was placed on a stirring plate in a refrigerator at 4° C, and the PSI protein-detergent solution was gently stirred during the concentrating steps. Less than 5 psi of pressure was applied to the concentrator to push filtrate

through the membrane during the concentrating steps. The filtrate was collected in a beaker as a precaution to protect against sample loss stemming from the rupturing of the membrane. To determine whether a small leak was present in the membrane, small amounts of the filtrate at different time points were taken and examined visually for green color.

If a green color was present, and the intensity of the green color did not decrease with time, UV-VIS spectra would be taken of the filtrate samples to determine the content of the filtrate. When a leak was discovered, a new membrane was immediately prepared by washing plus incubating in water for 10-20 min. The new membrane was immediately added to the ultrafiltration unit and two 100-mL portions of A_{100} would be flowing through the membrane under pressure, while never allowing the membrane to completely dry.

The sample was concentrated to a final volume of approximately 10 mL. The volume was measured by drawing the solution into a 10 mL pipette, and the volume was recorded. The used pipette is thoroughly wrapped with plastic wrap and stored at 4° C. One method to reduce sample loss would be to determine the density of the solution and measure the mass of the sample.

Final PSI purification through initial crystallization

The total chlorophyll concentration of the concentrated ($\sim$ 10 mL) sample is calculated using the known amount of chlorophyll added to the concentrator and the measured volume – the total amount of chlorophyll lost in the filtrate should be subtracted from the added chlorophyll; however, the amount of chlorophyll lost in the filtrate was always negligible. The PSI is crystallized by reducing the ionic strength of the solution until the attractive forces between the PSI trimers overcomes the PSI-solvent interaction. Consequently, the sample is diluted using G_0 until a low final salt concentration is achieved. Once the salt was diluted, the protein would be increased until crystallization.

The amount of G_0 buffer that is needed to dilute the sample to 10 mM $MgSO_4$ is calculated, estimating the initial solution as having 135 mM $MgSO_4$and using the dilution formula $M_1V_1 = M_2V_2$, where M_1 is the concentration of the salt before diluting at a volume of V_1 with desired salt concentration M_2 at a volume of V_2. In order to accurately know the final salt concentration, the ionic strength of the PSI solution should be measured directly using a conductivity meter and a standard curve.

To dilute the sample, the G_0 buffer was added drop wise to the sample with approximately one second between drops, as the sample was being gently mixed (to prevent bubbles) on a stir plate at room temperature. After thorough mixing, a 25-μL aliquot of the sample was removed for a (635 mg acetone/ 195 μL water/ 5 μL sample) chlorophyll assay in triplicate. The known volume and chlorophyll concentration was used to calculate the total amount of chlorophyll, which could be subsequently used to calculate the P_{700} concentration assuming 96 chlorophyll/ P_{700} (Jordan et al., 2001). Upon dilution, the sample was concentrated using an initial pressure less than 5 psi. The rate of concentrating was initially allowed to be quick, but it was slowed with time to allow for slower crystallization of the PSI sample. The pipette used to measure the protein mixture volume is washed using the diluted protein mixture. The filtrate was again collected and inspected for a compromised membrane.

After the sample was concentrated to 25-50 mL, the stirring was stopped, and the pressure line was removed from the concentrator. The pressure inlet of the stirred ultrafiltration unit was tightly covered using plastic wrap, and the pressure outlet and filtrate outlet were left open. The tubing of the filtrate outlet was taped to the inside of the beaker facing downward, allowing for hydrodynamic pressure to slowly concentrate the sample for 12-24 hr until 5-10 mL of total volume remained, corresponding to a chlorophyll concentration between 5 and 12 mM. Crystals of PSI will be present on the membrane immediately and are allowed to grow for 12 hours.

The Photosystem I crystallites were harvested from the membrane by washing the membrane with 1-mL aliquots of the retentate using a 1-mL pipette. The ultrafiltration setup was held at an angle up to 10° so the crystals washed from the membrane would settle in the remaining solution at the bottom of the setup. The ultrafiltration unit was slowly rotated to ensure that portions of the membrane were not allowed to dry out. Once all of the crystals were removed from the membrane, the PSI crystals re-suspended in the retentate would be removed for settling experiments.

The PSI sample was placed into eppendorf tubes in 1 mL aliquots for settling experiments. Two methods have been utilized for the settling experiments, one is to do the settling experiments in G_{10} (or whichever final solution the crystals were grown in) and wash the crystals after settling, while the second method is to wash all crystals together and place into G_0 buffer, at which point the settling is done with the crystals in G_0. The reasoning behind the different methods for the settling experiments will be given below, but the nature of the experiments will dictate which method is most effective and appropriate. The case for doing the settling experiments in G_0 is immediately evident: reducing the [$MgSO_4$] to zero in the solution will minimize the solubility of the PSI trimers. This would protect the nanocrystals from dissolving. The smallest microcrystals will have a higher thermodynamic driving force to dissolve than the larger PSI crystals, due to the smaller size and smaller radius of curvature at the interface between the different phases; consequently, the number of nanocrystals should be maximized in this experiment. However, the case for settling in G_{10} is to remove the smallest protein crystallites from the crystallite suspension, and this was the preferred method for the most recent femtosecond nanocrystallography experiments in Feb 2011 at the CXI beamline, see Section 3.7. Initial evaluation of femtosecond nanocrystallography from Dec. 2009 indicated that the smallest crystallites may not scatter to the high

resolution desired in the high-resolution structural studies. The presence of the smallest crystallites will cause the estimation of the crystallite density—3.7—to be less valid. Settling the crystals and crystallites in G_{10} was used as a method that would allow for the loss of some of the smallest microcrystals, effectively serving as a low-end filter for the PSI samples. Both methods have been used and have shown similar results. However, as higher-energy X-rays will be used in future femtosecond nanocrystallography experiments, the optimal size of the PSI microcrystals has to be experimentally determined.

PSI settling experiments

Once the PSI samples were placed, as 1-mL aliquots, into several eppendorf tubes, the samples were re-suspended. After 10 minutes, the supernatant was carefully removed from the settled pellet and placed into a new eppendorf tube. After 20 minutes, the supernatant was again removed from the settled pellet and placed into a new eppendorf tube. This procedure was repeated for 30 min, 40 min, and 60 min settling steps. Upon removal of the supernatant from settled pellets, the pellets were washed with G_3 and centrifuged at 1,000 *x g* for 1 min, with the supernatant discarded. The 10 min settled crystals were combined into three eppendorf tubes, and all other settled crystals were combined into one eppendorf tube per settling step. The PSI crystals were washed with G_0 and centrifuged at 1,000 *x g* for 1 min, with the supernatant discarded. As long as the centrifugation step is done in a 4° C room, the centrifugation time can be extended by a few minutes, in order to make a slightly denser pellet, so less sample is lost during the wash steps (as the crystallites most prone to be lost are the nano- and microcrystals that are of interest) The crystals were washed with G_0 again, and the after centrifugation, the crystals were stored at 4° C in the dark.

Remaining soluble PSI sample was added to the ultrafiltration cell to be further concentrated; no pressure was used to concentrate the sample. The gas inlet was

closed, the pressure release valve was opened, and the filtrate outlet was fixed in a completely vertical position, such that hydrostatic pressure could concentrate the samples very slowly. Upon becoming concentrated, the crystals were harvested from the membrane when the volume was maximally 3 mL G_{10}, and the same settling procedure, as described above, was performed. The ultrafiltration cell was kept assembled, and the membrane was stored under 50-100 mL of G_{100} to initially clean the membrane and prevent the membrane from drying out so that it could still be used in subsequent PSI preparations. The membrane is cleaned by passing approximately 100 mL of 0.01 M NaOH through, under pressure, followed by 200 mL of de-ionized water.

3.2 Re-crystallization of Photosystem I from crystal pellets

Materials

Sephadex G50 fine

Millipore water

1 mL syringes with removable needle

Cotton wool (cotton balls)

Test tubes

Preparing the columns

The re-crystallization implementing spin columns was used to mimic batch crystallization to low salt conditions, such as G_{0-5}. For discussion of why the technique was developed using gel-filtration media and why re-crystallization of PSI is necessary, please refer to Section 4.3.

An amount of 1 g of Sephadex G50 fine (GE Healthcare, Part No. 17-0042-01) was incubated for 24 hr at room temperature in 20 mL of Millipore water (always a 1 g: 20 mL ratio). Cotton wool was used to make a plug at the bottom of a 1 mL syringe,

with removable needle, using the syringe plunger (attempted multiple times until the cotton wool sticks to the bottom). The cotton wool weighed approximately 4 mg and occupied a volume less than 0.05 mL in the syringe. The syringe was placed into a glass test tube. The equilibrated resin was slurried, and using a Pasteur pipette, the syringe was filled to the top with the slurry (after settling, this was about 1mL of settled column material), by placing the suspension on top of the cotton wool plug (the tip of pasteuer pipette was placed near the bottom while filling to avoid air bubbles). The liquid was allowed to drip out of the syringe, and the eluant was monitored for turbidity. If the eluant was turbid, the column would have been rebuilt, as the G50 fine would be eluting out of the syringe through the cotton wool. The syringe was refilled, as needed, with slurried resin until the resin filled the syringe between 0.9 and 1.2 mL. The water was allowed to elute out without allowing the material to dry out. The resin was equilibrated with 2-3 mL of the final crystallization solution (G_{0-5}). The column was allowed to run until there was no eluant flowing out the bottom. The columns were centrifuged at 1500 rpm (478 *x g*) for 2 min at 22-24° C using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435). This caused the resin to shrink in volume to approximately 0.6-0.75 mL and to appear dry. The spin columns were covered and kept at 4° C until needed.

Preparing the PSI for re-crystallization

An eppendorf tube is labeled and the mass recorded. A volume of 100-500 μL of re-suspended PSI-crystal sample is added to the labeled eppendorf tube. The sample was centrifuged at 14,100 *x g* for 1 min at room temperature, and the supernatant was discarded. Subsequently, the sample was centrifuged at 14,100 *x g* at room temperature for 1 min, with the remaining traces of supernatant discarded two additional times. The mass of the eppendorf tube plus PSI pellet was recorded and the mass of the PSI pellet was calculated. An aliquot of G_{100} was added to the pellet such that the amount

of microliters of G_{100} was equal to the number of milligrams of crystal pellet. The solution was re-suspended to dissolve the crystals, and a 0.5-μL aliquot of the sample was visually inspected under an optical microscope to determine whether all crystals were dissolved. If crystals remained, 1-μL aliquots of G_{50} were added to the sample, with thorough mixing, until all crystals dissolved. A (635 mg acetone/ 199.5 μL water/ 0.5 μL sample) chlorophyll assay was performed in triplicate. The PSI sample was diluted to the desired concentration between 100 and 1000 μM chlorophyll using G_{50}, and 200-μL aliquots of the sample were added to each spin column. The spin columns and sample were immediately centrifuged at 1,500 rpm (478 *x g*) for 2.5 min at 22-24° C using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435). *NOTE: the setups must be transferred as quickly as possible to the centrifuge to avoid the formation of crystals or precipitate on the resin*. After centrifugation, 2-μL aliquots of the eluant were immediately inspected using an optical microscope, and images were recorded for future reference. The remaining eluant of the spin column re-crystallization trials was stored at 4° C and visually inspected with images taken after time delays ranging from 1 hr to 48 hr. A correlation between the frequency of imaging the re-crystallized sample and the growth rate was discovered, but the reader is referred to section 4.3 for details.

The re-crystallized samples were tested for crystallinity by imaging with cross polarizers as shown in Fig. 4.14, and SONICC as shown in Fig 4.16. The re-crystallization samples were monitored at various time points using optical microscopy and birefringence.

3.3 Crystallization of the Photosystem I-Ferredoxin complex

Materials

M_{100} buffer: (M_x: x mM $CaCl_2$, 100 mM HEPES pH 7.5, 0.02% (m/v) β-DDM)

M_{50} buffer:

1 M $CaCl_2$:

1 M HEPES pH 7.5:

K_{500} buffer: (K_x: x mM $CaCl_2$ plus 500 mM HEPES pH 7.5 solution

0.1% (m/v) β-DDM:

70 % (w/w) PEG 400:

Purified Ferredoxin (Fd) from *Synechocystis sp.* 6803:

Photosystem I crystals in G_0:

Preparing the protein and the crystallization solutions

Although the method to grow large PSI-Fd crystals was previously established (Fromme et al., 2002), the purpose of the PSI-Fd work was to use vapor-diffusion experiments to map out the phase diagram and to use this information to deduce new crystallization methods that could be used to produce nanocrystals of the PSI-Fd complex. The protocol used to grow the PSI-Fd co-crystals was adapted from (Fromme et al., 2002). An eppendorf tube was labeled and its mass recorded. A re-suspension of either the 10-min or 20-min settled PSI crystals in G_0 was made, and an approximately 200 μL aliquot of the PSI-crystal slurry was transferred to the labeled eppendorf tube. The sample was centrifuged at 14100 *x g* for one minute and the supernatant was removed.

Subsequently, the sample was centrifuged at 14100 *x g* with removal of the remaining supernatant two additional times; however, a 10-μL pipette was used to carefully remove the supernatant during the final two runs. The mass of the eppendorf tube plus sample was measured, and the mass of the PSI was calculated. A volume of M_{100} was added such that the number of microliters of M_{100} added was equal to the number of milligrams of PSI. The sample was re-suspended multiple times to try to ensure all of the Photosystem I crystals dissolved (*NOTE: the solubility of PSI in $CaCl_2$ solution is greater than in $MgSO_4$ solution; therefore the crystals should dissolve).* A 0.5-μL aliquot of the sample was visually examined under a light microscope to ensure all crystals were dissolved. If crystals were present, 1-μL aliquots of M_{50} were added until crystals were no longer observed under the microscope, but the crystals were found to dissolve in the initial M_{100} in every experiment. The sample is taken and centrifuged at 14100 *x g* for 2 min to remove bubbles, and a (635 mg acetone/ 199.5 μL water/ 0.5 μL sample) chlorophyll assay was performed in triplicate using the sample. The mean value was used as the chlorophyll concentration. The mass of the PSI sample was determined, and assuming a density of 1 mg/mL, the volume of the sample was determined. The P_{700} concentration was determined from the chlorophyll concentration. The PSI solution contained 50 mM $CaCl_2$, 50 mM HEPES pH 7.5, 0.02% (m/v) β-DDM, 0% (w/w) PEG400, and 0 mM Fd. The PSI solutions for the crystallization experiments were set up using 50 mM $CaCl_2$, 50 mM HEPES pH 7.5, 0.013% (m/v) β-DDM, 4.7% (w/w) PEG400, and a P_{700}:Fd ratio of 9:10, with the P_{700} concentration varied between 5 and 90 μM for vapor-diffusion experiments. To calculate the amount of stock solutions needed to be added to the PSI-containing solution to arrive at the final concentrations for the crystallization tests, the following formulas were utilized by being implemented in a spreadsheet program:

The amount of $CaCl_2$ stock solution to be added was calculated as follows:

$$V_{\mathrm{CaCl_2,\ stock}} = \frac{V_{\mathrm{total}} \cdot C_{\mathrm{CaCl_2,\ final}} - V_{\mathrm{P_{700}, initial}} \cdot C_{\mathrm{CaCl_2,\ initial}}}{C_{\mathrm{CaCl_2,\ stock}}} \tag{3.2}$$

where $V_{CaCl_2,stock}$ is the volume of $CaCl_2$ stock solution to be added, V_{total} is the final volume of sample for the crystallization experiments, $C_{CaCl_2,final}$ is the concentration of $CaCl_2$ in the crystallization mixture, $V_{P700,initial}$ is the initial volume of the P_{700} mixture, from the mass difference measurement, $C_{CaCl_2,initial}$ is the concentration of $CaCl_2$ in the P_{700} mixture, and $C_{CaCl_2,stock}$is the concentration of $CaCl_2$ in the stock solution.

The amount of HEPES pH 7.5 stock solution to be added was calculated as follows:

$$V_{\mathrm{HEPES,\ stock}} = \frac{V_{\mathrm{total}} \cdot C_{\mathrm{HEPES,\ final}} - V_{\mathrm{P_{700},\ initial}} \cdot C_{\mathrm{HEPES},initial}}{C_{\mathrm{HEPES,\ stock}}} \tag{3.3}$$

where $V_{HEPES,stock}$ is the volume of HEPES, pH 7.5, stock solution to be added, V_{total} is the final volume of sample for the crystallization experiments, $C_{HEPES,final}$ is the concentration of HEPES, in the crystallization mixture, $V_{P700,initial}$ is the initial volume of the P_{700} mixture, from the mass difference measurement, $C_{HEPES,initial}$ is the concentration of HEPES in the P_{700} mixture, and $C_{HEPES,stock}$is the concentration of HEPES in the stock buffer solution.

The amount of β-DDM stock mixture to be added was calculated as follows:

$$V_{\beta\mathrm{DDM,stock}} = \frac{V_{\mathrm{total}} \cdot C_{\beta\mathrm{DDM,\ final}} - V_{\mathrm{P_{700},\ initial}} \cdot C_{\beta\mathrm{DDM},initial}}{C_{\beta\mathrm{DDM,\ stock}}} \tag{3.4}$$

where $V_{\beta DDM,stock}$ is the volume of β-DDM stock mixture to be added, V_{total} is the final volume of sample for the crystallization experiments, $C_{\beta DDM,final}$ is the concentration of β-DDM in the crystallization mixture, $V_{P700,initial}$ is the initial volume of the P_{700} mixture, from the mass difference measurement, $C_{\beta DDM,initial}$ is the concentration of β-DDM in the P_{700} mixture, and $C_{\beta-DDM}$ is the concentration of β-DDM in the stock mixture.

The amount of PEG400 stock solution to be added was calculated as follows:

$$V_{\text{PEG400, stock}} = V_{\text{total}} \cdot C_{\text{PEG400, final}} C_{\text{PEG, stock}} \tag{3.5}$$

where $V_{\text{PEG400, stock}}$ is the volume of PEG400 stock solution to be added, V_{total} is the final volume of sample for the crystallization experiments, $C_{\text{PEG400, final}}$ is the concentration of β-DDM in the crystallization, and $C_{\text{PEG, stock}}$ is the concentration of β-DDM in the stock mixture.

The volume of water added was calculated as follows:

$$V_{\text{water}} = V_{\text{total}} - \sum_i V_{i,\text{ solutes}} \tag{3.6}$$

where V_{water} is the volume of water needed to reach the final volume, V_{total} is the total volume, and $\sum_i V_{i,\text{ solutes}}$ is a summation of the volumes of every solute added, assuming ideal mixing of the solutions.

The solutes were added in the following order: $CaCl_2$, HEPES, β-DDM, water, PSI, Fd, PEG400. Before the Fd is added, the setups must be taken into a room that had very dim light and narrow-band green-light filters placed upon any lights (green room)

in order to prevent the photoinduced dissociation of Fd from PSI. All subsequent steps must be done in a green room. The $CaCl_2$, HEPES, β-DDM, water, PSI and Fd were added and mixed with a pipette. The mixture was allowed to equilibrate for one hour in the dark in order to allow the Fd to dock with the PSI, and then the PEG400 solution was added. The mixture was centrifuged at 14100 *x g* for 2 min to remove debris and the supernatant was transferred to a clean eppendorf tube for use in subsequent crystallization experiments.

Photosystem I – Ferredoxin crystallization experiments

Vapor-diffusion crystallization experiments were used to gain insight into the phase diagram of the PSI-Fd co-crystallization. The [$CaCl_2$], [HEPES pH 7.5], [β-DDM], and [PEG400] in the crystallization drop were kept constant, and the [PEG400] or [$CaCl_2$] in the reservoir was varied. The crystallization droplet was taken from the pre-made PSI sample and not mixed with the reservoir solution. Aliquots of 0.5-1 μL of the sample were used per trial in 24-well hanging drop plates. The 0.5-μL droplets were used because the small volume causes a small radius of curvature, which produces a higher vapor pressure in the droplet than in the larger drops, causing the evaporation rate to be higher than in the case for larger droplets.

The crystallization trays were kept in the complete dark at 22° C and monitored every day for 2-3 days using an optical microscope with low-intensity green light, followed by examination once a week. Seeding experiments were carried out by crushing a preformed, large PSI-Fd co-crystal using a hair, and sliding the portion of the hair used to crush the crystal through another crystallization drop, before closing the well.

Once appropriate conditions were determined for producing nano- and micro-crystals of PSI-Fd, batch crystallization would be used to replicate the conditions in larger volumes more appropriate for the LCLS experiments. The batch crystallization was done using 10-100 μL glass capillaries that were sealed in wax, and additionally

in 0.65 mL and 1.5 mL eppendorf tubes. Batch experiments were performed by mixing the PSI mixture for crystallization experiments 1:1 with a mixture that was made such that the final concentration of the precipitants was identical to the final precipitant conditions in a vapor-diffusion trial, once mixed. When using glass capillaries, the mixtures were initially mixed on a clean piece of plastic wrap and taken up into capillaries through capillary action. The total volume would be between 2 and 50 μL, depending on the size of the capillary. When using eppendorf tubes to house the batch crystallization experiments, the mixtures were mixed directly in the eppendorf tube.

3.4 Isolation and Purification of Photosystem II for crystallization

Materials for cell lysing

Prep1 buffer: (20 m MES pH 6.0, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 10 μM tocopherol)

Prep2 buffer: (20 mM MES pH 6.0, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 500 mM D-mannitol, 10 μM tocopherol)

Prep3 buffer: (20 mM MES pH 6.0, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 500 mM D-mannitol, 20% (v/v) glycerol, 10 μM tocopherol)

PMSF solution: 500 mM PMSF in dimethylsulfoxide (DMSO)

Acetone:

Materials for solubilization and anion-exchange chromatography

Lyophilized n-dodecyl-β-maltopyranoside (β-DDM), Glycon

AMG_0 buffer: (AMG_X: X mM $MgSO_4$,20 mM MES pH 6.0, 10 mM $CaCl_2$, 10 mM $MgCl_2$, 500 mM D-mannitol, 20% (v/v) glycerol, 10 μM tocopherol, 0.02% β-DDM)

AMG_{150} buffer:

Materials for initial crystallization of Photosystem II

D_0 buffer:	(D_X: X% (m/v) PEG 2000, 100 mM (PIPES) pH 7, 5 mM $CaCl_2$)
D_{20} buffer:	(D_X: X% (m/v) PEG 2000, 100 mM (PIPES) pH 7, 5 mM $CaCl_2$)
C buffer:	(100 mM PIPES pH 7, 5 mM $CaCl_2$ 0.03% β-DDM)

Thylakoid preparation

Photosystem II (PSII) was, in principle, isolated as described previously (Zouni et al., 2001), but with modifications. *Thermosynechococcus elongatus* cells were grown, harvested, and stored, as described in Section 3.1. The PREP1 and PREP2 buffers are analogous to MCM and MMCM from Section 3.1, except for the addition of 10 μM tocopherol, used as an antioxidant. PREP3 is used because the added 20% (v/v) glycerol prevents the disassembly of the PSII during solubilization, and all subsequent purification steps use solutions that contain glycerol.

All lighting was turned off during the initial cell lysis, letting dim, natural sunlight, which was filtered through sun-filtering blinds, serve as the only light exposure. A total of 10 mL of 0.5 M PMSF solution was freshly made prior to starting the preparation. A 25 g quantity of frozen *T. elongatus* cells was defrosted and resuspended in a 250 mL centrifuge tube using $<$ 200 mL of PREP1 that had been warmed to 56° C using a water bath. The centrifuge tube was closed, and the centrifuge tube plus sample were repeatedly inverted until a mass of green was not present at the bottom of the centrifuge tube. The cells should not be allowed to reach 4° C for any length of time. The cell suspension was centrifuged at 7000 rpm (7400 *x* g) for 10 min – all centrifugation steps in the thylakoid preparation were carried out using the SLA-1500

fixed-angle rotor (Beckman Coulter) and a Sorvall RC-3C Plus centrifuge (Beckman Coulter) – at 24° C. During the centrifugation run, the microfluidizer was prepared by passing water (2x) followed by PREP2 (3x) through the system. The supernatant was discarded, and a brush was used to lightly remove the orange layer from the top of the pellet. The pellet was resuspended in the smallest quantity of PREP2 possible, typically 10-20 mL, and the solution was diluted to approximately 150 mL using PREP2, with 100 μL of PMSF stock. The cell suspension was passed through a wire mesh before addition to the microfluidizer, to prevent large aggregates of cells from blocking the channel. The cell suspension was passed through the microfluidizer, with a standing pressure of approximately 10.5 kpsi, and a working pressure of approximately 9.5 kpsi. The collected solution was passed through the microfluidizer a second time, using the same pressure settings. The lysed cells were immediately placed on ice, divided into two centrifuge tubes, diluted to 250 mL each, using PREP1 plus 200 μL of PMSF stock, total. The lysate was quickly centrifuged at 14,000 rpm (29,774 *x* g) for 10 min at 4° C. The supernatant was removed, except a small portion which was stored at -20° C for future characterization. Each pellet was resuspended in approximately 25 mL of ice-chilled PREP1 with 40 μL of PMSF stock solution. The solutions were diluted to 250 mL with ice-cold PREP1 and 160 μL of PMSF stock solution. The diluted solutions were centrifuged at 11,000 rpm (18,270 *x* g) for 10 min at 4° C. The pellets were resuspended again as described above and centrifuged at 11,000 rpm (18,270 *x* g) for 10 min at 4° C. The previous step was repeated one additional time. The supernatant was discarded, and the pellets, containing the thylakoids, were resuspended in 10 mL of PREP3, which is ultimately diluted to approximately 40-50 mL using PREP3.

Solubilization and extraction of PSII from thylakoids

A (635 mg acetone/ 199 μL water/ 1 μL sample) chlorophyll assay was performed in triplicate. The chlorophyll concentration was determined using Eq. 3.1. The chloro-

phyll concentration was taken as the mean value of the three measurements, and the standard deviation was used to evaluate whether the chlorophyll assay required repeating.

The sample was diluted to 0.75 mM chlorophyll concentration using PREP3 that contained enough β-DDM such that the final solution had a 1.0% (m/v) β-DDM concentration. *NOTE that the 70Ti rotor can maximally hold 8 centrifuge tubes containing approximately 200 mL total of solution. If the total volume to achieve 0.75 mM chlorophyll is greater than 200 mL, the chlorophyll concentration may be (slightly) increased until the total volume is less than or equal to 200 mL, which causes the solubilization step to be less harsh. The β-DDM conentration is kept at 1% (m/v). However, the solubilization step is predicated on a predetermined ratio of membranes, protein, and detergent, so the concentration of chlorophyll should never be greater than 0.80 mM. NEVER allow the chlorophyll concentration to be lower than 0.75 mM, as having too much detergent for a given amount of the membranes and protein produces much harsher solubilization conditions, which can lead to the disassembly of PSII.* The total solid β-DDM was dissolved in the volume of PREP3 to be added, and the PREP3 plus β-DDM was added slowly to the sample, while gently stirring (to avoid bubbles) in the dark, so as to avoid the denaturing of the protein. *NOTE that the protein solution should never be added to the PREP3 containing the β-DDM because this will create harsh conditions that could denature PSII.* After dilution, the solubilization was allowed to proceed for 45 min at room temperature, but the sample and stir plate were housed in a non-transparent cardboard box to avoid degradation of the PSII by light exposure. Upon completion of the solubilization step, the solution was centrifuged at 50,800 rpm (190,000 *x g*) for 1.5 hr at 4° C using a Ti 70 rotor (Beckman Coulter, Part No. 337922) and an Optima-100K centrifuge (Beckman Coulter, Part No. 393253). The supernatant, containing the solubilized PSII in the form of protein-detergent micelles, was removed and used for the FPLC purification steps.

Preparing the anion-exchange column for the PSII purification

Tentacle anion-exchange chromatography is used to purify the PSII from the protein mixture in the supernatant of the ultracentrifugation step using Toyopearl DEAE-650M resin (Tosoh Bioscience LLC, Part No. 07988). However, the used anion-exchange media is a weak ion-exchange media, as a strong ion-exchange media such as Q-Sepharose causes the disassembly of the PSII complex. The Toyopearl DEAE-650M has beads that attach the positively-charged groups to long, flexible linkers, possibly allowing for a less harsh interaction with the PSII that is bound or eluting through the resin.

The column used is 70 cm in length and 26 mm in inner diameter (XK 26/70, GE Healthcare, Part No. 18-1000-71) and has a total volume of 200 mL. The length of the resin bed inside the column is 60 cm in length. All washing and packing steps must be completed by the day before a PSII preparation to ensure the FPLC and column are ready for use.

The resin material is regenerated and cleaned by washing the entire system with 200 mL of each of the following solutions, in the order presented: 0.01 M NaOH, filtered de-ionized water, 0.01 M HCl, filtered de-ionized water, 2 M NaCl, filtered de-ionized water, filtered de-ionized water, 50% 2-propanol in water, and filtered de-ionized water. The aqueous solutions are passed through the resin at 10 mL/min and the 50% 3-propanol solution is run at 5 mL/min.

PSII purification using anion-exchange chromatography

During the ultracentrifugation step, the cold-water bath is connected to the outer jacket of the column, and the water temperature is set to 4° C. The column is equilibrated by washing the resin with 200 mL of AMG_{150} followed by washing the resin with 200 mL of the 90% AMG_0/ 10% AMG_{150} mixture.

Time (min)	AMG_{150} (% of total)	Flow rate (mL/min)
0.0	10	10
40.0	10	10
50.5	13	10
65.0	13	10
70.0	28	10
100.0	28	10
100.5	100	10
115.0	100	10
115.5	10	10
134.5	10	10

Table 3.2: **Elution gradient used PSII** The step gradient used to separate Photosystem II from cell lysate using a weak anion-exchange media (Toyopearl DEAE-650M).

The program for the PSII FPLC purification steps consisted of loading the sample onto the resin through the same pump used for AMG_0 (pump A). The program used to separate the protein samples has a flow rate of 10 mL/min and is a step gradient, as shown in Table 3.2. Fraction collection is programmed to begin at 75 min and end at 110 min (after the start of the AMG_{150} wash).

A (635 mg acetone/ 197 μL water/ 3 μL sample) chlorophyll assay was performed on the supernatant of the ultracentrifugation sample. An aliquot containing 20 μmole of chlorophyll, from the supernatant of the ultracentrifugation step, was diluted to 100 mL using AMG_0 and loaded onto the Toyopearl column that had been pre-equilibrated with AMG_0. The FPLC run was started where the column was initially washed with 90% AMG_0/10% AMG_{150} and finally washed with 100% AMG_{150} using a step gradient as described above. The second peak contains the PSII; see Fig. 4.4 for a reference chromatogram. The fractions containing the second peak were combined into a graduated cylinder, the total volume of the fractions was recorded, and a 200-μL aliquot was saved for a (635 mg acetone/ 150 μL water/ 50 μL sample) chlorophyll assay and SDS-PAGE characterization. The sample was immediately concentrated at 4° C, as described below. Care was taken to avoid collecting any fractions that would contain significant amounts of other peaks, with peak 1 containing the PSI monomer

as well as the PSII monomer and peak 3 containing the PSI trimer. A total of 4 FPLC runs are typical for a preparation starting with less than 25 g of *T. elongatus* cells.

Concentrating the protein

Fractions containing the second peak from all FPLC purification steps were collected and concentrated using a centrifugal filter unit with a 100 KDa cutoff filter (Millipore, Ultracel – 100K, UFC910024). The samples were concentrated by centrifugation at 4,000 rpm (2870 *x g*) at 4° C using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435). Four centrifugal filtration units are filled with 15 mL aliquots of sample and centrifuged for 11 min. More sample is added to the centrifugal filtration unit to a final volume of 15 mL, and subsequently, the samples are centrifuged for 16 min (increased time is due to more concentrated protein decreasing the flow rate through the membrane). The next centrifugation steps are done at 22 min and 25 min, with all remaining steps being centrifuged maximally for 25 min and the volume being brought to 15 mL as long as sample remains. *NOTE that the sample should never be concentrated to such an extent that the membrane becomes exposed to air, as this will compromise the integrity of the membrane and may cause the membrane to leak or burst.* Upon concentration to a total volume less than 1 mL, the retentate was collected into an eppendorf tube and homogenized. A (635 mg acetone/ 199 μL water/ 1 μL sample) chlorophyll assay was performed in triplicate. The filtrate was examined for color indicating a rupture of the membrane, and if colorless, was discarded. If color was present in the filtrate, the filtrate and retentate (the retentate was used to wash the membrane) were added to a new centrifugal filter unit, and the concentration step was restarted.

Final PSII purification through initial crystallization

At this step, the PSII is precipitated using PEG2000. The volume of the PSII sample was determined by adding the protein to an eppendorf tube of known mass and calculating the mass difference. The density of the protein solution was assumed to be 1 g/mL. The PSII sample was added to a 50-mL vessel and buffer C was added to reach a chlorophyll concentration of 0.75 M. While stirring, a volume of buffer $D_{15.0}$, equal to the volume of buffer C used for dilution, was added dropwise. The sample was covered and placed on ice for a 40-min incubation. The sample was centrifuged at 5100 rpm (5525 *x g*) for 4 min at 4° C using a swinging bucket rotor (TS-5.101500, Beckman Coulter, Part No. 368308) and an Allegra 25R centrifuge (Beckman Coulter, Part No. 369435). The supernatant was discarded, and the precipitate was dissolved in a volume of buffer C equal to the initial amount of buffer C added. While stirring, a volume of buffer $D_{13.0}$, equal to the volume of buffer C used for dilution, was added dropwise. The sample was covered and placed on ice for a 40-min incubation. The sample was again centrifuged at 5100 rpm (5525 *x g*) for 4 min at 4° C using the previous setup. The second supernatant was discarded, and the second precipitate was dissolved in a volume of buffer C equal to the initial amount of buffer C added. While stirring, a volume of buffer $D_{11.0}$, equal to the volume of buffer C used for dilution, was added dropwise. Upon addition of the $D_{11.0}$, the sample was covered and stored at 4° C until the sample was needed for experiments. The sample was tested for crystallinity using SONICC, as discussed in Section 3.5, and UV microscopy.

3.5 Experiments to characterize the sample and determine the nature of the precipitates

The techniques used to characterize the sample and determine the nature of the precipitate samples will be explained in this section. The results of using the techniques listed

in this section will be reported for the various samples that were used in the results section associated with the sample.

SDS-polyacrylamide gel electrophoresis characterization

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) characterization was done to ascertain the purity of the proteins isolated and purified during the protein purifications. SDS-PAGE was done utilizing both a manual method and an automated method. The automated method was done using high-density gels and the PHAST system (Phastsystem, GE Healthcare, Part. No. 18-1018-23), while the manual method made use of a 20% polyacrilamide gel, as described below.

Tricine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) techniques described by (Schagger, 2006) were used for the analysis of the protein samples in the manual method. Separation of the protein subunits could be obtained with a 20% polyacrylamide separating gel layered with a 1-cm 12% polyacrylamide stacking gel. A 30% solution of acrylamide and bis-acrylamide solution, 29:1, (Bio-Rad, Prod. No. 161-0156) was used to make the gels. Gels were usually run at a constant voltage of 80 V during migration in the 12% stacking portion, followed by an increase to 150 V during migration in the separating gel portion (Initially, the protein subunits were separated using 290 V, but under these conditions, too much heat was generated, leading to porrer separation and evaporation of the running buffers). The Bio-Rad Mini-PROTEAN Tetra Cell electrophoresis system (Bio-Rad, Prod. No. 165-8000) was used for preparing and running the gels. Precision Plus Standards (Bio-Rad, Prod. No. 181-0374) were used as ladders for all gels. Samples were prepared for loading on the gel by mixing with 4x 73 SDS Sample Loading Buffer (26.7% sodium dodecyl sulfate, 133 mM Tris-HCl pH 8.0, 2.5 M 2-mercaptoethanol, 26.7% glycerol, 133 μM bromophenol blue). The 4x SDS sample loading buffer was added to samples in a volume that corresponds to 1x final concentration of the total volume (25% v/v). Denaturation and

solubilization of the samples was achieved by heating samples at 92°C for 10 minutes prior to loading.

The gels were initially stained using a Coomassie blue staining protocol. Coomassie Blue Staining was done by equilibrating the polyacrylamide gel in heated Coomassie Blue Stain (0.1% Coomassie Brilliant Blue (CBB) 250, 30% methanol, 10% acetic acid) for 5 minutes, followed by equilibration in heated Destain I (50% methanol, 10% acetic acid) for 5 minutes, and final equilibration in Destain II (5% methanol, 10% acetic acid) until bands were well resolved. For the heating steps, the gels and solutions were heated together in a microwave (0.95 kW) for 30 seconds. Alternatively, the gels can be allowed to incubate in room temperature Coomassie blue stain for a 2-3 hours, but the additional time made this protocol impractical.

Silver-staining was ultimately chosen as the staining technique due to a higher sensitivity and a lower limit of detection than Coomassie blue staining. Silver Staining was done by equilibrating the polyacrylamide gel in 12.5% glutaraldehyde solution for one hour, followed by a 5 minute rinse in H2O. The rinsing step was done by washing the gel 2-3 times with water and pouring the water off. The gel was incubated in the final wash of water for 5 min. The gel was subsequently equilibrated in a 1.0% $AgNO_3$ solution for one hour, which was also followed with a 5 minute rinse in H_2O in a manner identical to the one described above. The gel was then transferred to Developer Solution (0.25% 74 formaldehyde, 6.25% (m/v) Na_2CO_3) and equilibrated until bands could be distinctly resolved, at which point the gel was immediately transferred to Fixing Solution (10.0% (m/v) glycerol, 10% (v/v) acetic acid) to preserve development. All silver staining steps are carried out at room temperature.

Although silver staining has a much lower limit of detection than Coomassie blue staining, there was one issue with silver staining. The ferredoxin band was never resolved when using a silver staining procedure, and consequently, in order to observe whether ferredoxin is present in a sample, Coomassie blue staining must be done. The

recommendation is to do two gels in parallel, one gel can be stained with Coomassie blue, while the second gel is silver stained. One staining method that has been shown to have as low of a limit of detection as silver staining, but requires the same amount of time as Coomassie blue staining, is zinc staining (Cong et al., 2010) and may present an alternative approach to observing the ferredoxin band while retaining the low limit of detection for the low molecular mass subunits and contaminants.

Birefringence

Birefringence is a first test of the crystallinity of a sample but can only be used for crystals greater than 1 μm in size. Birefringence is measured by re-suspending a 10-μL volume of the precipitated sample and placing 0.5-1 μL of the re-suspended sample on a microscope slide. The slide is placed on a light microscope between two polarizers with orthogonally aligned polarization directions.

Images can be recorded on the stereomicroscope using the SPOT Advanced Imaging program (SPOT imaging solutions, Prod. No: SPOTA). The exposure times that were found to be the best were between 20 and 30s. For the highest quality birefringence images, the lights in the room should be dimmed as well. The brightfield image was also taken for the sample.

SONICC

All precipitate samples were tested for second-harmonic generation (SHG) using the SONICC technique in collaboration with the group of Dr. Garth Simpson, who has developed this technique (Wampler et al., 2007). After an initial visit to the lab of Dr. Simpson to learn the technique, further samples were sent to the group of Dr. Garth Simpson at Purdue University. A schematic of the SONICC setup used is shown in Fig. 3.5.

The samples were monitored for second-harmonic generation in the forward

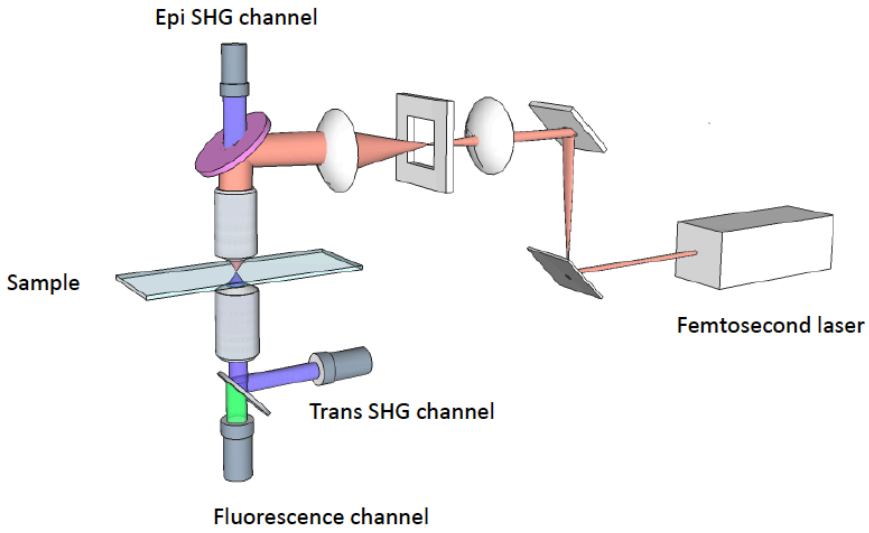

Figure 3.5: **Schematic of the SONICC instrument.** A schematic of the SONICC instrument showing the trans and epi second-harmonic generation (SHG) channels and fluorescent channel, as well as the sample position within the instrument. Image provided by Ellen Gualtieri and Garth Simpson of Purdue University.

(trans) and backward (epi) directions. Narrow-band filters were used to select the wavelength of the second harmonic for the SHG channels. The smallest crystalline samples would tend to have a larger epi signal relative to the trans signal than the $>$ 500 nm samples. The sample was also monitored for broadband fluorescent signal, by measuring for fluorescent signal in a wavelength range surrounding the wavelength of the second-harmonic. The fluorescent signal was used to discriminate false positive SHG measurements from large fluorescent signals; a large fluorescent signal relative to the total SHG signal would be indicative of fluorescent signals being recorded as SHG. However, the PSI, PSII, and PSI-Fd samples were never shown to have significant fluorescent signals.

The samples were sent for SONICC measurements by overnight mail. The samples were stored in eppendorf containers sent in Styrofoam boxes with gel packs that had been equilibrated to the correct temperature. For example, PSI-Fd co-crystals were to be sent at room temperature, and therefore, gel packs were equilibrated to room temperature and used as a temperature buffer for the samples. Samples were loaded

into 0.65-mL eppendorfs in 10-μL volumes (for vapor-diffsuion crystallization trials) or 0.3-0.5-mL volumes (for crystallization screens that produce larger volumes, such as batch), and plastic wrap was used to tightly seal the lid. Once the sample-containing eppendorf had been placed into the eppendorf holder, the eppendorf was taped into place.

Samples for SONICC were initially tested using a chlorophyll concentration of the re-suspended precipitate samples of approximately 100 μM. However, the procedure was changed to only re-suspending the top portion of the pellet of the sample, and imaging this portion with SONICC, because the smallest crystals would be found on the top layer of any pellet. The sample was added to the sample well of a sitting-drop plate and covered. The reason that the first trial was used with a sample that had not been quantified was that the nature of the SONICC experiments only requires a thorough knowledge of the sample concentration when no or little SHG signal is observed. If the sample was positive for SHG and negative for fluorescence, it was very likely crystalline. The amount of SHG was compared to a bright-field image of the sample to compare the number of particles to SHG positive regions.

The above method could be prone to false negatives because amorphous precipitates are less dense than crystals. A better method of testing the sample would be to use the completely re-suspended sample to avoid only sampling the amorphous precipitate. However, completely re-suspending the sample would make identification of the smallest crystals difficult, as the measured signal would overwhelmingly consist of SHG from the larger crystals. Consequently, the preferred method would be to initially only scan the top portion of a pellet for crystallinity so as to measure the smallest crystalline samples, and if the sample does not show SHG signal, the sample may be thoroughly re-suspended to determine whether only a portion of the material was amorphous.

The settings on the instrument that can be altered are the magnification of the

objective lens, the power of the laser, the wavelength of the laser, and the number of scans signal-averaged before being read out. The 10x objective was exclusively used for all experiments. The power of the laser was adjusted between 50 and 150 mW, with the most common setting being 50 mW. The wavelength used was 800 or 1000 nm, with 1000 nm being chosen for the protein-pigment containing samples. The sample was measured through rastering across the sample, and in order to improve the signal to noise, typically 2000 scans were performed before readout.

3.6 Serial powder diffraction experiments using PSI crystallites

Brief overview of serial PD experiment and facility used

Experimental details in this section are taken from (Hunter et al., 2011) with additional details provided. X-ray diffraction data were collected at beam line 9.0.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory, using photon energies of 520 eV and 1560 eV.

Details of the PSI samples and the on-site preparation

The PSI samples used in the serial powder diffraction experiments were initially grown by ultrafiltration using a final crystallization buffer of G_6. However, for the results presented in this section, the PSI crystallite samples were grown using a final buffer of G_8. During the initial serial powder diffraction experiments, the re-suspended PSI crystallite samples were inline filtered without using settling experiments to remove the largest PSI crystals. The settling experiments—Section 3.1—were developed during the time of these experiments as a method to reduce the loss of large crystals in the PSI sample to the inline filters and to coarsely select the crystallite size. The PSI samples used for the serial powder diffraction experiments consisted of the supernatant of the 60 min settling steps of re-suspended PSI crystallite samples from six different PSI preparations. The serial powder diffraction experiments were carried out over the course of three years.

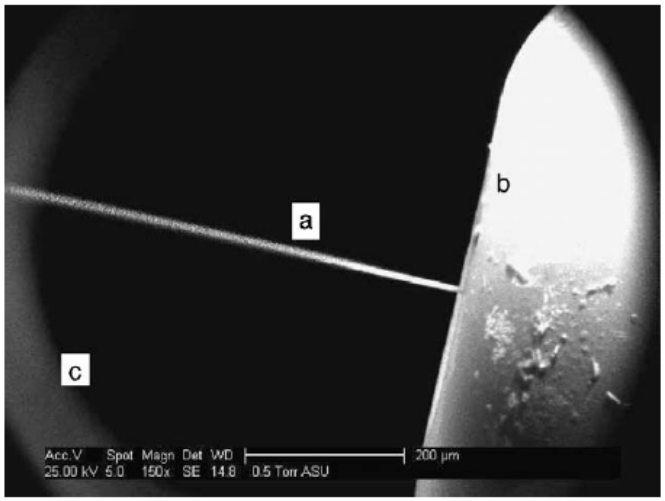

Figure 3.6: **ESEM image of a liquid jet.** Environmental scanning electron microscope image of a liquid jet (a) emerging from the nozzle (b) as viewed through the 500-μm pressure-limiting aperture. Water jets of diameter up to 700 μm have been imaged in the fashion. Figure and caption taken from DePonte et al. (2008).

PSI crystallite samples were stored at 4° C prior to use in the serial powder diffraction experiments. Samples of the PSI nanocrystals were re-suspended and were pre-filtered immediately before use. The samples were pre-filtered by first being loaded into a 1-mL reservoir with a 0.5-μm inline filter at the outlet port. Gas pressure was used to flow the sample into the primary reservoir that was associated with the injector, which likewise had an inline filter. A 25-μL aliquot of the sample was taken and a (635 mg/195 μL/5 μL) chlorophyll assay was performed in triplicate to determine the concentration. The sample concentration was between 0.5 mM chlorophyll and 1 mM chlorophyll for the experiments. However, the exact concentration is irrelevant in the case of these experiments, as increasing the number of patterns collected will compensate for lower PSI-crystallite concentration.

Sample delivery and overall experiment design

The sample reservoirs were made from acrylic cylinders with 2-mm bores. The reservoirs had to be kept vertical, due to the use of gas pressure to drive the liquid flow. The liquid line of the nozzle was attached to the bottom of the reservoir and a gas line, em-

anating from a gas cylinder, was attached to the top of the reservoir. The reservoir was attached to the end of the injector so that fiber optic lines would be sufficiently long to reach the chamber. The liquid flow was driven by gas pressures typically between 300 and 1000 psi. The gas pressure for the coaxial gas was typically between 100-300 psi, but as is the case for the liquid pressure, the necessary pressure varied wildly between the different nozzles. Additionally, slow clogging of the inline filter often required the gas pressure to be increased in order to maintain a constant flow rate. Flow rates were monitored by estimating the liquid consumption using graduations machined into the acrylic reservoirs.

The PSI samples were introduced to the interaction region using a 10 μm diameter liquid jet produced by the gas-dynamic virtual nozzle, shown in Fig. 3.6, as detailed previously (DePonte et al., 2008) utilizing the serial crystallography technique (Spence and Doak, 2004). The diffraction patterns were obtained from the continuous area of the jet, before Rayleigh breakup occurs, using the scheme shown in Fig. 3.7. This reduces the background in the diffraction patterns produced by the shape transform of the jet to a streak perpendicular to the jet, as shown in Fig. 3.7. The X-ray illuminated volume contained many crystals in different orientations at any time, producing the observed powder diffraction patterns shown in Section 4.6. Diffraction patterns were collected using an on-axis CCD detector or an off-axis CCD detector, which doubled the achievable resolution from 2.8 nm to 1.4 nm. The 9^{th} harmonic of a 10 cm period undulator was focused into a 50-μm beam spot by an off-axis zone plate segment, providing a total incident photon flux of 10^{12} photons/second with a bandwidth of $\sim$1% for the 1560 eV X-rays.

The experiments were performed *in vacuo* with a base pressure of 10^{-6} Torr and a working pressure (when using the jet) between 10^{-4} and 10^{-3} Torr ($1.3\ x\ 10^{-7}$ to $1.3\ x\ 10^{-6}$ atm). The chamber pressure was closely monitored during the experiments for several reasons: Because soft X-rays will interact with matter more than hard X-

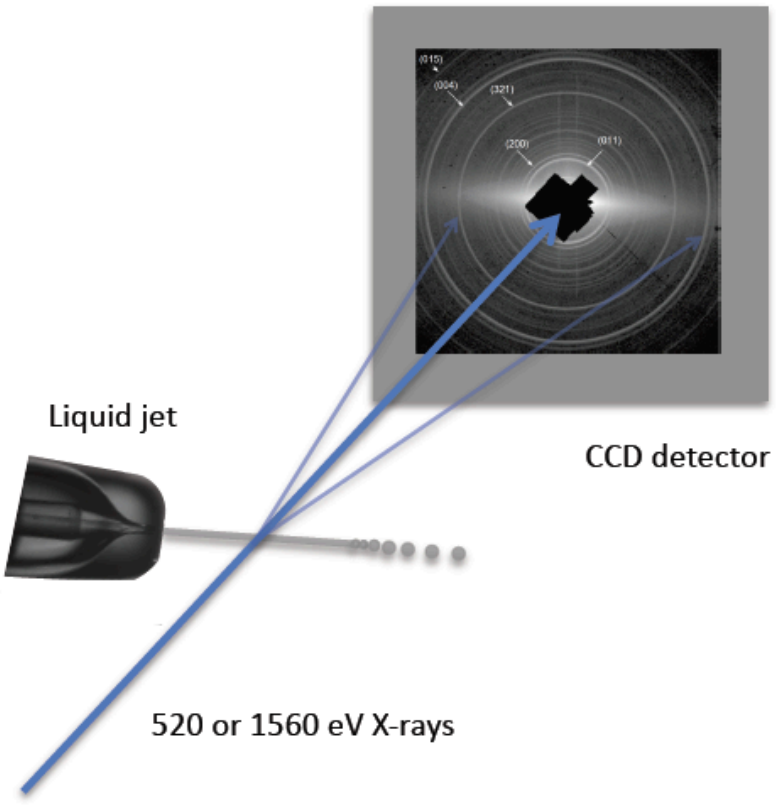

Figure 3.7: **Serial powder diffraction.** Overview of the serial powder crystallography experiment. An aqueous suspension containing many PSI crystallites passes through an incident X-ray beam, creating a powder diffraction pattern. Figure and caption taken from Hunter et al. (2011).

rays, in order to protect the sensitive detector from having vapor condense onto the chip, and to avoid tripping vacuum systems upstream of the endstation. To reduce the influence of the coaxial gas and liquid on the chamber pressure, CO_2 was used as the accelerating gas, and a cold trap was used to condense the CO_2. Although vacuum pumping is faster using helium or nitrogen, due to the diffusion constants, freezing the CO_2 was more effective at keeping the vacuum load minimal, and the cold trap froze the water from the liquid jet. The cold trap was made using a copper pipe that was fed with liquid nitrogen. Inside the chamber, multiple concentric rings of the copper pipe were made to increase the surface area of the cold trap. The chamber was vented during off-shifts by turning off the liquid nitrogen flow and only pumping on the chamber with a roughing pump (typically defined as the vacuum pump used to initially reduce the

chamber pressure).

The injector was mounted in a housing containing a gate valve to allow for the quick replacement of a nozzle while maintaining the main chamber vacuum. The injector would be withdrawn behind the gate valve, with the gate valve subsequently closed in order to replace a nozzle.

The injector was mounted on a *xyz*-stage to allow for precise control of the position of the injector relative to the X-rays. In order to find the jet, the injector was moved completely in toward the X-ray beam and translated vertically until the signal from the X-rays was blocked. The injector was slowly backed out until the signal returned. The injector would be moved vertically until scattering from the nozzle was detected, after which the injector position was optimized by maximizing the streak from the liquid jet on the detector.

Detector settings and resolution

The detector used in the experiment was a CCD detector (Princeton Instruments, Part No. MTE2-1300B) and was set at a working distance of 9 cm (Shapiro et al., 2008). The detector has a 1340 *x* 1300 array of 20-μm pixels and the largest scattering angle recorded was 12° with the detector on-axis. The detector was moved diagonally away from the central axis for some experiments, allowing for the maximum recordable scattering angle to be 25°. Using a 9-cm working distance, the detector provided an angular resolution of 0.01° when on-axis (0.02° diagonally), which corresponds to approximately 100 pixels per peak FWHM with 1560 eV X-rays, and consequently 4 *x* 4 binning of the data was used.

Using the on-axis detector, powder diffraction rings to a resolution of 2.8 nm could be recorded with a total exposure of 200 s using a total sample solution of 74 μL in volume, corresponding to 143 μg of PSI for the crystal suspension filtered through a 500-nm filter. To avoid pixel saturation and improve the counting statistics at higher

scattering angles with the detector moved off-axis, repeated 30-second exposures were taken and averaged during post-processing. The maximum exposure time for a single CCD exposure is determined both by the strength of the background signal and by the strength of the scattered signal at small scattering angles. A high background signal and a large scattered signal at low spatial frequencies will cause detector saturation and signal bleeding—which is when the accumulated charge spills into neighboring wells of the CCD detector—for exposures longer than 30 seconds. The statistics of the signal at large scatterings angles are improved by averaging many such CCD recordings; this continuous signal is readily subtracted from the semi-discrete powder rings.

3.7 Femtosecond nanocrystallography at LCLS

Calculating the crystallite density

For the femtosecond nanocrystallography experiments, the ideal condition allows one LCLS pulse to hit at most one crystal, with the diffraction pattern being read out before the subsequent pulse arrives. In order to assure that the number of multiple hits, defined as diffraction patterns containing contributions from more than one crystallite, was low, the average hit rate was set so that one out of every four interaction volumes would contain a crystallite. The interaction region volume needed to be calculated so that the crystallite density could be appropriately set. The interaction volume is the intersection of a Gaussian beam profile and cylindrical jet, so the volume was approximated, as described below. In the case of the jet being bigger than the X-ray focus (AMO experiments 2010, CXI experiments 2011), the following formula was used:

$$V_1 = \pi(r_\perp r_j)d_{jet} \tag{3.7}$$

where V_I is the volume of the interaction region, r is the beam radius in the orthogonal

direction to the flow of the jet and the z-axis, and r_j is the beam radius in the direction parallel to the flow of the jet. In the case of the X-ray focus being larger (AMO experiments 2009), the following formula was used:

$$V_{\mathrm{I}} = \pi \cdot r_j (\frac{d_{jet}}{2})^2 \tag{3.8}$$

All the lengths in the calculations are measured in consistent units, but nanometer is usually the most convenient length unit. The formulae given above will overestimate the interaction region volume. The exact volume could be determined through a volume integral of the interaction region, but due to the uncertainties generated by the various assumptions (see below) of the crystal density calculation, the gain of accuracy would be minimal.

Once the interaction volume was estimated, the crystal density would be set such that there would ideally be one crystallite for every four interaction-volumes:

$$\rho_{\mathrm{cryst}} = \frac{1}{4V_{\mathrm{I}}} \tag{3.9}$$

where ρ_{cryst} is the crystal density. If all lengths are given in nm, then the calculated volumes will be in femtoliters (fL).

In order to calculate the number of crystallites in a given volume, a few methods could be utilized. For a suspension that contains only crystals visible under a light microscope, the sample could be re-suspended, and a defined volume could be placed onto a microscope slide and imaged, giving a value to directly determine ρ_{cryst}. However, this method is not feasible for most protein crystals, even in samples containing a few large crystals, as without the use of a cover slide, the depth of field restricts the cer-

tainty of the calculation, and cover slides could not be used in the case of PSI because the crystals would be crushed. The counting method is also excluded for nanocrystals samples, as these samples mainly contain crystals that cannot be visualized with a light microscope. EM would be another option, but PSI crystals cannot be imaged by EM as they dissolve upon the evaporation of water that is required for the thin ice films used in cryo-EM, and negative staining would dissolve the PSI crystals. Thereby, the crystallite density was coarsely approximated using the chlorophyll concentration of the sample and assuming an average size of the crystals. A (635 mg/ 195 μL/ 5 μL) chlorophyll assay was performed in triplicate on the PSI crystallite suspension, and the average was used as the chlorophyll concentration of the solution.

In the case of PSI, the exact size distribution of the samples was unknown, and a size assumption was made by one of the following two ways. In the case of PSI crystals that were larger than 1 μm in size, the mean of the size distribution was estimated using a light microscope. For the small PSI crystallite samples, the mean of the size distribution was set such that it coincided with the middle value within the range of sizes that pass through the filter, *e.g.*, for 0.1—2-μm PSI crystallites, the size was assumed to be 1 μm. The data collection and observed hit rate at the LCLS was proof that using the midpoint in the crystal-size range is a good estimate in setting the crystallite density. However, it should be noted again that for samples that are not as sensitive as PSI, much more exact methods, such as EM, should be feasible.

To understand the formula used for a hexagonal crystal, the derivation will be presented. The area of a parallelogram with edges of length a and b is ab, with the corresponding hexagon of area $3ab$. A hexagonal needle such as PSI with a height of c would have a volume of $3abc$. Therefore, the volume of the crystal of given size was calculated using the following formula for a PSI crystallite (a hexagonal system) assuming a hexagonal crystallite:

$$V_{\text{crystal}} = 3l_c(\frac{d_{\text{ab}}}{2})^2 = 3\frac{d_{\text{ab}}^3}{4}R_\Lambda \tag{3.10}$$

where l_c is the length of the crystal in the direction of the c-axis, d_{ab} is the diameter of the crystal in the direction parallel to the ab-plane, and R_A is the aspect ratio (defined in this case to be the ratio of the length of the crystal in the c-direction to the diameter of the crystal) of the crystal. The term d_{ab} corresponds to the diameter of the crystal, measured in the ab-plane of the crystal. The largest value that d_{ab} can assume is the filter pore size, but l_c can be larger than the pore size by a factor of the aspect ratio R_A.

The total number of unit cells in the crystal is $V_{crystal}$ /$V_{unitcell}$. The total number of chlorophyll molecules in the crystal is equal to the number of unit cells multiplied by (2 trimer/unit cell)·(288 chlorophyll molecules/trimer) = 576 chlorophyll molecules per unit cell (Jordan et al., 2001). Multiplication of the total number of chlorophyll molecules in the crystal by ρ_{cryst} generates the total number of chlorophyll molecules in the mixture of crystallite density ρ_{cryst}. The next step is to divide the number of chlorophyll molecules by Avogadro's number to arrive at the total number of moles of chlorophyll in the sample. Dividing by the volume generates the chlorophyll concentration necessary in order to have a crystallites density of ρ_{cryst} with all crystallites of the assumed size.

In the case of PSI-Fd, the size distribution was also unknown. The formula for the crystallite volume of PSI-Fd was different than in the case of PSI because the PSI-Fd space group is orthorhombic ($P2_12_12_1$, a = 194 Å, b = 208 Å, c = 354 Å) (Fromme et al., 2002). Consequently, the crystal volume was calculated as follows for the orthorhombic case:

$$V_{\text{crystal}} = s_a s_b s_c \tag{3.11}$$

where s_a is the length of the side of the crystal in the **a** direction. The length of a side of a PSI-Fd crystal that passes through a filter of pore diameter d_{filter} is $d_{filter}/2^{1/2}$.

The total number of unit cells in the crystal is simply $V_{crystal}$ /$V_{unitcell}$. The total number of chlorophyll molecules in the crystal is equal to the number of unit cells multiplied by (4 trimers/unit cell)·(288 chlorophyll molecules/trimer) = 1152 chlorophyll molecules per unit cell (Fromme et al., 2002). The remainder of the calculation is identical to the case for PSI.

To calculate the PSII crystallite density, the PSI-Fd crystallite density calculation is used as a template, since PSII also grows as with orthorhombic space group ($P2_12_12_1$, a = 127.69 Å, b = 225.40 Å, c = 306.11 Å). The only difference is the calculation of the total number of chlorophyll per unit cell. The total number of chlorophyll molecules in the crystal is equal to the number of unit cells multiplied by (8 dimers/unit cell)·(70 chlorophyll molecules/dimer) = 560 chlorophyll molecules per unit cell (Loll et al., 2005). Again, the remainder of the calculation is analogous to the case for PSI.

The sample volume is adjusted to meet the chlorophyll concentration of the calculation. The chlorophyll concentration of the sample is calculated after the prefiltration step, and the sample must be thoroughly homogenized.

The procedure described for setting the crystallite density is far from ideal, with a large number of approximations. A direct method to determine the crystallite density is needed, but empirically, the method described above reproducibly allows for a hit rate of 5-20% at femtosecond nanocrystallography experiments. The method assumes that the number of crystals decreases away from the chosen mean in a similar manner on both sides of the mean. However, the presence of a large amount of small crystals would yield the above calculation less able to predict the crystal density. The method was developed with the assumption that the crystallite size distribution is Gaussian, and

although the shape of the distribution is unknown, the calculation makes a reasonable initial approximation. The calculation described above will work better with smaller size distributions.

Experiments at AMO in beamline in December 2009

The experimental details in this section are published in (Chapman et al., 2011) with additional details provided.

The femtosecond nanocrystallography experiments (Chapman et al., 2011) were performed at LCLS at SLAC at the AMO (Atomic, Molecular and Optical Science) beamline (Bozek, 2009) *in vacuo* (base pressure: 2×10^{-7} Torr, pressure jet injection conditions: 10^{-5} Torr) in the CFEL-ASG Multi-Purpose (CAMP) endstation (Struder et al., 2010), placed 3 m downstream of the nominal AMO instrument. The experiments were performed with the setup shown in Fig. 3.8.

X-ray pulses, generated at a repetition rate of 30 Hz, were focused to a 7-μm diameter spot with a pulse intensity of 900 J/cm^2, corresponding to a dose of 700 MGy per pulse to PSI and a peak power density in excess of 10^{16} W/cm^2 at 70-fs duration. The pnCCD detectors can operate at a maximum frame rate of 200 Hz and here were read out at the 30-Hz rate of the delivered LCLS pulses. Every X-ray diffraction pattern was digitized and stored. Each detector panel consists of 512×1024 pixels of 75×75 μm^2. The back detectors, located 564 mm from the jet, record low-angle scattering from 0.1° to 4.0° in the vertical scattering plane, and the front detectors, located 68 mm from the jet, cover 4.6° to 40.5° in the same vertical plane. The largest scattering angle magnitude accepted by the front detector was 45.7°, corresponding to a resolution d of 8.9 Å at the edge of the detector using 6.9-Å wavelength. X-ray fluorescence from the water jet was filtered by a 8-μm thick polyimid film in front of the pnCCDs.

A liquid microjet (Hunter et al., 2011; DePonte et al., 2008) was used to inject the nanocrystal suspension into the FEL beam. The microjet was emitted from

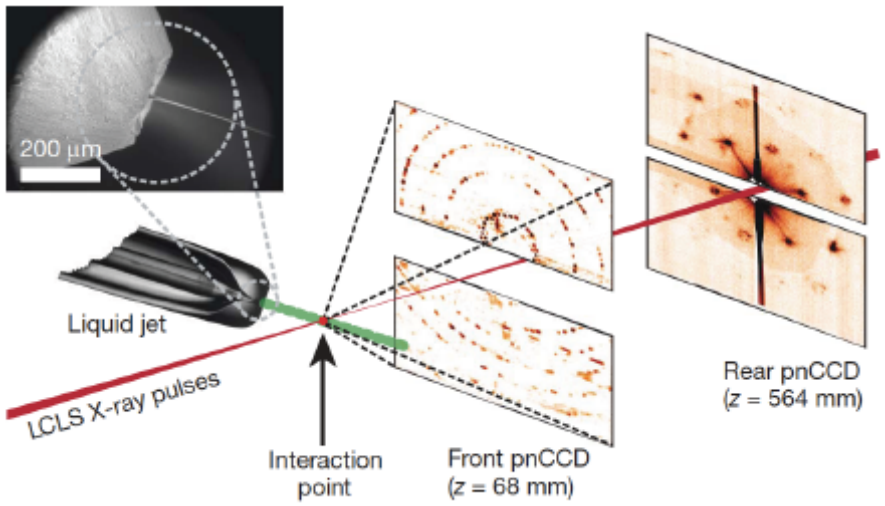

Figure 3.8: **Femtosecond nanocrystallography.** Photosystem I nanocrystals flow in their buffer G_0 in a gas-focused 4-μm diameter jet, perpendicular to the X-ray beam that is focused on the jet. The insert shows an environmental scanning electron micrograph of the nozzle, flowing jet, and focusing gas (DePonte et al., 2008). Two pairs of high-frame rate pnCCD detectors (Struder et al., 2010) record low and high-angle diffraction from single X-ray FEL pulses, at a rate of 30 Hz. Crystals arrive at random times and orientations in the beam, and the probability of hitting one is proportional to the crystal concentration. Figure and caption from (Chapman et al., 2011) with slight modifications.

a 40-μm-diameter capillary, and focused down to a 4-μm diameter column by a gas focusing, achieved by a coaxial flow of helium. The gas focusing technique avoids clogging of crystallites, as the capillary exit is much larger than the nanocrystals. The narrow jet diameter confines the crystallites to the most intense part of the focused X-ray beam and reduces X-ray attenuation by the water to at most 30% (the attenuation length of water is 12 μm at 1.8 keV photon energy). The interaction region of the X-rays and crystallites is located in the continuous liquid column, prior to the Rayleigh breakup of the jet into drops, so that most of the X-ray scattering from the liquid is confined to a narrow vertical streak in reciprocal space.

Photosystem I crystallites from two different protein batches were used in the collection of the data at AMO in Dec. 2009. The PSI nanocrystals were grown in batch at low protein concentration (30 μM P700 = 10 μM Photosystem I trimer) and low

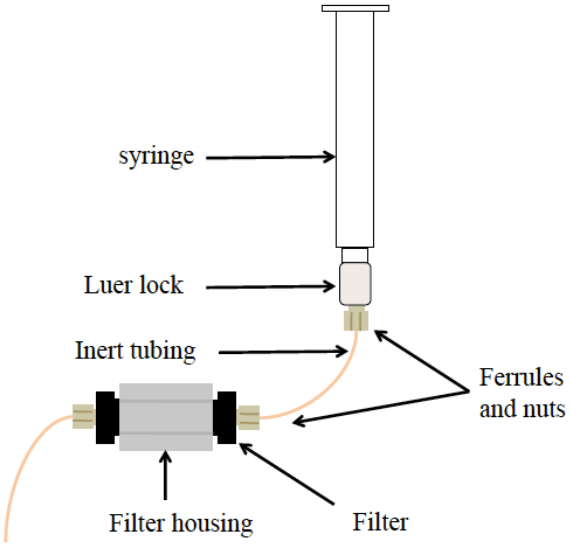

Figure 3.9: **Nanocrystal pre-filtration setup.** Schematic showing the setup and pieces used to pre-filter the samples before use at the LCLS. The sample would be pushed through using the smallest amount of force necessary to provide flow.

ionic strength (G_{10} buffer) at 4° C, as described in Section 3.1.

The size of the PSI crystallites was restricted by pre-filtering using a 2-μm inline filter (Inline filter, Upchurch). The sample was loaded into a syringe and passed through an inline filter of desired size, as shown in the schematic of Fig. 3.9.

The filtered crystallites were stored at 4° C in G_0 buffer. The PSI nanocrystals were suspended in G_0 and a (635 mg/ 195 μL/ 5 μL) chlorophyll assay was performed to determine the protein concentration. The volume of the sample was adjusted to set the protein concentration to 1 mg/mL. In order to ensure that a constant flow was running through the nozzle at all times, a dual-line system was designed, as shown in Fig 3.9, that had manual valves that could be switched from a sample containing line to a buffer (or water) containing line. The nanocrystal suspension was filled into the sample loop of the liquid jet injector, while the solvent loop was filled with G_0 buffer.

The constant liquid flow was necessary to prevent freezing of the nozzle caused by termination of flow. The dual-line system also allowed fast switching to the buffer

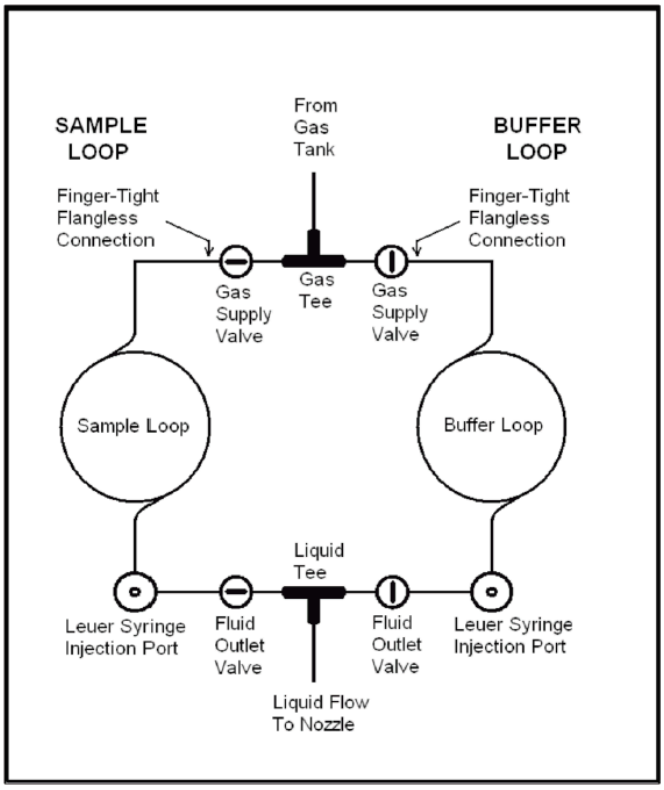

Figure 3.10: **Dual-line setup utilized in the Dec 2009 femtosecond nanocrystallography experiments.** The scheme maintains constant flow through the nozzle when changing samples to prevent freezing of the nozzle that occurs if the liquid stops flowing. The diluted Photosystem I nanocrystals (at 1 mg/mL) are contained in a "sample loop" of up to 10-mL volume. A "solvent loop" was integrated so that the sample need not be consumed during X-ray alignment adjustments and to provide constant flow through the jet when changing the sample suspension. Figure and caption taken from (Chapman et al., 2011).

solution during instrumental adjustments of the X-ray laser. Unfortunately, there was not a differential pumping section in the first version of the injector; therefore, there was no way to remove a frozen nozzle without bringing the entire chamber to atmospheric pressure, which would require bringing the detector up to room temperature. Replacing the nozzle would require a loss of approximately 6 hours of beamtime. Due to careful sample preparation, no clogging of the nozzle occurred during PSI data collection. However, as described in Section 3.7, alterations to the injector design were made to allow for the nozzle to be changed, as necessary, without involving the entire sample chamber.

The constant liquid jet of PSI nanocrystals (crystal size 0.2 μm – 2 μm, upper size restricted by filtering) was adjusted by the gas flow to a diameter of 4 μm, which is slightly smaller than the 7-μm focus of the X-ray beam. The PSI crystallite concentration was chosen such that the probability of hitting a crystal was about 25%, in order to reduce double hits; the nanocrystal suspension contained, on average, one crystal per 400 μm^3, which is four times the intersection of the jet and X-ray beam volumes, corresponding to 2.5×10^9 crystals/ml.

Data collection at the AMO beamtime in Dec. 2009

The experimental details in this section are provided in (Chapman et al., 2011) with slight modifications or additional details included.

Some high intensity peaks are broadened due to saturation of the detectors, as can be seen for some peaks in Figs. 4.13 and 4.17. The detector pixels have a full well capacity of 282,000 electrons, corresponding to 571 X-ray photons of 1.8 keV (λ = 6.9 Å). When more than 571 photons arrive in a pixel on the front detectors, the generated electrons "spill" into neighboring pixels. The integrated photon count in the peak is correct, but the shape of the peak is lost in that region. The charge collected in the CCD wells is transferred to read-out registers and digitized into a 14-bit signal. The rear detectors were operated with a gain setting that produced 10.5 digital counts per X-ray photon, in order to match the well depth to the digitization range. In order to optimize the detection of weaker peaks on the front detectors, they were operated with a higher gain setting that produced 167 digital counts per X-ray photon, which saturated the analogue to digital converter (ADC) at 78 X-ray photons, and gave a readout noise less than a photon.

X-ray fluorescence from the water jet is significant and is a potential source of background of the diffraction patterns. Due to the small dynamic range of the detector after switching to a higher gain setting, the fluorescence signal needed to be minimized.

For a 4-μm diameter jet, 210 oxygen K-shell fluorescence photons (525 eV) will be expected per pixel per pulse on the front detectors, whereas the smaller, solid angle of the rear detectors will collect 2.9 fluorescence photons per pixel per shot. The fluorescence photons are attenuated by an 8-μm thick polyimid foil placed between the water jet and the front pnCCDs. This foil has a transmission of 10^{-5} at 525 eV, whereas the 1.8 keV diffraction from the crystals is only attenuated by 0.5. A 4-mm diameter hole was punched in the polyimid filter for the direct FEL pulse to pass through. A shadow of the hole caused by the fluorescence photons can be faintly seen on the rear detectors, as seen in the patterns in Fig. 4.32. This shadow is magnified by a factor of 14, as given by the source-to-foil and source-to-detector distances. The fluorescence signal level in the unattenuated region is about 5 digital counts, in agreement with the abovementioned expectation.

Typical noise levels due to readout and diffuse photon background were lower than 1.3 photons/pixel. All patterns have been processed by removal of signal from known bad pixels, applying a pre-characterized gain (flat-field) correction, and subtracting a background. The latter step is carried out by initially identifying frames with no crystal hits by simple threshold discrimination. A background estimate is determined from a moving-window selection of 50 nearby no-hit frames to ensure similar jet and exposure conditions. For each pixel, the background is taken as the median of values at that location over the 50 frames. After background subtraction, the frames are again searched for diffraction spots by a threshold analysis and aperture photometry. For the quantitative analyses of indexing 4.29, 4.32 and averaging intensities 4.31, the saturated pixels were identified, and the charge that leaked into neighboring pixels was then integrated. This step essentially increases the dynamic range of the detector, but loses the shape information of the affected reflections.

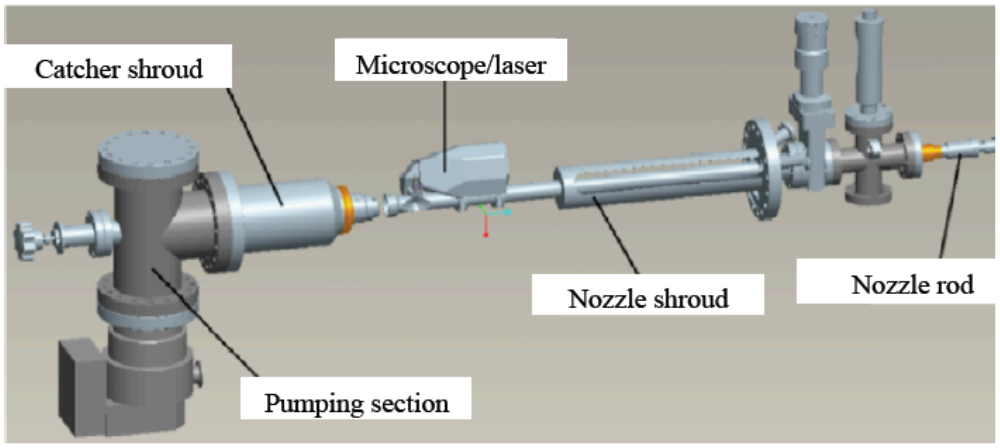

Figure 3.11: **Injector design for initial LCLS pump-probe experiments.** Schematic overview of the injector design used for pump-probe experiments utilizing the femtosecond nanocrystallography technique.

Experiments at AMO beamline in June 2010

The principle setup for the femtosecond nanocrystallography experiments in June 2010 was similar to the Dec. 2009 (Chapman et al., 2011) and the experiments were done *in vacuo* using the CAMP chamber (Struder et al., 2010). However, several improvements to the experimental design and LCLS operation were implemented in time for the June 2010 beamtime that were not available in Dec. 2009 and are discussed below.

The repetition rate of the LCLS was increased to 60 Hz, and 2 keV X-rays were available and used at the AMO beamline of the LCLS with a 27-μm^2 focus. Due to the changes in the focus and X-ray energy, the samples were exposed to a fluence of 4000 J/cm^2.

The injector was redesigned for the pump-probe experiments. One critical change made to the injector was the introduction of a gate valve, which would allow for the removal of a nozzle from the main chamber without causing the entire chamber to be brought to atmosphere, as was the case in the initial femtosecond nanocrystallography experiments. The redesigned injector is shown in Fig. 3.12. The major features

incorporated into the new injector design were an optical microscope in the shroud region, which allowed the nozzle and jet to be monitored visually from the control room, and a muti-channel fiber optic that ran from the outside of the chamber to the shroud region, which allowed an optical laser to be directed onto the liquid jet within the shroud. The inclusion of the optical microscope allowed the imaging of the incident pumping laser and X-ray free-electron laser on the liquid jet, as shown in Fig. 3.12, which could be used for alignment purposes.

One of the major objectives of the June 2010 femtosecond nanocrystallography experiments was to collect pump-probe data on PSI-Fd co-crystallites and PSII crystallites using an optical pump laser. Optical lasers were available and were used to irradiate the liquid jet using the fiber-optic in the injector. The laser options were an 790-nm Ti-sapphire laser, a 1064-nm Nd:YAG laser, and the frequency-doubled wavelengths of 395 nm and 532 nm, with the 532 nm wavelength ultimately chosen for the experiments.

One of the pump-probe experiments done in June 2010 was designed to excite PSI-Fd crystallites with a 532-nm optical pump laser. The aim was to collect X-ray diffraction data of the light induced process—Section 1.2—that occurs upon exposing the complex of PSI and ferredoxin to light (PSI charge separation electron transfer from PSI to ferredoxin conformation changes leading to movement and ultimately undocking of the ferredoxin from PSI).

An overlay of the laser wavelengths and the room temperature PSI absorbance spectrum is shown in Fig. 3.13. The chosen optical pump laser, which was a frequency-doubled Nd:YAG laser, operating at 532 nm wavelength, was ultimately used for all pump-probe experiments. The optical laser was brought in through a fiber-optic cable in the injector housing, and was used in continuous mode.

For PSI-Fd, -10-μs, 5-μs, and 10-μs delays between the optical laser and the

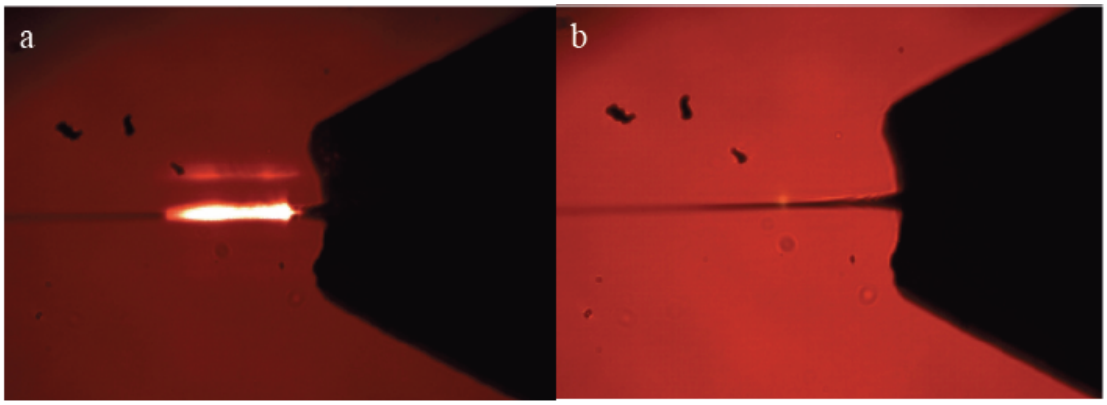

Figure 3.12: **Incident pump laser and probe laser on the liquid jet.** Images showing (a) the alignment of the incident pump laser on the liquid jet and (b) the faint plasma radiation caused by the interaction of the X-ray free electron laser with the liquid jet.

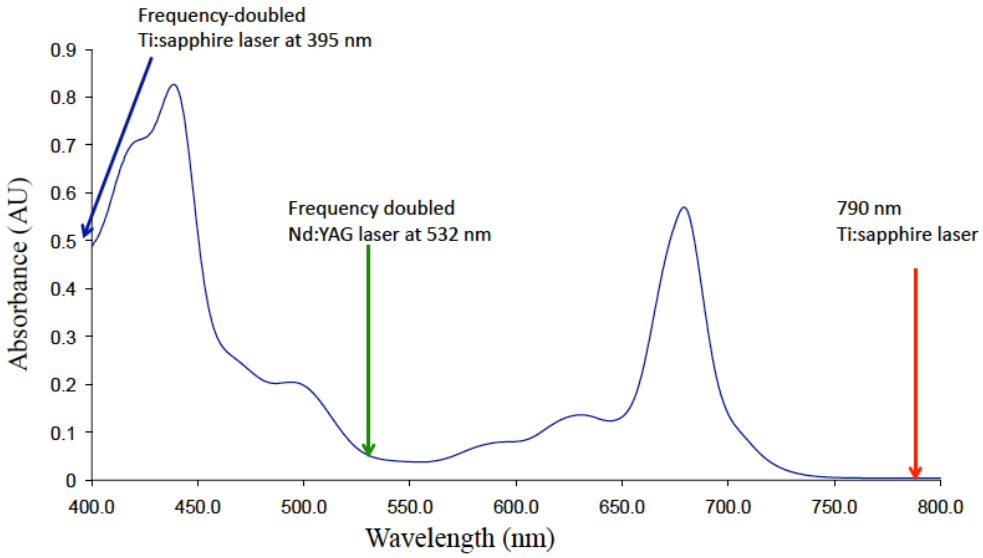

Figure 3.13: **Room temperature absorbance spectrum of Photosystem I.** The room temperature absorbance spectrum of Photosystem I is shown with the laser wavelengths available for the initial pump-probe experiments on Photosystem I-ferredoxin co-crystallites. The frequency-doubled Nd:YAG was ultimately chosen due to considerations of molar absorptivity as well as total laser power.

LCLS X-ray pulse were used as the time points, with a 70-fs pulse duration of the LCLS pulse. The PSI-Fd crystallites were inline filtered using a 2-μm or 5-μm filter, and all experiments were carried out in the dark. Green LEDs were used to provide some background light but were used sparingly; generally, the light was reflected from walls and onto the sample, never direct exposure. All red lights in the AMO hutch were covered in order to protect the sample from light exposure.

Pump-probe experiments were also designed for Photosystem II using the setup in exactly the same way as for the PSI-Fd experiments. For PSII, -10-μs and 5-μs delays between the optical laser and the LCLS X-ray pulse were used as the time points, with a 70-fs pulse duration of the LCLS pulse. The PSI-Fd crystallites were inline filtered using a 5μm filter, and all experiments were carried out in the dark.

In addition to the pump-probe experiments, femtosecond nanocrystallography experiments were carried out using PSI in order to collect data for various LCLS X-ray pulse durations. The experiments were done in order to study the sample damage incurred by exposing the sample to the intense X-ray pulses. The data were collected without an optical pump laser and pulse durations of 70 fs, 100 fs, 150 fs, 200 fs, 250 fs, and 300 fs.

High resolution femtosecond nanocrystallography

Femtosecond microcrystallography experiments were carried out at the Coherent X-ray Imaging (CXI) beamline (Boutet and Williams, 2010) of the LCLS using 6.8 (λ = 1.8 Å) and 9.3-9.4 keV (λ = 1.3 Å) X-rays, with pulse durations between 10 and 50 fs in Feb. 2011 using Photosystem I as a test case. These were the first experiments at the new CXI beamline, and the beamtime was done during the commissioning of the beamline.

The experiments were performed at 120 Hz using pulse energies between 0.3 and 3 mJ with photon flux between 2 x 10^{11} and 2 x 10^{12} photons/pulse and were done

in vacuo (base pressure: 10^{-7} torr, working pressure: 10^{-6} torr). The X-ray focus at the sample was approximately 6 μm^2. The fully hydrated nanocrystals were delivered to the interaction region using the liquid injector setup as described above for the June 2010. The size of the PSI crystallites was restricted by pre-filtering and inline filtering the sample before injection into the interaction region using a 5-μm filter. Photosystem I crystallites from eight different protein batches were used in the collection of the data, as well as re-crystallized protein.

A novel instrument to prevent sample settling—developed by Dr. Robert Shoeman of the Max Planck Institute in Heidelberg—was tested during the experiments. This anti-settling device consisted of 360° rotatable mount on which a syringe-driven pump was installed. The nanocrystal and microcrystal suspensions of PSI were loaded into 1-mL or 2-mL stainless steel syringes. Since the re-suspended PSI nanocrystals and microcrystals were in the sample reservoir the majority of the experiment, the rotating syringe would prevent the sample from settling appreciably (although this device would not prevent settling in the lines). The metal syringes were necessary to prevent sample loss in the case of the liquid injector clogging, as results from previous experiments showed that the syringe-driven pump will break the glass syringes in the case of clogging—due to syringe pumps being driven by flow rate, and not a constant force—causing a loss of sample.

A new type of detector developed by researchers from Cornell University was used during the initial experiments at the CXI beamline, and was operated with a read-out frequency of 120 Hz. The detector used a modular design, with 32 CMOS (individual tiles) (Koerner et al., 2009).

Table 3.3 summarizes the different femtosecond nanocrystallography experiments. Now that the LCLS has a beamline operating at wavelengths of the order of a carbon-carbon bond, most of the future LCLS work regarding bioimaging and nanocrystallography will be done at higher energies.

Beamtime	Dec 2009	June 2010	Feb 2011
Beamline	AMO	AMO	CXI
Instrument	CAMP	CAMP	CXI
Detector	pnCCD	pnCCD	Cornell-SLAC
Number of detectors	2; front, back	1	1
Detector readout (Hz)	Up to 200	Up to 200	120
X-ray energy (keV)	1.8	2.0	6.8-9.4
Repetition rate (Hz)	30	60	120
Nominal pulse duration (fs)	10-250	10-300	10-50
Flux (photons/pulse)	10^{11}-10^{12}	10^{12}-10^{13}	10^{11}-10^{12}
Irradiance (W/cm^2)	10^{16} (70 fs)	10^{17} (70 fs)	10^{17} (40 fs)
Total energy per pulse (mJ)	0.3	3	2-3
Fluence (J/cm^2)	900	4000	25000
Samples	PSI	PSI, PSI-Fd, PSII	PSI
Size cutoff (μm)	2	5	5

Table 3.3: **Comparison of the fs nanocrystallography experiments to date** Comparison of the experimental parameters of the femtosecond nanocrystallography experiments completed so far. Most of the future work of femtosecond nanocrystallography will be done utilizing the CXI instrument.

Chapter 4

RESULTS AND DISCUSSION

The difficulty of growing large, well-ordered protein crystals and X-ray-induced radiation damage are the major issues of X-ray protein crystallography. Crystallization screens more commonly result in conditions that produce protein nanocrystals and microcrystals, but because of radiation damage, the crystallites cannot be used for structure determination using conventional X-ray crystallography. All experiments reported here were done to establish the use of nanocrystals and microcrystals of membrane proteins for structure determination utilizing the LCLS and other fourth-generation X-ray sources, alleviating the necessity to grow large protein crystals and avoiding radiation damage. Many different experiments were necessary during the course of the work, including experiments pertaining to protein purification, protein crystallization, sample injection, sample preparation, data analysis and quality control and assessment.

In order to establish the use of protein nanocrystals and microcrystals for structure determination, the reproducibility of nanocrystal formation had to be verified starting with a protein harvested from natural sources. The femtosecond nanocrystallography technique was envisioned for membrane proteins which are notoriously difficult to crystallize. Photosystem I (PSI) was chosen as a model system because it is the largest and most complex protein crystallized to date. The large size (MW = 1,035,000 Da) and complex nature of the protein (36 protein subunits and 381 cofactors per trimer) (Jordan et al., 2001) would ensure that PSI was a difficult test case, and consequently, success with PSI would indicate general applicability of the technique.

4.1 General Photosystem I purification

Motivation

Photosystem I (PSI) is an integral-membrane protein of high natural abundance in photosynthetic organisms, and purification and crystallization procedures for large single crystals has been established (Fromme and Witt, 1998). In cyanobacteria, PSI can exist as a monomer (PSI_M) or as a trimer (PSI_T), but the focus of the work was the isolation and purification of PSI_T. A summary of the procedure is described below. An overview schematic of the overall process is shown in Fig. 4.1, with experimental details provided in Section 3.1.

PSI preparation experimental overview

The Photosystem I purification procedure was optimized in order to produce large yields of highly pure PSI_T for use in nanocrystallization and microcrystallization experiments. In order to isolate large amounts of the protein, *Thermosynechococcus elongatus* cells were grown under well-defined light and nutrient conditions with constant monitoring of the growth rate and cell vitality. Cells were harvested in the exponential growth phase and frozen in a dense pellet without cryo-protectants. The cells were lysed using microfluidics to prepare the thylakoids, and multiple centrifugation steps were conducted to remove proteins in the cell lysate from the thylakoids. The PSI protein complex was extracted from the thylakoids by incubating the thylakoid-containing sample in a solution with mild detergent to solubilize the membrane-protein complexes in the form of protein-detergent complexes. After solubilization, the sample was centrifuged with a large centrifugal force in order to sediment the PSI trimer together with the thylakoid-membrane fragments leading to a supernatant containing the remaining solubilized membrane proteins. Fast-protein liquid chromatography purification using a strong anion-exchange resin was used to separate the different membrane protein

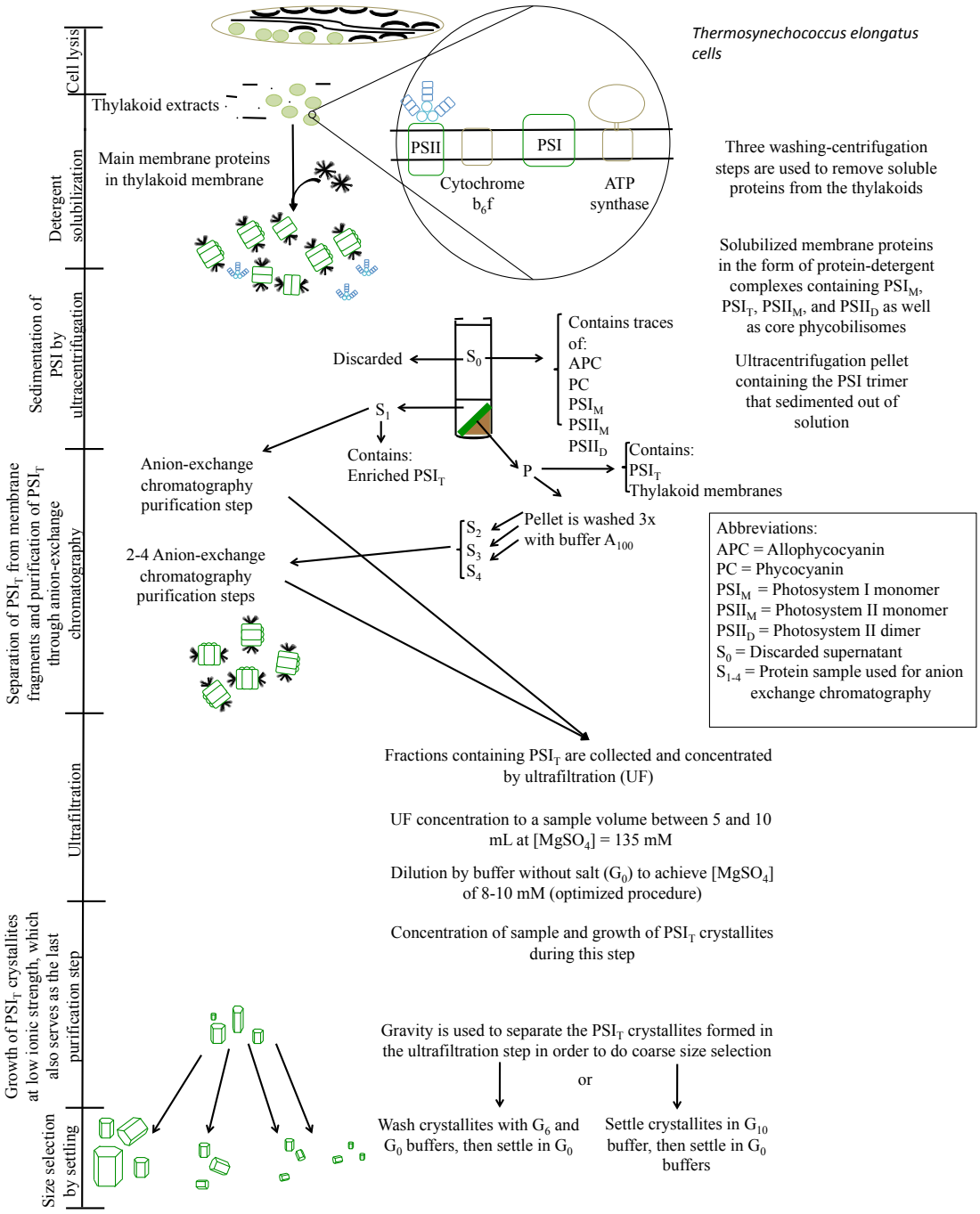

Figure 4.1: **Overview of PSI work** In order to produce Photosystem I crystallites, cells are taken and lysed using microfluidics. The thylakoids are cleaned and the PSI is solubilized from the thylakoid membranes by detergent extraction. The PSI-detergent complex was sedimented out of solution by ultracentrifugation. Anion-exchange chromatography is used to purify the protein-detergent complexes contained within the detergent extract. The fractions containing PSI are collected and concentrated before being crystallized in low salt buffer. The PSI crystallites are segregated into different sizes by settling experiments, and the samples are used for femtosecond nanocrystallography.

components of the sample. The fractions containing the PSI trimer were collected and concentrated using ultrafiltration.

PSI preparation results

Summary

In general, the overall quality of the PSI nanocrystals and microcrystals (including yield, size, and diffraction quality) produced—Section 4.2—were found to be highly dependent on the quality of the protein being used in the crystallization experiments. Many experimental parameters were important to maintaining a consistently-high protein quality, with the cell vitality and physiological state upon freezing serving as the most difficult steps for the reproducibility and maintenance of constant conditions for the PSI crystallization experiments. The reason is that the environmental conditions (light conditions, Fe and P content of media, etc.) will have a direct effect on the PSI_M-to-PSI_T ratio. Cells under optimal low-light conditions have a PSI:PSII of 8:1, with the majority of the PSI in the trimeric form. The amount of ultracentrifugation supernatant used and the separation on the resin during the FPLC purification steps were also important to the quality of the PSI nanocrystals. Other factors that were important for highly pure protein were the wavelength of maximum absorption in the wavelength range of 670-682 nm (which indicates the PSI (680 nm) to PSII (672 nm) ratio in the thylakoid membrane), the length of the ultrafiltration step, as well as the separation achieved during the anion-exchange chromatography purification step; all were proven to be critical for high quality PSI sample preparations.

Cell growth

PSI trimer was isolated from thermophillic cyanobacterium (*T. elongatus*) cells. The cells were grown in a large, 122-L bioreactor at 56°C, with air—that had been enriched with 2% CO_2—bubbled at 10 L/min.

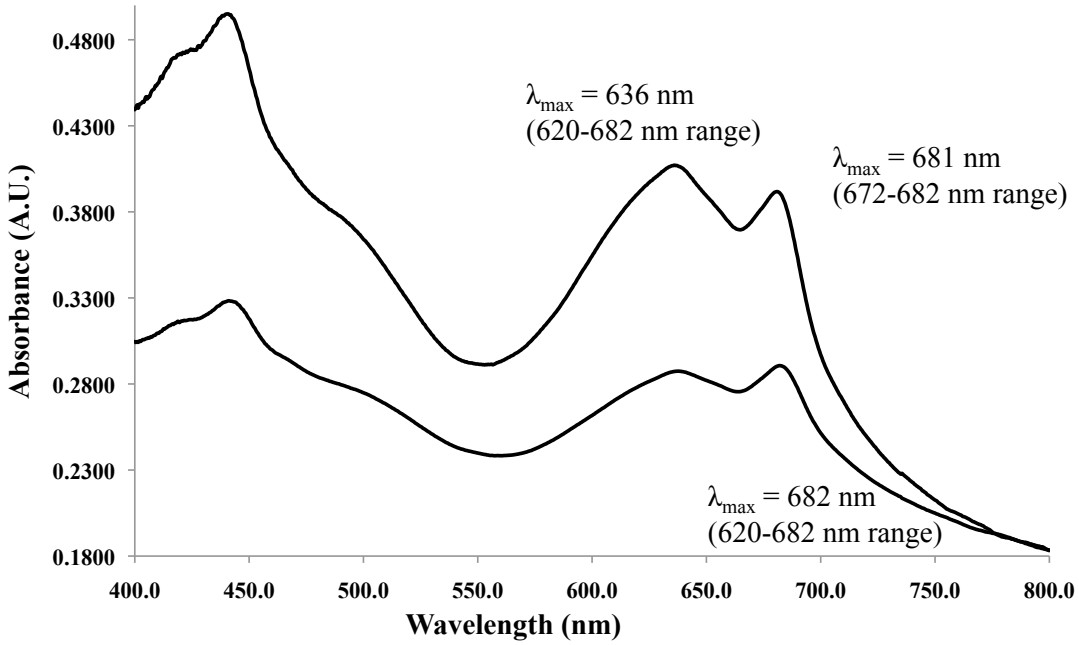

Figure 4.2: **Cell spectra** The cell spectra of two different cultures of *T. elongatus cells*. The peak at 440 nm corresponds to the Soret bands of the chlorophyll molecules, the peak at 630 nm corresponds to the phycobiliproteins, and the peak at 670-680 nm corresponds to absorbance of chlorophyll *a* molecules. The top plot was produced from cells of higher density than the bottom plot. The two spectra shown here were chosen to show one case of a cell culture with appropriate light intensity (bottom, as indicated by a A_{630}:$A_{680} \leqslant 1$) and one case of a cell culture that has too low of light intensity (top, as indicated by a A_{630}:$A_{680} \geqslant 1$).

The light intensity is crucial to the stabilization of trimeric PSI in native membranes. In low-light conditions, the cells produce trimeric PSI, with a PSI-to-Photosystem II ratio of approximately 8:1. However, in higher-light conditions, the cells degrade the trimeric PSI to monomeric PSI, and a decreased ratio of PSI to Photosystem II (PSII) results.

The ratio of PSI:PSII is indicative of the physiological state of the cells and can be monitored by determining the local maximum in the spectrum between 660 nm and 690 nm. Isolated PSII has an absorbance maximum in the red at 672 nm, whereas isolated PSI has an absorbance maximum in the red at 680 nm. Due to the differing protein environments for the chlorophylls of PSI and PSII, cells will have a

local maximum around 673-674 nm for stress conditions in which the PSI_T is degraded, whereas the cells will have a local maximum around 681-682 nm for conditions in which the PSI_T is the primary photosystem in the membrane. The local absorbance maximum could shift between 674 nm and 681 nm, depending on the light conditions. If the local absorbance maximum shifted even slightly to the blue, the light intensity was immediately decreased, while monitoring the peak shift back to 680-681 nm over a period of days.

In order to maintain the vitality and physiological state of the cells, the cells were monitored by taking absorbance spectra between 400 and 800 nm. Two spectra of *T. elongatus* cells are shown in Fig. 4.2. Prominent peaks are seen at approximately 440 nm (with shoulders at approximately 410 nm and 470 nm), 630 nm, and 672-681 nm. The three peaks correspond to absorbance due to the Soret bands of chlorophyll molecules in the sample (440 nm); absorbance due to the phycobiliproteins, with the primary contribution coming from phycocyanin for *T. elongatus* (630 nm); and absorbance by chlorophyll *a* molecules (672-681 nm). The top plot of Fig. 4.2 is a cell culture of higher density than the bottom plot. The light conditions could be monitored by observing the relative peak heights of the 630 nm and 672-681 nm peaks; for too low of light, the 630 nm peak would become large relative to the 672-681 nm peak (a ratio of the absorbance at 630 nm to 680 nm greater than one is to be avoided). In the case of the top plot of Fig. 4.2, the light intensity would be increased.

Thylakoid isolation

The thylakoid isolation was done to quickly separate the PSI-containing thylakoids from soluble cell fractions that contain proteases. The amount of time taken to accomplish the isolation should be minimized in order to diminish the degradation of the PSI by the liberated proteases.

The protocol utilized for the isolation and purification of PSI was taken and

adapted from (Fromme and Witt, 1998). The thylakoids were isolated by rupturing the cells through sheer force, followed by washing the thylakoids by re-suspension and centrifugation in isolation buffer, with the supernatant removed.

Table 4.1 shows the amount of cells used as well as chlorophyll yields during a number of PSI preparations. In total, 24 PSI-preparations were conducted during the course of the femtosecond nanocrystallography project. The amount of cells used was rarely above 35 g, because five FPLC runs were required for this mass of cells, with either four or five FPLC runs completed in every preparation, as discussed in Section 4.1.

Modification and optimization of the thylakoid-isolation protocol were done to further improve the results of the protein preparation. The amount of washing steps between the cell lysing and solubilization step was varied to include an additional washing and centrifugation step, to increase the washing-buffer-to-pellet ratio in an attempt to remove more of the soluble and membrane peripheral proteins, such as rubisco, proteases, allophycocyanin, and phycocyanin, from the thylakoid fraction before solubilization. A high content of the contaminant proteins reduces the effectiveness of the anion-exchange chromatography, as explained in Section 4.1. However, the extra washing steps must be balanced against the extra amount of time necessary to re-suspend the pellets for additional washing steps. Unfortunately, the increased number of washing steps did not have an effect on later purification steps, but the increased amount of time could cause the protein to be subject to degradation.

Solubilization of Photosystem I by detergent extraction

PSI was solubilized in the form of detergent micelles by extraction with the mild, non-ionic detergent β-dodecylmaltoside. The PSI sample was subsequently centrifuged at high speed in order to produce a pellet containing the thylakoid-membrane fragments and sedimented PSI_T, separate from the other, smaller solubilized membrane-proteins

Prep Identifier	P	O	U	R	Q	S	T
mass of cells (g)	22.3	19.9	40.7	32.5	25.2	32	35.7
Chl amount after cell lysate washes	144	92.3	370	183	124	207	164
Chl amount in S1 (μmole)	13.6	6.07	10.8	15.0	13.4	14	10.6
Chl amount in S2 (μmole)	27.9	19.5	32.4	41.3	26.6	45.7	39.4
Chl amount in S3 (μmole)	9.12	9.74	6.37	14.2	12.3	14.8	12.9
Recovered Chl after FPLC runs (μmole)	36.7	16.6	45.6	42.1	37.0	59.8	38.2
Chlorophyll loaded on FPLC (%)	50.6	35.3	86.2	70.5	52.3	74.4	38.2
Chlorophyll recovered/chlorophyll loaded on FPLC (%)	75.9	47.0	86.2	81.7	70.9	80.3	63.3
Chlorophyll recovered/mass of cells (μmole/g)	1.65	0.863	1.12	1.30	1.48	1.87	1.07

Table 4.1: **Cell mass used and yields achieved for PSI preparations** The cell masses used and yields achieved for seven PSI preparations used in the serial powder diffraction and femtosecond nanocrystallography experiments. The preps were identified using a naming scheme that differentiated the preparations based upon a letter of the alphabet.

such as PSI_M, the Photosystem II monomer ($PSII_M$), the PSII dimer ($PSII_D$), and ATPsynthase. The top 20 mL of the supernatant of each centrifuge tube was discarded, whereas the remaining 2-5 mL of the supernatant was retained and used for later FPLC purification steps, as shown in Fig. 4.1. The pellet was washed to remove the PSI_T from the top of the pellet, increasing the total yield of PSI extracted during the PSI preparation.

The PSI trimer was partitioned between the supernatant and top layer of the pellet of the ultracentrifugation run. While removing the supernatant of the ultracentrifugation step, care was taken to save the supernatant that visibly had less red color and more of a green color, which was typically the remaining 2-5 mL above the pellet. The spectra of S_0 amd S_1 are shown in Fig 4.3. The red fluorescence observed by eye from S_0 is indicative of large amounts of phycocyanin because phycocyanin exhibits red fluorescence when exposed to light. The spectrum of S_0 also shows the presence of a much higher concentration of carotenoids in S_0 than in S_1. Additionally, PSI monomer and PSII monomer were present in more abundance in S_0 than S_1, and consequently, most of the supernatant was discarded, but approximately 20-50 mL total volume that was in the 2-5 mL directly above each pellet was retained, as shown in 4.1.

PSI_T is the largest of the protein-detergent complexes, leading to a large sedimentation coefficient. Therefore, the PSI_T sediments during the ultracentrifugation step and is partitioned between the supernatant and the top portion of the pellet. As a result, in order to increase the yield of PSI_T, the PSI_T was extracted from the pellet by sequential washing steps. The washing steps were allowed to occur until the next sample was needed for contiguous FPLC purification steps—Section 4.1.

The pellet was washed with A_{100} buffer; the solution was swirled while not allowing for bubble formation. The pellet was allowed to incubate in A_{100} buffer for at least one hour at 4°C (but usually during the entire concurrent FPLC run), with constant shaking of the solution above the pellet by placing the centrifuge rack directly

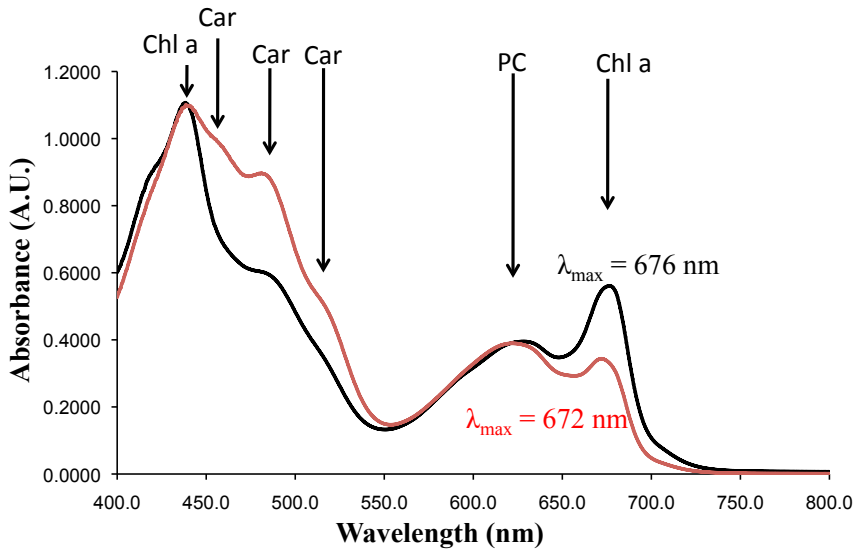

Figure 4.3: **Ultracentrifugation supernatant spectra** UV-VIS spectra of S_0 (red plot) and S_1 (black plot). The plots were normalized with respect to absorbance at a wavelength of 440 nm. The increased red fluorescence of S_0 that is observed by eye is due primarily to phycocyanin. Phycocyanin has a blue color (absorbance at 630 nm) but shows strong red fluorescence that is visible by eye. As a result, S_0 has a large absorbance at 630 nm relative to the absorbance 680 nm; the 630-nm:680-nm ratio will depend on the protein preparation and the physiological state of the cells. Additionally, the larger shoulder at approximately 470 nm in the S_0 spectrum is indicative of larger numbers of carotenoids. The case of the spectrum of S_0, λ_{max} = 672 nm indicates a higher percentage of PSII, whereas in the case of the spectrum of S_1, λ_{max} = 676 nm indicates an increase in the percentage of PSI relative to PSII. The absorbance bands are labeled by the molecule excited at the particular wavelength, with Chl a standing for chlorophyll a, Car standing for carotenoids, and PC standing for phycocyanin.

on a shaker operating at 20-30 RPM.

Although the S_1 solution was enriched in PSI_T, there were still many smaller soluble proteins and smaller protein-detergent complexes contained within the solution, such as APC, PC, PSI_M, $PSII_M$, and $PSII_D$. All of these proteins could contribute to poorer separation in the FPLC purification steps—Section 4.1—where increased purity of the PSI_T sample was desired. One way to achieve higher purity would be to decrease the partitioning of the PSI_T in the supernatant, and attempt to sediment more of the protein in the pellet, while keeping the smaller soluble proteins and protein-detergent complexes in S_1. The ultracentrifugation step was optimized by lengthening

the spin duration in order to decrease the amount of PSI_T in the supernatant. Increasing the duration of the centrifugation from 90 min to 105 min increased the total amount of chlorophyll within the second pellet wash, while reducing the amount of chlorophyll in the supernatant, and keeping the amount of chlorophyll in the first pellet wash approximately the same. The increased amount of chlorophyll within the pellet washes was found within the PSI_T peak, as discussed in Section 4.1. Resultingly, the protocol was changed to accommodate these results.

Anion-exchange chromatography purification of PSI

Anion-exchange chromatography (AEC) was used to purify the PSI from the other solubilized membrane proteins and soluble proteins after the ultracentrifugation step. The anion-exchange resin separated the different protein species based upon affinity for the positively charged functional group of the resin.

The column was initially equilibrated with a buffer containing 100 mM $MgSO_4$. Under these conditions, APC and PC did not bind to the anion-exchange resin and were eluted during the application of the sample to the media, whereas PSI and PSII were bound to the column. The Photosystems were eluted from the column by a linear concentration gradient from 100 mM to 150 mM $MgSO_4$. The PSI-trimer eluted as the third peak, as shown in Fig. 4.4.

Figure 4.4 shows a set of typical chromatograms recorded during a PSI preparation for S_1 as well as the pellet washes S_{2-4}. The first peak in the chromatogram contains mainly phycobiliproteins, specifically PC and APC. The second peak, when resolved, contains the PSI_M, inactive $PSII_M$—there was no Ca^{2+} in the running buffers, causing the disassembly of the active $PSII_D$—whereas the third peak contains the PSI trimer. Although the peak separation became better with each subsequent FPLC run during a preparation, baseline resolution – defined here as an absorbance between two peaks that is 10% of the max absorbance of the peaks – was not always achieved. Con-

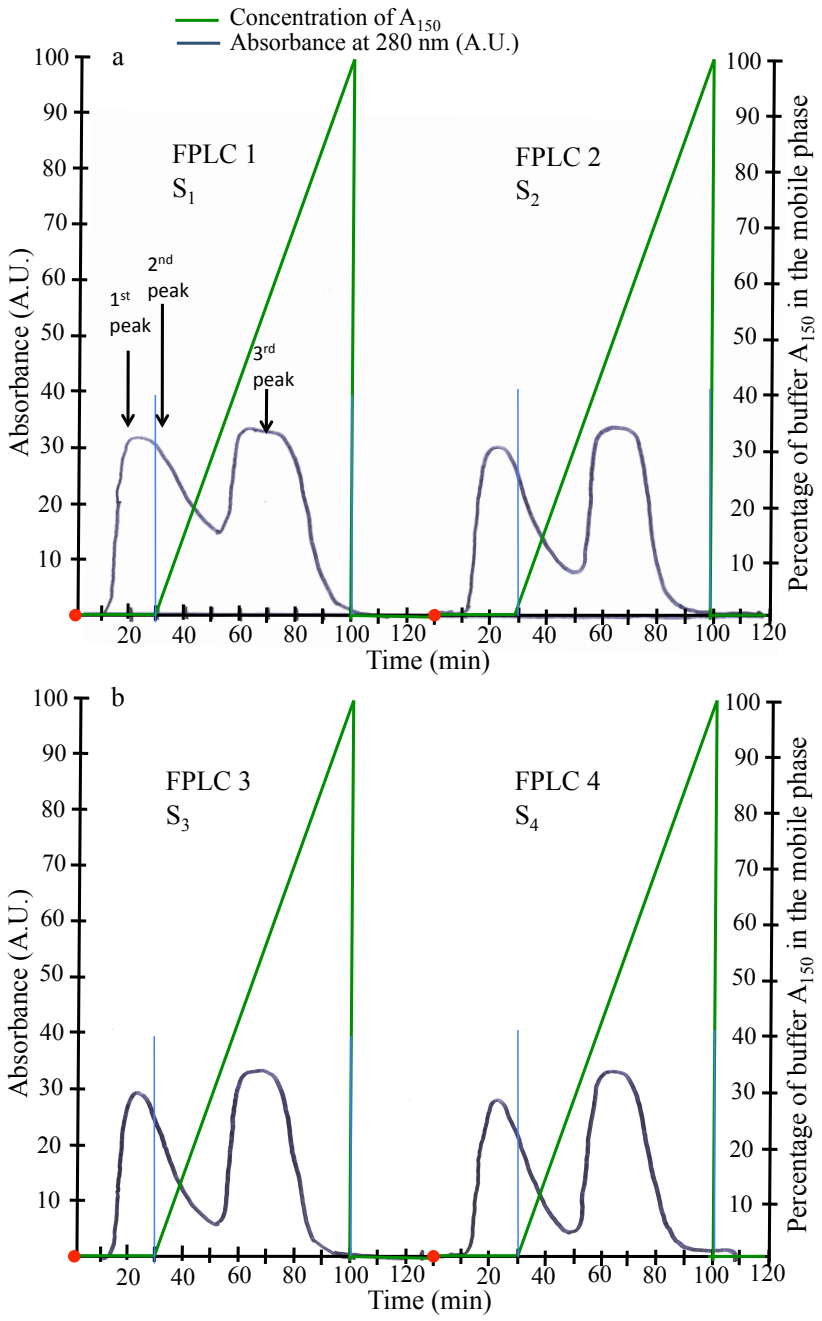

Figure 4.4: **Chromatogram of Photosystem I separation** The chromatogram show the separation of the PSI sample during four subsequent FPLC purification steps using anion-exchange chromatography. The 1st peak corresponds to phycobiliproteins, most notably phycocyanin; the 2nd peak, which is often a shoulder and poorly resolved, corresponds to monomers of PSI as well as monomers and dimers of PSII; and the 3rd peak corresponds to the PSI trimer peak. In the case of the chromatograms, the second peak cannot be resolved from the first peak.

sequently, the first one or two fractions of the third peak (totaling 20 mL) were omitted from collection for any FPLC run that did not achieve the aforementioned baseline resolution. The reason the fractions were omitted was the possible presence of PSI monomer in the second peak. A PSI monomer to PSI trimer ratio of 1 to 10,000 is sufficient to inhibit the growth of PSI trimer crystals (Fromme and Witt, 1998). Although the AEC protocol could have been modified by lengthening the gradient, a protease can bind with PSI_T on the column, degrading the PSI_T. As a result, the length of time of PSI_T on the column had to be minimized and the elution time was minimized. Therefore, no step gradient with increased washing times was introduced.

The percentage of the chlorophyll initially loaded onto the column that was contained within the PSI_T peak changed dramatically between the supernatant FPLC run and all pellet washes. For instance, the percentage of loaded chlorophyll that eluted in the PSI trimer band was typically 60% for the ultracentrifugation supernatant runs, whereas the percentage reached 80% for the first pellet wash and up to 90% for the second pellet wash. Ergo, as many pellet washes as needed were conducted, until the washes were significantly reduced in chlorophyll content, typically by the third wash.

The first FPLC run of each PSI prep typically had the worst peak separation, as evidenced by the chromatograms in Fig. 4.4. S_1 was used for the first FPLC run because there was some PSI trimer contained within the volume and to allow the "washes" of the ultracentrifugation pellets to proceed, which contained a much higher amount of PSI trimer. However, care was taken to avoid fractions collected toward the beginning of the third peak of the separation, that is if the second and third peaks were not resolved well enough. If the separation for a particular FPLC purification run was very poor, none of the eluted peaks would be saved because the entire sample could be contaminated with PSI monomer. SDS-PAGE characterization of the collected fractions of the AEC purification step is discussed in Section 4.1.

The amount of additional soluble and peripheral membrane proteins in S_1 could

potentially make the binding of the PSI on the anion-exchange column less efficient, leading to a broader PSI-trimer peak. The protein added to the column will be bound by the positively charged amine groups at the top potion of the resin, thereby diminishing the binding capacity at the top of the resin. The protein will travel further into the column before being bound, causing a broadening of every peak. Although this could explain the lack of baseline resolution in the chromatograms, changing the amount of washing steps prior to solubilization did not lead to better separation. These results could be interpreted as indication that the core of the phycobilisomes may not be removed during the washing steps of the thylakoid preparation.

By the fifth FPLC run, the separation on the column resin became worse and the resin looked orange due to binding of carotenoids and lipids. Before more FPLC runs were to be done, the column would need to be subjected to the proper cleaning procedure. As a result, the PSI preps were usually done with $\leqslant$ 35 g of cells to allow for a maximum of 5 FPLC purification runs.

SDS-PAGE of PSI eluant and concentrated protein

SDS-PAGE characterization was done on the collected fractions of the AEC purification steps in order to ascertain the purity of the Photosystem I sample with respect to other protein contaminants, such as Photosystem II, rubisco, and ATP synthase. SDS-PAGE characterization was done using a 20% polyacrylamide gel in order to separate the small and medium molecular-weight subunits of PSI; for details of the SDS-PAGE protocol, see Section 3.5.

The collected volume of fractions from the third peak of each AEC purification run was homogenized and characterized by SDS-PAGE. Results of the three FPLC purification steps made during a representative PSI prep are shown in Fig 4.5. Lane 2 shows the sample from the first AEC purification step, which corresponded to the retained supernatant of the ultracentrifugation step. Lanes 3 and 4 show the samples

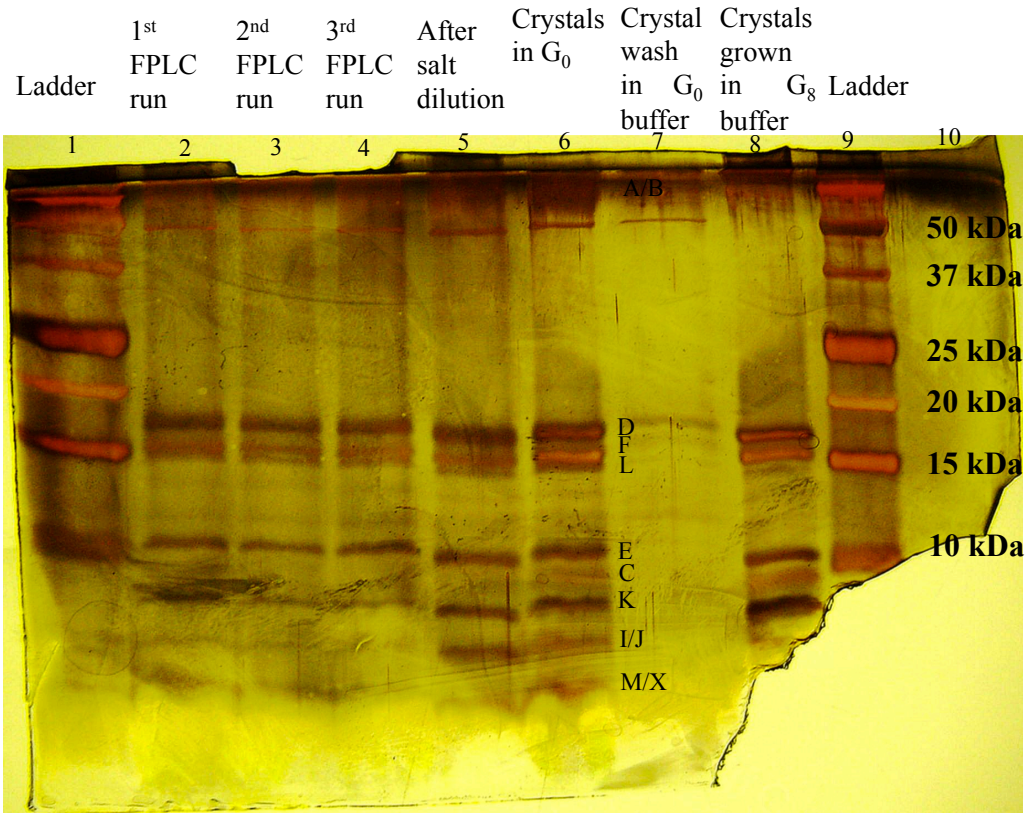

Figure 4.5: **SDS-PAGE characterization of purified PSI** An SDS-PAGE characterization showing the purity of the protein collected from the third peak of the anion-exchange chromatography purification steps described in 3.1.

from the second and third FPLC purification steps, which corresponded to the first pellet wash. Lane 5 shows the combined protein sample after the $MgSO_4$ was diluted to 8 mM in the ultrafiltration step, and lane 8 shows the results for dissolved PSI crystals that had been grown in G_8, as described in Section 4.2.

The bands in the AEC fraction are from the PSI subunits PsaD, PsaF, PsaL, PsaE, PsaC, PsaK, PsaI, PsaJ, PsaM, and PsaX subunits. The sample is quite pure and is an appropriate starting point for crystallization experiments.

Concentrating the PSI using ultrafiltration

Ultrafiltration was utilized to initially increase the PSI trimer concentration and subsequently reduce the $MgSO_4$ concentration of the PSI solution by addition of buffer

without salt (G_0). The $MgSO_4$ concentration was reduced in order to lower the ionic strength of the solution, facilitating the salting-in crystallization of PSI explained in 4.2.

The PSI was concentrated by applying pressure, with nitrogen gas, to the PSI mixture in a 400-mL ultrafiltration stir-cell (Amicon) using a 100-kDa cutoff membrane. The solution was concentrated until a volume of 5-10 mL was reached, and subsequently the solution was diluted with G_0 buffer to a final $MgSO_4$ concentration that was varied between 4 mM and 12 mM.

4.2 Nanocrystal and microcrystal growth of Photosystem I

Motivation of microcrystal production of PSI

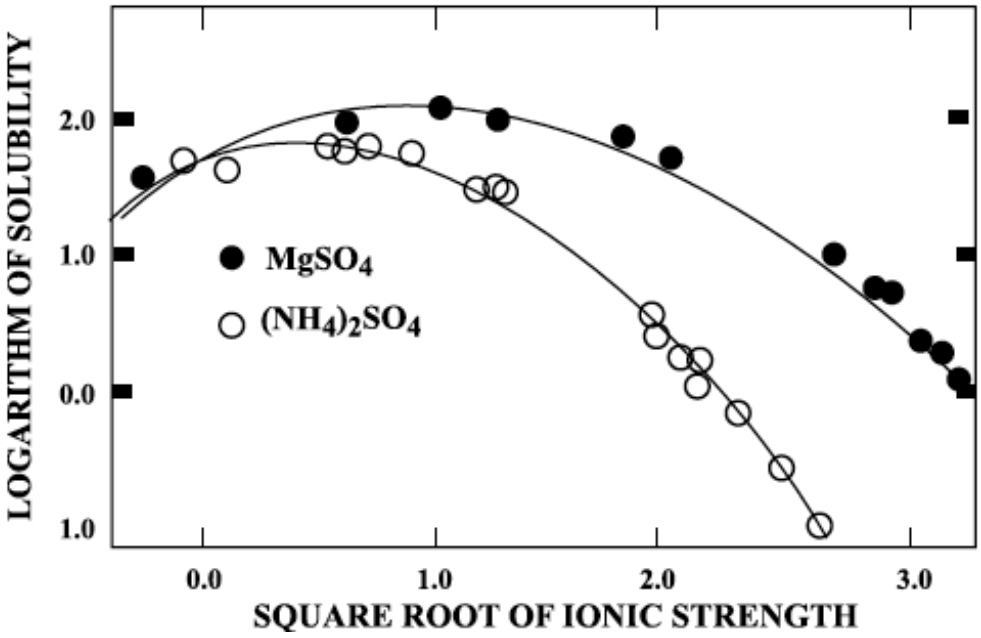

Figure 4.6: **Solubility of a typical protein versus ionic strength** The solubility of a typical protein, enolase, is shown here as a function of ionic strength produced by two different, widely used salts. The regions of the end points of the curves where solubility decreases are called, at low ionic strength, the "salting in" region, and at high ionic strength, the "salting out" region. Both provide opportunities for the creation of supersaturated macromolecular solutions and crystal growth. Figure and caption taken from (McPherson, 2004).

Photosystem I nanocrystals and microcrystals were specifically grown in order

to establish conditions for the growth of nanometer- and micrometer-sized membrane protein crystals for structural studies. PSI is crystallized using a method called salting in the protein, in which the protein is crystallized by moving to the low ionic strength side of the solubility curve, as shown pictorially in Fig. 4.6. The solubility curve is for the protein enolase, but the trends shown in the solubility curve are typical for many proteins in solution. Proteins will exhibit a maximum solubility in a moderate ionic-strength solution. The protein molecules can be made to interact either by reducing the ionic strength or by increasing the ionic strength. The precise form of the protein solubility curve as a function of ionic strength will depend on the particular protein as well as the chosen salt. However, it is more common to approach crystallization experiments utilizing the high-salt side of the solubility curve.

The PSI crystals were grown by increasing the protein concentration of the crystallization solution at low ionic strength. As low ionic strength is approached, the protein surface is depleted of counter ions, which causes the PSI trimers to form direct ionic interactions between the PSI molecules, and ultimately precipitate into a new phase. The terminology of "salting in" is used because the addition of salt causes the dissolution of the protein crystal on this side of the solubility curve.

The primary method for producing the PSI crystallites was ultrafiltration, in which the protein concentration is slowly increased through retention of the protein by a molecular-weight cutoff membrane, as shown schematically in Fig. 4.7. The ultrafiltration crystallization was done with constant $MgSO_4$ concentration while gently stirring the protein solution. Several different scenarios are shown in Fig. 4.7, but of the four shown, only condition *d* would produce conditions that would not result in crystallization, as the solution has not reached the nucleation zone. However, simply increasing the protein concentration further would allow *d* to be a viable crystallization pathway, and in addition, seeding condition *d* with a small PSI crystal would allow crystal growth. Conditions (a), (b), and (c) would result in a shower of crystals, with

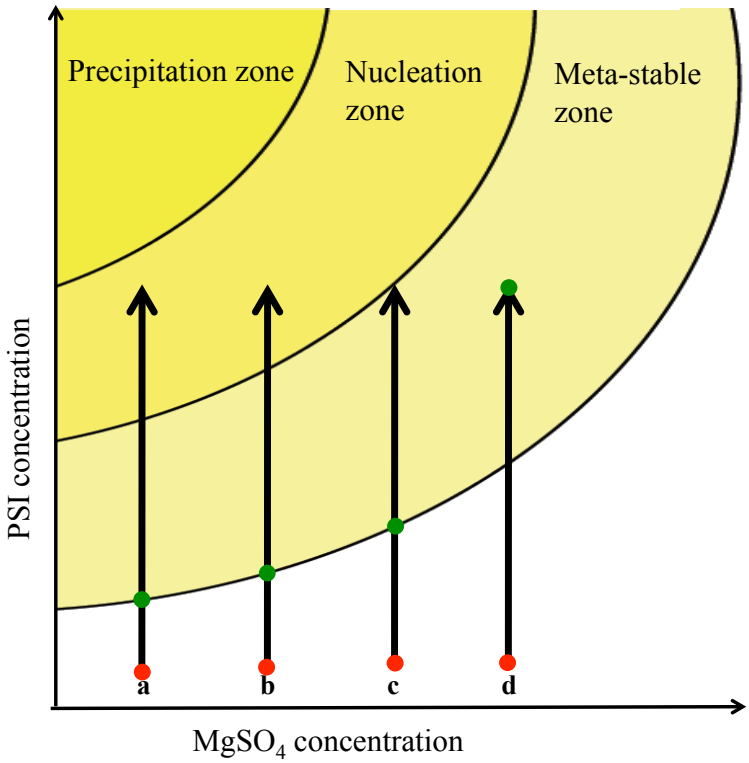

Figure 4.7: **Schematic phase diagram of Photosystem I** Schematic phase diagram of Photosystem I showing the ultrafiltration crystallization method. Lines a-d indicate different starting $MgSO_4$ concentrations for the crystallization experiments, in which the red circles indicate the starting position of a crystallization experiment and the green circles show the equilibrium positions of the sample in the phase diagram. The crystallization experiments (a,b,c) would lead to a shower of crystals and crystallites, whereas (d) would need further concentrating to produce crystalline sample. (a) would produce the conditions that led to the largest number and fastest growing crystals, whereas (c) would produce the smallest number of crystals that grew slowly comparative to (a).

the nucleation rate and crystal growth rate becomming the highest in (a) and lowest in (c) due to the difference in supersaturation.

Once the protein solution was concentrated to the desired level, and the crystallization experiments were allowed to proceed for 12 hours, a mat of PSI crystals and crystallites was found on the membrane of the ultrafiltration setup, with additional PSI crystallites found in the liquid above the membrane. The crystallization should not be permitted to occur for more than 24 hours, as any remaining contaminants should be removed from the crystal suspension with the retentate.

PSI crystallite production experimental overview

PSI was initially crystallized using ultrafiltration, a process in which the protein concentration was increased at a constant salt concentration. PSI crystallization experiments using ultrafiltration were performed using a variety of salt concentrations, by using the buffers G_4, G_5, G_6, G_8, G_{10}, and G_{12} – where G_x is a buffer containing x mM $MgSO_4$ – as a final crystallization condition, as detailed in Section 3.1.

The 5-10 mL of retentate was used to wash the membrane and remove the PSI crystals on the mat. The ultrafiltration setup was slowly rotated in order to ensure that the entire membrane was washed, as washing the entire membrane ensured that the membrane could be used for additional PSI ultrafiltration crystallization experiments. The rotation of the setup allowed the crystal mat to receive cleaning from fluid flowing in multiple directions. Upon removing the PSI crystal mat from the membrane, the sample was transferred to multiple eppendorf tubes for settling experiments. Approximately 3-5 mL of G_{10} buffer was added to the ultrafiltration unit to ensure the membrane was hydrated, and the entire ultrafiltration setup was covered and stored at 4°C. Long-time storage of the membrane should not be done in buffer; the membrane should be washed with G_{100}, followed by water, 0.1 M NaOH, and finally water.

Once the crystals were harvested from the ultrafiltration membrane and added to the eppendorf tubes, the crystals were selected by size through settling experiments. The crystals and crystallites were allowed to sediment for time durations of 10-60 min, using 10-min intervals. Two different procedures were used for settling experiments, one in which the sample was cleaned and stored in buffer G_0 before settling, and one in which the settling experiments were carried out in the G_{10} retentate of the crystallization experiment.

The G_{4-8} ultrafiltration crystallization experiments produced a larger number of smaller crystals than the experiments in higher salt (G_{10} and G_{12}), as expected, due to the higher supersaturation. However, the crystals produced in G_{4-8} appeared to be of lesser quality than those grown in G_{10-12}. More protein would dissolve in the supernatant of crystal washes using G_0 buffer than the crystals grown with the higher salt concentrations. Additionally, the crystals grown in lower salt did not survive for as long in storage; having more protein dissolve in the supernatant than the PSI crystals grown in the higher-salt conditions showed an increased solubility of the protein. Both of these results indicate that the protein crystals grown in lower-ionic-strength conditions were of poorer quality.

The solubility of PSI in the $MgSO_4$-containing solutions depends on the quality of the protein, and in addition, the solubility of PSI in the crystalline form will also depend on the order (quality) of the crystals. The order of the crystals can be correlated to the supersaturation borderline, which indicates the quality of the protein plus the quality of the crystals. The reason that the quality of the crystal can be correlated to the supersaturation borderline is related to the number of crystal contacts; in the case of an ideal crystal of PSI, each PSI_T molecule has six crystal contacts with its neighbors (with each crystal contact being the formation of four salt bridges between the PSI_T molecules). A higher ionic strength would be needed to dissolve the PSI_T molecules in the ideal crystal than in the case of a crystal of identical size but containing many defects (the large number of defects leads to less crystal contact sites). Heterogeneities in the protein preparations, such as partial protein degradation of the PsaF subunit that occurs under a change in light conditions, can lead to an increased solubility of PSI. Therefore, the solubility of PSI at low ionic strengths is a direct indication of the quality of the PSI preparation and crystals.

Even in the case of a PSI preparation devoid of heterogeneities, crystal quality depends on the number of defects in the crystal. The number of defects in the crystal can be correlated to the rate at which crystal growth occurs, which is directly related to the supersaturation of the protein in solution. The higher the supersaturation, the higher the rate of crystal growth, and the higher crystal growth rate can lead to increased surface defects (admolecules, surface vacancies, or slight misalignment of the molecules within the crystal). Thereby, the crystals with a higher number of defects will dissolve at a lower ionic strength (and even in G_0), whereas the higher quality crystals grown at higher ionic strength are more stable at low ionic strength. Consequently, the washing steps with G_0 buffer presented an indication of the quality of the crystals.

The crystal-washing steps indicated that the crystal quality increased as the salt-concentration was increased from G_4 to G_8, with (approximately) constant protein concentration during the crystallization process in the ultrafiltration cell. Reducing the supersaturation of the solution for the crystallization condition was required in order to produce higher quality crystals. However, the reduction in the supersaturation could be achieved by either increasing the ionic strength of the solution while keeping the protein concentration the same, or by decreasing the protein concentration while keeping the ionic strength of the solution the same. The method chosen to reduce the supersaturation was to increase the ionic strength of the crystallization experiment, while keeping the chlorophyll concentration at approximately 5 mM.

As expected, the G_{12} produced PSI crystallites that had the lowest solubility in the G_0 buffer washes. However, the yield of crystallites in the G_{12} crystallization experiments was the lowest of the attempted conditions, due to the lowest supersaturation. Although PSI crystallites grown in G_6 and G_8 were used for preliminary serial powder diffraction experiments, G_{10} was chosen as the crystallization condition for all future experiments, due to a combination of crystal quality and a crystallite size distribution that produced a large amount of high-quality microcrystals, as discussed in the below

discussion of settling experiments.

The PSI crystallites for femtosecond nanocrystallography experiments were prepared by ultrafiltration crystallization in G_{10} by concentrating the PSI to 5-12 mM chlorophyll. Usually the nucleation was allowed to occur in a total volume of 5-10 mL of G_{10}, with subsequent growth allowed for 12-24 hours. Crystals ranging in size from submicron to greater than 50 microns in size are produced during the ultrafiltration crystallization experiments.

Harvesting the PSI crystallites from the membrane

The PSI microcrystals harvested from the ultrafiltration membrane were coarsely separated by size in subsequent settling experiments, discussed below, and used for X-ray diffraction experiments—Sections 4.6 and 4.7—as detailed in Section 3.1. The crystallites were removed by using a 1-mL pipette and gently washing the membrane while the ultrafiltration setup was kept at a slight angle of approximately 10° to allow the crystals and crystallites to flow into the retentate pooled at the bottom of the setup.

Photosystem I crystallite settling experiments

Settling experiments were devised in order to minimize the amount of PSI crystals lost in the serial crystallography experiments by segregating the PSI crystals and microcrystals by size based upon settling time. The settling experiments were completed by re-suspending 1-mL aliquots of the PSI crystal and crystallite samples and allowing the larger crystals to settle to the bottom of the eppendorf tube, forming a loose pellet. The supernatant was transferred to a new eppendorf tube, and the settling experiments were continued with a longer time scale.

Images of the crystals remaining after 10-, 20-, 30-, and 40-min settling steps are shown in Figure 4.8a-d, respectively. The images show the contents of 1-μL droplets of PSI crystallites that had been re-suspended from the pellet by gently mixing

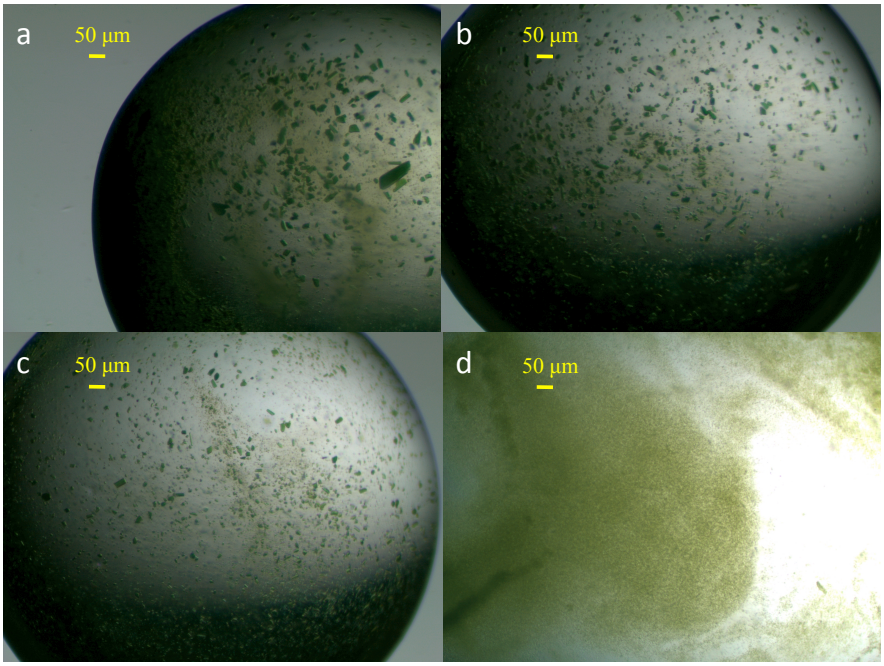

Figure 4.8: **Photosystem I crystallite size distribution** Images of the crystals contained within the Photosystem I crystal suspensions after the solution is allowed to settle for 10 min (a), 20 min (b), 30 min (c), and 40 min (d).

with a 200-μL pipette. Fig. 4.8a-d shows the sample with a total magnification of 40X.

As can be seen from the images in Figure 4.8a-d, the amount of large crystals decreases with successive settling experiments. Additionally, the average size of the crystals in the distribution decreases with longer settling experiments. Although the average size of the crystals can be controlled using the settling experiments, some large crystals will still be present within the 40-min-settling sample, making inline filtering necessary.

From the images in Fig. 4.8, the size distribution of the crystals in the pellet remaining after each settling experiment can be estimated. The smallest crystals in each of the settling steps were $\leqslant 5$ μm in size, but likely smaller than 2 μm in size. In the 10-min settled crystals of Fig. 4.8a, the size distribution is measured to be 2-100 μm in size, whereas in the 20-min PSI crystals of Fig. 4.8b, the size distribution is measured as 2-30 μm in size. The size distribution of the 30-min settled PSI crystals

Prep identifier	M	O	P
wet* mass of crystals in 10 min (mg)	245.5	243.2	349.9
Percent of total mass in 10 min(%)	60.1	60.45	65.61
wet* mass of crystals in 20 min (mg)	40.2	41.3	61.4
Percent of total mass in 20 min(%)	9.98	10.3	11.5
wet* mass of crystals in 30 min (mg)	18.7	16.5	19.3
Percent of total mass in 30 min(%)	4.64	4.10	3.6
wet* mass of crystals in 40 min (mg)	98.3	101.3	102.7
Percent of total mass in 40 min(%)	24.4	25.18	19.26

Table 4.2: **Comparison of the PSI crystallite during settling** Photosystem I crystal mass collected in the sediment of the different settling times for three representative PSI preparations. Most of the total wet* protein mass was present in the largest crystals within the 10 min settling step. 100 mg of wet mass of PSI crystals corresponds to approximately 16 mg of Photosystem I, but for a full description of wet protein mass, refer to the main text, Section 4.2.

of Fig. 4.8c is measured to be 2-20 μm in size, and the size distribution of the 40-min settled PSI crystals of Fig. 4.8d is measured to be 2-5 μm in size.

Alternatively, the crystals collected from the membrane could be washed and stored in G_0 prior to the settling experiments, which may prevent the dissolution of the smallest crystals due to the lower solubility of PSI in G_0. The initial serial powder diffraction experiments—discussed in Section 4.6—and the initial LCLS experiments in Dec 2009—discussed in Section 4.7—were carried out using crystals in the 60-min sedimentation step that had been first transferred to buffer G_0 before settling experiments. The higher-energy femtosecond nanocrystallography experiments—discussed in Section 4.11—used crystals in the 20-min and 30-min sedimentation steps that had been kept in buffer G_{10} during the settling experiments.

Table 4.2 shows the wet mass of PSI crystals collected in each settling step. The wet mass of the crystals is determined by measuring the mass of the crystals after centrifuging the crystals into a pellet and removing all supernatant. A quantity of 10 mg of PSI crystals can be dissolved in 10 μL of buffer G_{50}, leading to a chlorophyll concentration of 20 mM—using 96 chlorophyll molecules per PSI monomer (Jordan et al.,

2001) this correlates to a chlorophyll concentration of 20 mM being equal to 78 mg/mL of PSI—in solution. The chlorophyll concentration of the crystal is approximately 40 mM, leading to a PSI concentration of approximately 160 mg/mL. This corresponds to roughly 160 mg of PSI per 1000 mg of pellet (or 16% of the crystal mass is PSI). Consequently, the 102.7 mg of wet pellet in the 40-min settling step contains approximately 17 mg of PSI protein.

An inspection of Table 4.2 indicates that the PSI crystallite size distribution is bimodal, with one maximum in the 10 min settling size range, and one maximum in the 40 min settling size range. The large size distribution of the PSI crystallites produced in the ultra-filtration crystallization, as shown in Fig. 4.8, indicates that the critical nuclei were formed at various times during the subsequent incubation, with the size distribution being related to the time difference between the critical nuclei formed at the beginning and at the end of the experiments. Specifically, many PSI crystallites could be seen forming at the membrane during the concentration of the protein while the volume was larger than the incubation volume. The reason for the crystals forming at the membrane is that the mixing is not thorough, leaving a layer on the membrane surface that is not mixed, thereby increasing net concentration and the nucleation rate, due to increased supersaturation.

The nuclei formed at the membrane surface formed the largest crystals harvested from the filter and may represent the first peak in the size distribution that corresponds to the 10-min settling step. Nucleation occurring away from the membrane could be attributed to the smaller crystals and the second peak in the size distribution that corresponds to crystallites in the 40-min settling step.

The larger crystals grown directly on the membrane nuclei likely contribute most prominently to the large variation of the total size distribution and maximum crystal size harvested. The largest crystals were not used for the nanocrystallography experiments, but they can be dissolved for re-crystallization experiments, as discussed

in Section 4.3.

The amount of crystals and the size distribution of the PSI can be highly dependent on the particular protein preparation, the protein concentration, the rate of the volume reduction during the concentrating step, and the final salt concentration. Since most of the ultrafiltration crystallization was performed in G_{10}, the total amount of PSI isolated and purified during the preparation, as well as the final volume, will dictate the position of the crystallization experiment in the phase diagram. The highest percentage of crystals would be in the 10 min settling step, regardless of preparation, but the second highest percentage of crystals was found in the 40 min step.

One point to consider is that the membrane (and possibly the stir bar) could act as a nucleation point for the PSI crystals, allowing for heterogeneous nucleation at the membrane surface. The heterogeneous nucleation pathways, if available, would likely increase the variability in the size distribution, and possibly lead to crystals with larger imperfections, such as higher mosaicity. A systematic study of the morphology and internal order of the crystals found at the membrane surface could be done to further characterize the largest crystals. Additionally, the crystals grown on the membrane likely suffer from mechanical removal from the membrane surface, which increases the defects of the crystals.

SDS-PAGE results for Photosystem I crystallites

SDS-PAGE experiments were conducted on the crystalline samples produced from the ultrafiltration crystallization experiments. The SDS-PAGE characterization was examined in order to verify that the protein isolated and crystallized was PSI, and that limited amounts of contaminant protein were contained within the sample, as well as to determine whether any degradation of PSI has occurred.

A high-density polyacrylamide gel was used with an automated PHAST system for the crystalline samples, as detailed in Section 3.5. In order to detect small quantities

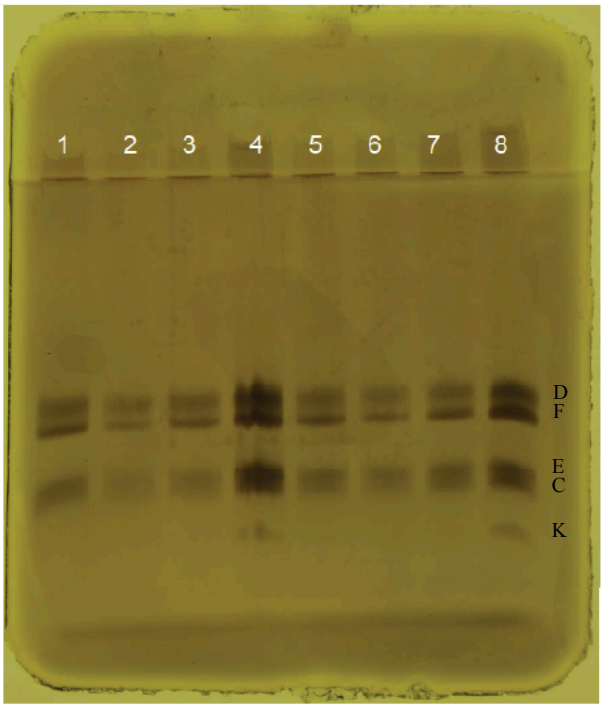

Figure 4.9: **SDS-PAGE results for Photosystem I crystallites** SDS-PAGE results of the Photosystem I crystallites produced from seven preparations. All samples were set to approximately 300-μM chlorophyll concentration, with lane 8 containing the same sample as lane 4, with 150 μM chlorophyll concentration. The designations D,F,E,C,K on the left indicate the Photosystem I subunit associated with the particular band.

of contaminants, silver staining was utilized to stain the gel.

The SDS-PAGE results, using a high-density polyacrylamide gel, for seven PSI crystallite preparations are shown in Fig. 4.9. The major bands present for any of the preparations correspond to the D, F, E, C, and K subunits of PSI, as shown to the right of the gel in the figure. Subunits A and B are retained at the top of the separation gel. All bands could be identified as protein subunits of PSI, without detectable levels of protein subunits from PSII, ATP synthase, or rubisco.

An important quality control and assessment measure for the quality of the PSI crystallites is to observe SDS-PAGE gels for bands corresponding to degradation products. The most prominent degradation product is a partially degraded PsaD subunit of Photosystem I, which manifests as a separate band below the normal PsaF band.

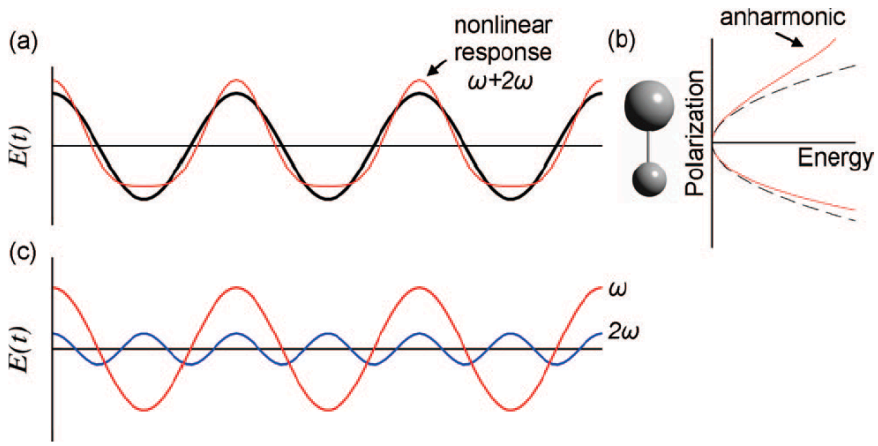

Figure 4.10: **Pictorial Representation of Second Harmonic Generation** Classical anharmonic oscillator model describing second harmonic generation. Linear and nonlinear responses (a) and the anharmonic polarizability (b) for a hypothetical heteronuclear diatomic molecule. The nonlinear response shown in (a) in the time domain is recovered by the summation of ω and 2ω (c). Image and caption taken from (Wampler et al., 2007).

Degradation products of PsaD could not be detected in the gel for crystallites of any PSI preparation. In the case that degradation products of PSI are detected in an SDS-PAGE gel, the crystals produced would be re-crystallized, but no sample containing the degraded protein would be used directly in structure determination experiments.

SONICC results from the Photosystem I crystallites

SONICC was used as a method to determine whether the precipitates formed in the ultrafiltration crystallization experiments were crystalline or amorphous PSI_T precipitate, *i.e.* an unordered state. Characterization by SONICC is essential for all samples that only contain potential nanocrystals ($\leqslant 1\ \mu$m size), as an example, see Fig 4.8.

SONICC utilizes the ability of crystalline samples with chiral space groups to produce the second harmonic of an intense incident laser—second harmonic generation (SHG), as shown in Fig. 4.10, is a nonlinear optical effect—and produces very high signal-to-background for crystalline samples. The measurements were done by inundating the PSI samples with a 1000 nm laser operating at a power between 50 mW and

150 mW. The sample was monitored for SHG in both the forward (trans) and backward (epi) directions, as shown in Fig 3.5.

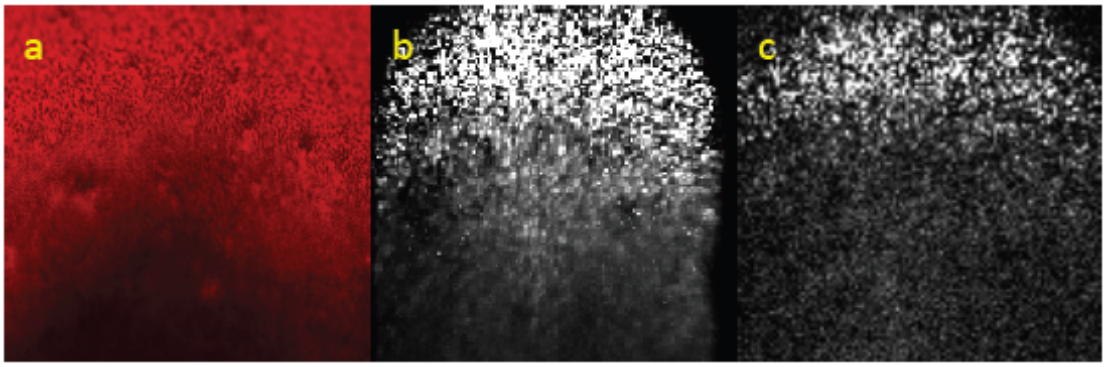

Figure 4.11: **SONICC images of Photosystem I crystallites** SONICC images of PSI crystallites from the 40 min settling step after being filtered using an inline 2 m filter. (a) Shows the brightfield image with the trans and epi SHG signals shown in (b) and (c), respectively. The large signal to background indicates that crystalline sample is likely. The fading of the image at the bottom of each image is due to depth of focus.

The 40-min crystallites were tested using SONICC and were positive for second harmonic generation, as shown in Fig. 4.11. The relative amount of SHG can be used to qualitatively show whether a precipitate is crystalline or amorphous precipitate (or crystals from any of the other protein-replete phases that can be found within protein crystallization screens). The SHG signal in the trans direction was measured with a mean of 3529 counts, with a range of 22-12565 counts, whereas the background was measured with a mean of 33 counts, and a range of 2-180 counts. The SHG signal in the epi direction was measured with a mean of 44 counts, with a range of 0-270 counts, whereas the background of the epi images was measured with a mean of 1.1 counts and a range of 0-4 counts. The fluorescence signal was negligible compared to the SHG signal in both the trans and epi directions. The SHG signal was over 100 times the background level and 44 times the background level in the trans and epi directions, respectively. From the large ratios of signal to background, it is clear that the sample is comprised of crystalline PSI material.

General discussion of ultrafiltration crystallization

The ultrafiltration crystallization method was further developed to produce the PSI crystallites in large amounts for the LCLS experiments based on the use of the method explored in the past as a last purification step for PSI. However, the method was never used to produce diffraction quality crystals. The ultrafiltration method allowed for the crystallization of large quantities of protein, which was vital due to the use of PSI crystallites in subsequent femtosecond nanocrystallography experiments.

Another reason why the ultrafiltration crystallization method was further developed to produce the PSI crystallites for the LCLS experiments *was* the large size distribution of the produced PSI crystals. The size distribution allowed for crystals of different sizes grown in the same ultrafiltration crystallization experiment using (approximately) the same experimental conditions, such as protein preparation, salt, etc. Comparison of the different sizes of PSI crystallites, produced in the same ultrafiltration crystallization experiment, using femtosecond X-ray protein nanocrystallography experiments, as discussed in Section 4.7, would allow a more direct comparison of the scattering strength and internal order of the crystallites by removing other variables such as protein preparation, FPLC separation, etc. Until the serial powder diffraction experiments – Section 4.6 – and the initial LCLS experiments – Section 4.7 – utilizing nanocrystallography, the resolution to which the different crystal sizes would scatter was unknown, and the size distribution allowed a clear examination of many different crystal sizes from the same batch.

4.3 Re-crystallization of Photosystem I from crystal pellets

Motivation

The PSI re-crystallization experiments serve two major goals. The first goal was to produce a nanocrystal and microcrystal sample with a reduced size distribution com-

pared to the large size distribution of the PSI nanocrystals and microcrystals collected in each of the settling time steps from the ultrafiltration-crystallization experiments. The second goal was to develop a method that would allow for the production of fresh PSI nanocrystals and microcrystals on site at the LCLS for femtosecond-nanocrystallography experiments.

In regard to the PSI size distribution produced during the ultrafiltration method, the settling experiments were a coarse method of size selection of the nanocrystals and microcrystals. Inline filters needed to be used to restrict the maximum size of the microcrystals, and there was no method to finely control the smallest size of the PSI crystallites. For femtosecond nanocrystallography experiments, having a narrow (and preferably known) size distribution would potentially reduce the necessary number of patterns for Monte Carlo integration of the intensities (Kirian et al., 2010) but was not a prerequisite for data evaluation (see Section 4.7).

Furthermore, when visually examining some of the larger PSI crystals produced by the ultrafiltration method, it was observed that many of the crystallites were jagged and often appeared to be fragments of larger crystals. The morphology of the crystals could indicate that the removal of the PSI crystals from the membrane was damaging the crystals, or that the low-speed centrifugation steps used during the cleaning of the crystals after settling, was damaging the PSI crystals. These poorly formed larger PSI crystals might not diffract very well to higher resolution. Therefore, the re-crystallization experiments were developed in an attempt to achieve a more uniform size distribution, as well as to develop a method to grow fresh crystals directly at the LCLS, prior to data collection. Furthermore, as discussed later in Section 4.7, well formed nanocrystals show clear shape transforms that could be used to solve the crystallographic phase problem. In addition, a consistent size and morphology of the crystallites would facilitate SAD or MAD phasing of nanocrystallography data.

The ability to generate PSI nanocrystals and microcrystals on site at the LCLS

directly before femtosecond nanocrystallography beamtimes was desired because earlier experiments done using large PSI crystals at conventional synchrotron sources indicated a loss of 0.2 Å of resolution in the diffraction pattern for every day that the large crystals were not frozen at liquid nitrogen temperatures. As a result, for the high-resolution LCLS experiments, a method to re-crystallize the PSI crystals within 2 days of their use in LCLS experiments was necessary to develop.

The lowered solubility of PSI in low $MgSO_4$ mixtures makes the preparation of batch experiments—which explore high protein concentration and low salt conditions in this case—through mixing of two solutions difficult, because there is a minimum ionic strength. Unlike in the case of high-precipitant conditions where higher concentrations of precipitants can be directly added to the solution, this does not work for lowering the ionic strength, as the solutions cannot be made such that there is lower salt concentration than zero. This makes the preparation of the solutions with low final salt concentrations and high protein concentrations impossible through the mixing of two solutions. Consequently, gel-filtration media was used to take a high protein concentration in G_{50} and quickly move the protein into G_{0-5}, as shown in Fig. 4.12.

The PSI re-crystallization was done using spin columns, in which gel-filtration media was used to allow the PSI samples to be desalted from G_{50} to G_{0-5} in 2.5 min.

Re-crystallization experiment overview

The re-crystallization experiments were designed in the manner described in Section 3.2 in order to allow pseudo-batch crystallization experiments of PSI using 0-5 mM $MgSO_4$ to be performed. Briefly, the gel-filtration media was added to a 1-mL syringe that had a cotton plug placed at the outlet to make the spin column. The gel-filtration media was incubated and washed with $2x$ the bed volume with the buffer of desired final low-salt concentration. The spin column was centrifuged in order to collapse the gel-filtration beads, which resulted in the exclusion of the PSI_T from percolating through

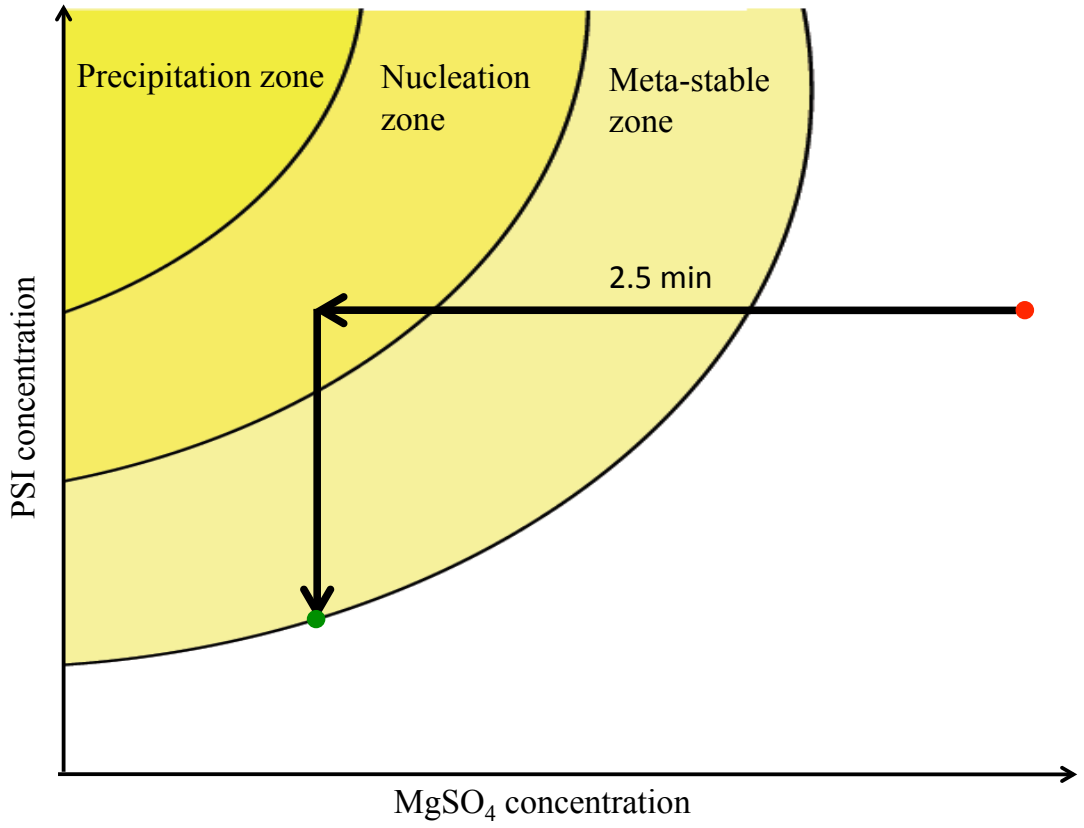

Figure 4.12: **Phase diagram for Photosystem I re-crystallization** Schematic diagram of the phase diagram of PSI showing the crystallization pathway of the pseudo-batch crystallization experiments conducted during the PSI re-crystallization experiments. The red circle indicates the location in the phase diagram where the crystallization experiment is initiated, whereas the green circle indicates the location of the system upon reaching equilibrium during the crystallization experiment.

the beads while still allowing for the small cations and anions of salts to percolate through the beads. The large PSI_T molecules would move around the beads and quickly elute form the column. The quick elution of the PSI_T ensured that a minimal amount of the protein precipitated onto the column material, and hence the technique can be considered pseudo-batch crystallization.

After centrifuging the PSI into the new salt condition during the re-crystallization experiments, the sample was stored at 4°C to allow slow crystal growth. A 5-μL aliquot of the homogenized sample was immediately removed, and images were taken using a light microscope. The samples would then have images taken every 1-6 hr until growth

stopped.

The parameters explored during the re-crystallization experiments of PSI using spin columns include determination of the growth rate of the re-crystallized sample, measurement of the effect of the PSI concentration on the resultant crystal sizes and shapes, and the measurement of the effect of the final salt concentration on the produced nanocrystals and microcrystals. PSI crystals from multiple preparations were used to produce the protein solution used in the re-crystallization experiments.

Results of PSI re-crystallization

Time-step measurements during re-crystallized of PSI

The growth of crystals during the re-crystallization experiments of PSI was monitored every hour for eight hours, in order to determine the growth rate of the re-crystallized PSI crystals. Crystals were re-suspended and a 1-μL aliquot was removed for imaging. The samples were stored at 4°C.

The conditions monitored were two samples with 600 μM chlorophyll being crystallized in G_0 buffer. The images for hours 1 through 8, in one-hour increments, as well as the 24-hour time point, are shown in Fig. 4.13. The images were not taken on a stereo-microscope, and as such, the crystal images are shown as gray-scale images. As can be seen, a slow but steady growth of the PSI nanocrystals and microcrystals was observed over 8 hours. Before 8 hours of growth, the re-crystallized PSI samples had nanocrystals that were $\leqslant 2$ μm in size. Upon being allowed to row for 24 hours, however, 5-μm PSI microcrystals grew. The slow growth made quenching the crystallite growth at the desired crystallite size feasible.

Interestingly, the sampling frequency appeared to have an effect on the growth rate and sizes of the re-crystallized PSI nanocrystals and microcrystals. When the re-crystallization samples were frequently imaged, the resulting crystals generally grew at a faster rate and became larger than crystals from the samples that were not imaged

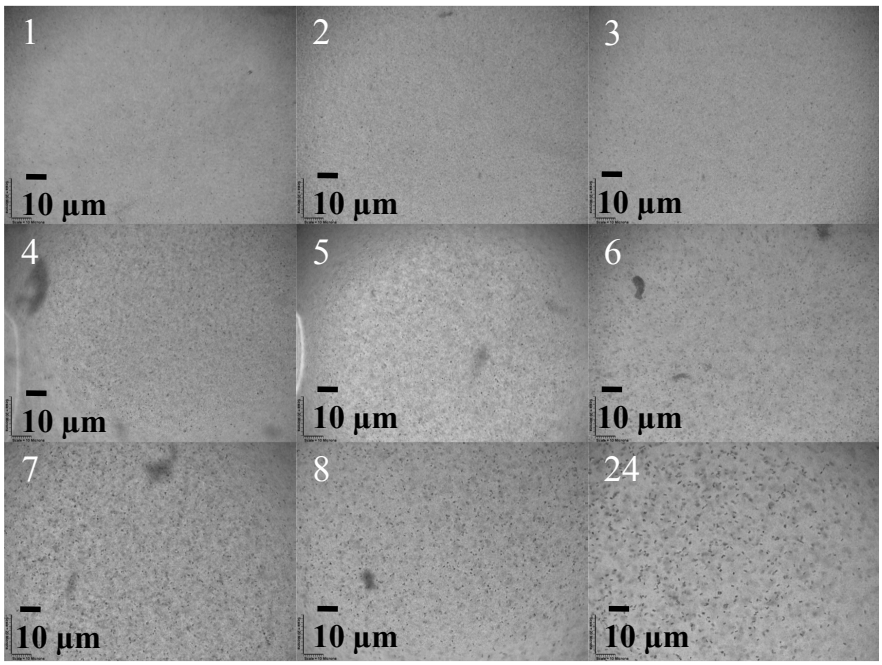

Figure 4.13: **Re-crystallization of Photosystem I time points** The figure maps the path of the crystal growth of PSI crystals in G_0 with time points of 1-24 hours, as indicated on the figure, at 600 μM chlorophyll.

as frequently. A possible explanation is that the frequently imaged samples had the crystals re-suspended in the solution, allowing the faster growth of the nanocrystals, as the crystal growth was facilitated by bringing additional PSI molecules to the crystal surface during re-suspension. In the non-perturbed samples, crystals sediment, thereby allowing only a small amount of soluble PSI access to the crystallite surface, which could possibly decrease the growth rate.

Variation of the Protein Concentration

Upon inspecting the freshly made sample, very tiny microcrystals occasionally became visible in the samples that contained the lower concentration of PSI and higher concentrations, but typically, the solution appeared turbid, with a green hue. The higher the chlorophyll concentration for a given salt concentration, the more turbid the initial sample would appear.

Figure 4.14a-f shows images of crystals produced using G_0 buffer with chlorophyll concentrations of 200, 400, and 600 μM, after 12 hours. As can be seen in the images presented in Fig. 4.14a-f, the number of microcrystals increases and the average crystallite size decreases with increasing chlorophyll concentration for a given incubation time. Examining the images closely, it is evident that a very narrow size distribution is present within the re-crystallized samples and that the morphology is very similar for all of the crystals produced. These results are in sharp contrast to results of the ultrafiltration crystallization experiments. However, upon letting the crystals grow to final size, it became apparent that the crystals, regardless of initial chlorophyll concentration, all finally grew to a fairly similar size of approximately 50 microns in length and 25 microns in width. In order for the crystals to be used in a desired size range, the crystal growth was monitored using a light microscope until the crystals were of sufficient size. Once the desired size was achieved, the supernatant, containing soluble PSI trimer, was removed, and fresh G_0 buffer was added, such that the volume of G_0

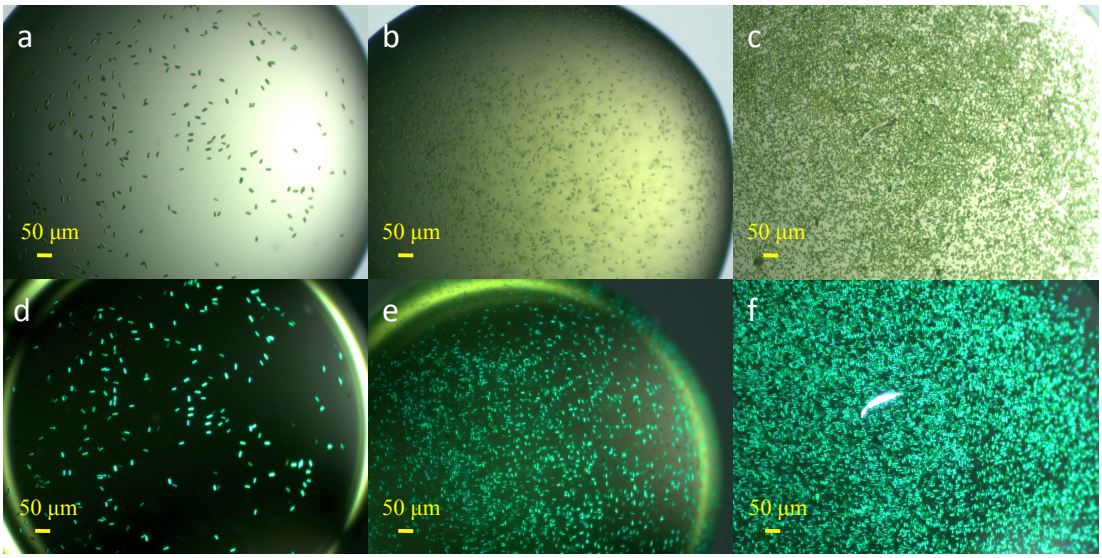

Figure 4.14: **Re-crystallized Photosystem I crystal images** Images of PSI crystallites formed using the re-crystallization procedure and G_0 as the precipitant solution with (A) and (D) being 200 μM chlorophyll, (B) and (E) using 400 μM chlorophyll PSI sample, and (C) and (F) using 600 μM chlorophyll PSI sample. Top row shows brightfield images and bottom row shows birefringent images.

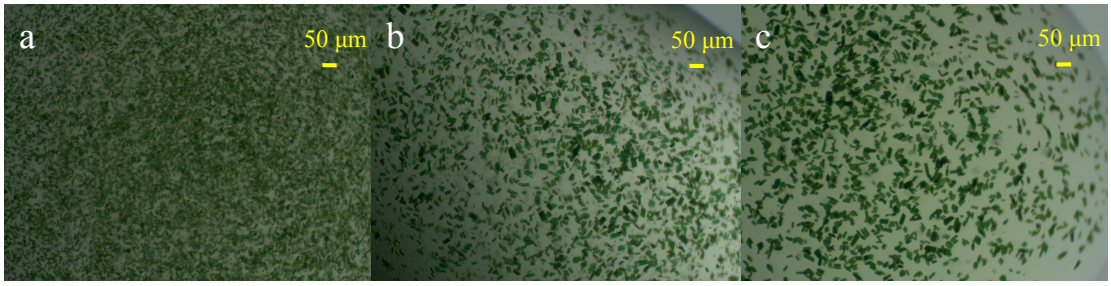

Figure 4.15: **Salt effects on re-crystallized PSI samples** Results of re-crystallization experiments in which PSI sample containing 600 μM chlorophyll was crystallized in buffer containing (a) 0 mM $MgSO_4$, (b) 3 mM $MgSO_4$, and (c) 5 mM $MgSO_4$ after 36 hours. The size of the PSI crystals is shown to be acutely dependent on the final salt condition.

buffer was approximately five times the amount of remaining supernatant. The crystals were monitored for dissolution, but this did not occur. Instead, the observation was that the crystals continued to grow slightly larger, due to PSI_T that remained in solution. The continued growth of the PSI nanocrystals and microcrystals was prevented using a second quenching step by replacing the first supernatant with fresh G_0 buffer.

Variation of the Final Salt Concentration

Figure 4.15 shows the crystals resulting from crystallization experiments in three different salt conditions, after the crystals were allowed to grow for 36 hours. A clear trend emerges that decreasing the salt of the final precipitant solution of the re-crystallization experiments leads to a larger number of smaller (on average) crystals. In some of the higher salt trials, 5-10 micron sized crystals grew within an hour. The PSI crystals produced using the 5 mM $MgSO_4$ had very sharp edges, and appeared to be well ordered, despite growing in such an expedited fashion.

Unfortunately, for the purpose of the high-resolution LCLS experiments in which 2-5 micron crystallites were desired, most of the experiments using G_{3-5} produced crystals with too high of a growth rate, and ultimately crystals that were 20-30 μm in size. The high growth rate was problematic because fine control over the size

was desired, and consequently, crystal growth over many hours was preferred. The size distribution of PSI crystals produced in the 3-5 mM $MgSO_4$ conditions was very narrow, and the produced crystals were strongly birefringent, indicating that the crystals were well formed.

SONICC results of the re-crystallized PSI samples

The re-crystallized PSI samples were tested with SONICC, discussed in Section 3.5, and the results indicated that the sample contained crystals, as shown in Fig. 4.16. The images show the re-crystallized PSI sample with 600 μM chlorophyll in G_0, after six hours of growth. The growth of the crystals was quenched after six hours by washing the crystallites with G_0 until most of the soluble protein was removed, indicated by the supernatant apparently colorless.

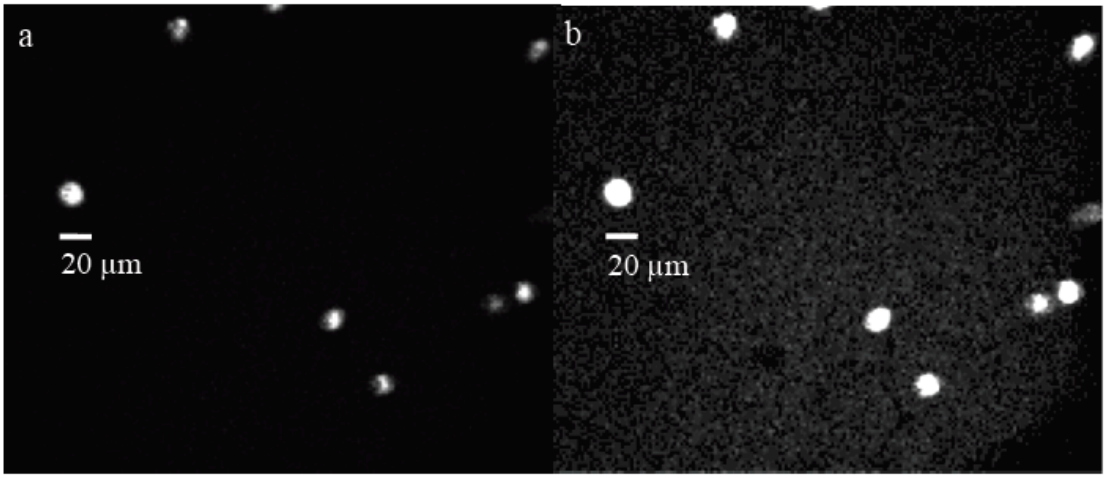

Figure 4.16: **SONICC images of a re-crystallized PSI sample** (a) SONICC image of a 600 μM chlorophyll re-crystallization of PSI using G_0. The scale bar is approximately 20 μm in size, and the image clearly indicates the presence of 5 to 15-μm crystallites of PSI. (b) 6x contrast of the image in (a) showing background counts.

SONICC images displayed in Fig. 4.16a, b are from the same sample with an increased contrast of 6x for (b). The increased contrast is used to highlight the background in the images. The signal-to-background is 429 in Fig 4.16, but the background was larger, and subsequently noisier, than in most SONICC images of PSI. The back-

ground is likely produced by very small protein nanocrystals or amorphous precipitate that appear in the re-crystallization experiments of PSI in G_0, as discussed below.

Challenges of the PSI re-crystallization procedure

Using 0 mM $MgSO_4$ as the final salt concentration for the re-crystallization experiments is a very harsh condition for the growth of PSI nanocrystals and microcrystals. Although the initial experiments using 0 mM $MgSO_4$ as the crystallization condition produced PSI nanocrystals that continued to grow with time, additional experiments, some experiments performed on site at the LCLS labs, presented in Section 4.11, produced small PSI precipitates that did not continue to grow with time. The small PSI precipitates produced in the re-crystallization experiments do not show birefringence, and are not readily detected in the SONICC experiments, as shown in 4.16b. Viewed using a high-magnification visible microscope, the re-crystallization experiments were shown to produce 5-μm PSI crystals as well as smaller precipitates that do not continue to grow. The possibility arises that the G_0 buffer re-crystallization was producing such a high supersaturation that amorphous precipitates were initially formed, followed by nanocrystals, once the supersaturation became low enough to allow controlled nucleation and growth. As a result, the re-crystallization experiments require thoroughly examinion to determine whether the conditions used so far are too harsh to allow reproducible nanocrystal and microcrystal growth.

One interesting note is that whereas the production of amorphous precipitates in the ultrafiltration crystallization method is very challenging due to the way ultrafiltration moves through the phase diagram, the fast movement through the phase diagram by the use of spin columns would make it possible to reach the precipitation zone, thereby leading to amorphous aggregates. Therefore, the re-crystallization experiments should be optimized in higher final salt concentrations, such as 1-2 mM $MgSO_4$. Ideally, a SONICC instrument would be installed at the LCLS in order to directly correlate the

SONICC results to the X-ray free-electron laser results.

The storage time of the PSI crystals, produced from the ultrafiltration crystallization, and used in the re-crystallization experiments had an effect on the final size of the re-crystallized PSI nanocrystals. PSI nanocrystal batches that had been prepared within two weeks of the re-crystallization experiments caused a higher growth rate of the crystals and produced a larger yield and larger number of crystals. The size dependence on the freshness of the PSI can be explained by considering that β-DDM is prone to hydrolyzing, producing maltose and N-dodecanol. The N-dodecanol is hydrophobic and may disrupt the lamellar phase of the detergent in the PSI crystals that stabilizes the crystal in the **a**-**b** plane.

Summary of the Re-crystallization Procedure for PSI

The PSI re-crystallization procedure was established in order to grow freshly-made PSI nanocrystals and microcrystals that could be made within two days. A narrow size distribution was highly desired for future work involving phasing of femtosecond X-ray protein nanocrystallography data.

The re-crystallization procedure allowed for the production of nanocrystals of PSI for a variety of conditions. The PSI crystal growth could be monitored and quenched as necessary, and any salt condition could be used for crystallization experiments. Samples were made onsite at the LCLS for femtosecond X-ray protein nanocrystallography experiments.

4.4 PSI-Fd preparation

Motivation for the PSI-Fd Preparation

Crystallization experiments involving the Photosystem I-Ferredoxin complex were done in order to establish conditions that could produce nanocrystals and microcrystals of the PSI-Fd complex for use in time-resolved X-ray structure experiments at the LCLS, as

discussed in Section 4.8. Although crystallization conditions had previously been established for the growth of large crystals of the PSI-Fd complex (Fromme et al., 2003), the phase diagram was previously never systematically explored for the production of PSI-Fd nanocrystals and microcrystals.

The phase diagram was initially explored through the use of hanging-drop vapor-diffusion experiments by varying the PEG400 concentration and total protein concentration, and a schematic of the phase behavior with representative crystallization experiments is shown in Fig. 4.17.

The PSI:Fd ratio was constant at 1:1.1 throughout the experiments. After the initial exploration of the PSI-Fd complex concentration as a function of PEG400 concentration, batch crystallization experiments, as shown schematically in Fig. 4.17, were established that mimicked the final conditions of successful vapor-diffusion experiments. The batch crystallization experiments were done with glass capillaries and eppendorf tubes as crystallization vessels, which allowed for the use of larger volumes of crystallization experiments needed for LCLS work.

PSI-Fd preparation results

Photosystem I-ferredoxin vapor-diffusion crystallization experiments

The PSI-Fd co-crystallites were made by first examining the phase diagram for crystallization of the PSI-Fd complex utilizing vapor-diffusion experiments. All incubation steps and crystallization experiments were performed in a green room, to prevent dissociation of the Fd from the PSI trimer. Table 4.3 shows the results from one of the initial vapor-diffusion crystallization experiments done in a 24-well plate using 0.5-μL sample volumes. In Table 4.3, two rows of the 24-well plate were devoted to each P_{700} and PEG400 concentration. The duplication of vapor-diffusion crystallization experiments was done to verify results.

Figure 4.18 shows the results of a select set of different crystallization con-

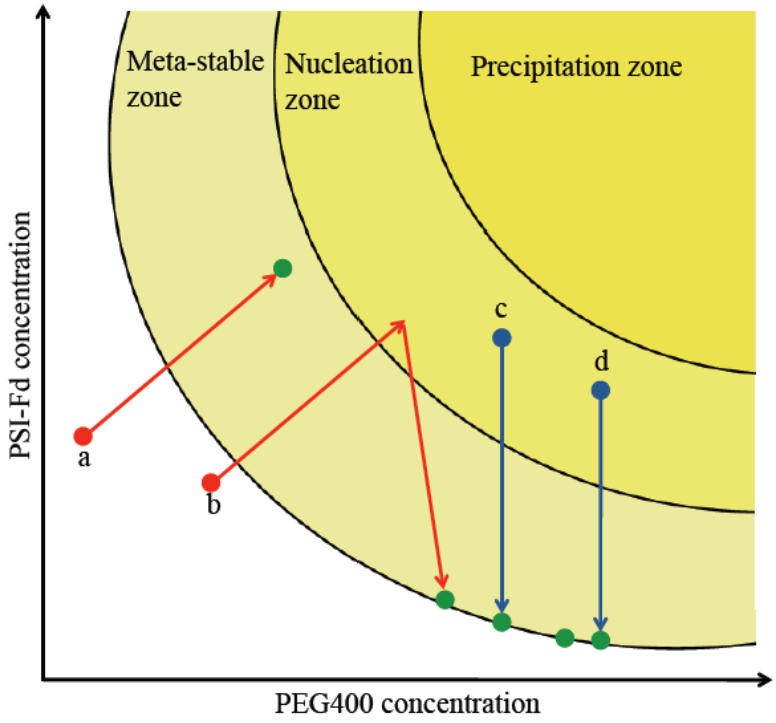

Figure 4.17: **Schematic of the phase behavior of the PSI-Fd complex as a function of PEG400 concentration** Schematic of the PSI-Fd with PEG400 phase diagram, showing the two types of crystallization experiments done for the PSI-Fd system, in which the red arrows represent vapor-diffusion experiments and the blue arrows represent the batch crystallization experiments. The red circles indicate the starting position of a vapor-diffusion experiment, the blue circles represent the starting point for a batch crystallization experiment, and the green circles indicate the point at which the system returns to thermodynamic equilibrium during the crystallization. (a) shows a vapor-diffusion experiment that does not reach the nucleation zone, and as a result, the system is left in a meta-stable zone, in which the thermodynamic driving force is for crystallization to occur, but a kinetic barrier prevents the formation of the critical nuclei. (b) shows another vapor-diffusion experiment that would produce crystals. (c) and (d) represent batch crystallization experiments that produce crystals.

	1	2	3	4	5	6
A; P_{700} = 90 μM	PEG400 = 14.0%	PEG400 = 14.5%	PEG400 = 15.0%	PEG400 = 15.5%	PEG400 = 16.0%	PEG400 = 16.5%
B; P_{700} = 90 μM	PEG400 = 14.0%	PEG400 = 14.5%	PEG400 = 15.0%	PEG400 = 15.5%	PEG400 = 16.0%	PEG400 = 16.5%
C; P_{700} = 80 μM	PEG400 = 14.5%	PEG400 = 15.0%	PEG400 = 15.5%	PEG400 = 16.0%	PEG400 = 16.5%	PEG400 = 17.0%
D; P_{700} = 80 μM	PEG400 = 14.5%	PEG400 = 15.0%	PEG400 = 15.5%	PEG400 = 16.0%	PEG400 = 16.5%	PEG400 = 17.0%
A; P_{700} = 90 μM	One large crystal	A few large crystals	A few large crystals	A few large crystals	$\geqslant$ 10 crystals (15-25 μm)	A few large crystals%
B; P_{700} = 90 μM	One large crystal	$\geqslant$ 10 crystals (15-25 μm)	A few large crystals%	$\geqslant$ 10 crystals (15-25 μm)	A few large crystals	$\geqslant$ 10 crystals (15-25 μm)
C; P_{700} = 80 μM	A few large crystals	One large crystal	A few large crystals	$\geqslant$ 100 crystals ($\leqslant$ 15 μm)	$\geqslant$ 10 crystals	$\geqslant$ 10 crystals (15-25 μm)
D; P_{700} = 80 μM	A few large crystals	A few large crystals	A few large crystals	$\geqslant$ 10 crystals (15-25 μm)	$\geqslant$ 10 crystals (15-25 μm)	$\geqslant$ 100 crystals ($\leqslant$ 15 μm)

Table 4.3: **Example of PSI-Fd vapor-diffusion crystallization experiment.** Example of one of the vapor-diffusion experiments on the PSI-Fd complex. The protein concentration was varied across the rows, whereas the PEG400 concentration was varied down the columns. The top four rows indicate the PEG400 in each well, whereas the bottom four rows indicate the resultant crystallite sizes for the crystallization experiments shown in the top four rows. The PSI:Fd ratio was kept fixed at 1:1.1, in accordance with (Fromme et al., 2002). Large crystals generally mean crystals over 25 μm in size.

ditions using the hanging-drop crystallization technique. The major trend that was observed in the PSI-Fd co-crystallization experiments was that smaller co-crystals produced in larger numbers were observed as the PSI concentration was reduced and as the PEG400 concentration in the reservoir was increased.

The conditions using the highest PSI concentrations and lowest PEG400 concentrations often produced one to a few large co-crystals in the screen. The large co-crystals typically showed very sharp edges and formed within two days at 20-22°C. Ideally, one of the large co-crystallites would have been examined using an in-house X-ray source, in order to determine the unit cell parameters, but the in-house X-ray equipment at ASU was not operating during the time of the PSI-Fd co-crystallization experiments.

Seeding experiments were performed by crushing the large co-crystals formed in the high PSI-Fd concentration trials using a hair, with the hair subsequently dragged through a crystallization drop containing lower concentrations of PSI. The seeded experiment results were compared to non-seeded results for a variety of PSI concentrations and PEG400 concentrations. In general, it was noted that crystallization trials that were seeded produced more numerous, but smaller, co-crystals.

Determining the effects of HEPES and $CaCl_2$

Generally, increasing the $CaCl_2$ concentration was found to have a similar, but less dramatic, effect as increasing the PEG400 concentration. There was no effect from varying the HEPES pH 7.5 concentration between 100 mM and 200 mM. Ultimately, the stock solution of 1.0 M $CaCl_2$ plus 1.0 M HEPES pH 7.5 was used to produce the batch crystallization mixtures to allow for larger volumes of 70% PEG400 to be added to the mixtures. Lipid additives were previously known to increase the quality of PSI-Fd crystals. Due to time constraints, however, crystallization trials to determine the effects of lipids were limited, and in addition, limited quantities of ferredoxin made a

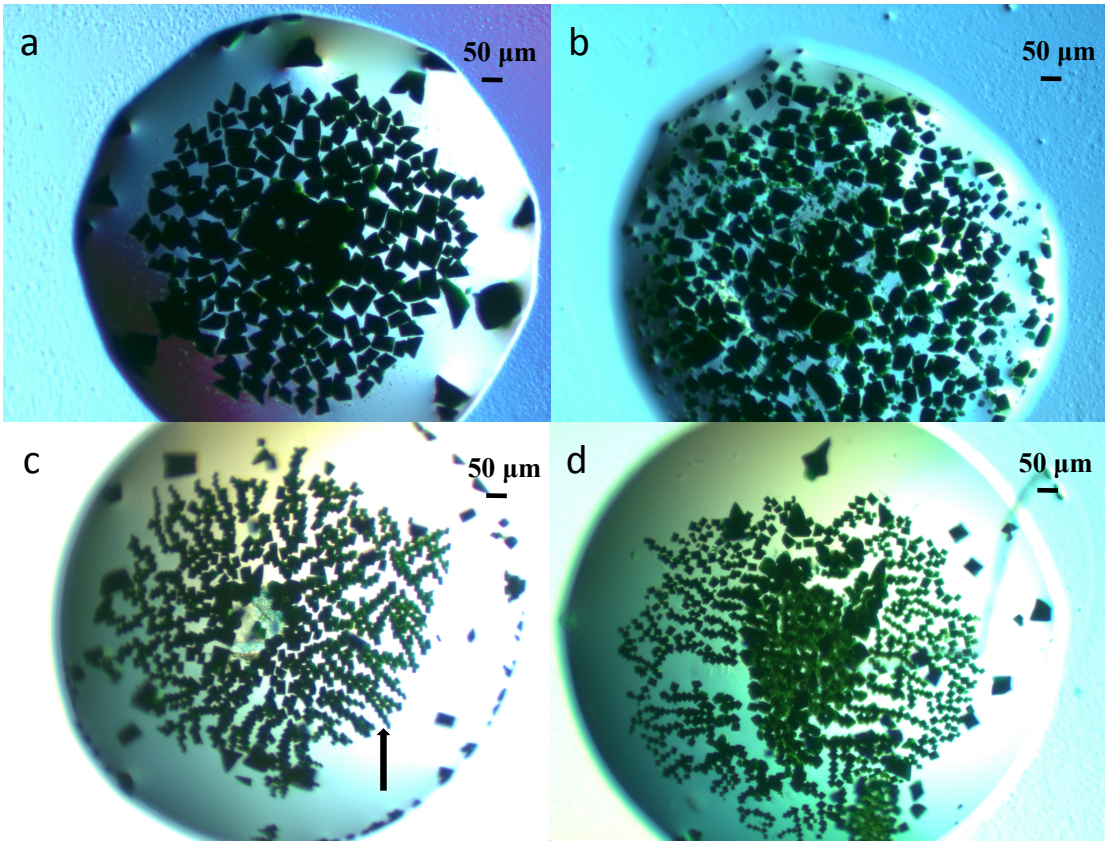

Figure 4.18: **Photosystem I-ferredoxin co-crystal images.** The figure shows a representative selection of results featuring a decrease in crystal size and an increase in crystal number density as the PSI concentration is lowered and the PEG400 concentration is increased. Top row shows results from a crystallization screen where the PSI concentration is decreased and the PEG400 concentration is increased from 90 μM P_{700} and 15.5% PEG400 (a) to 80 μM P_{700} and 16% PEG400 (b). Bottom row, shows results from a crystallization screen where the PSI concentration is decreased and the PEG400 concentration is increased from 10 μM P_{700} and 18% PEG400 (c) to 5 μM P_{700} and 18.5% PEG400 (d). The arrow in (c) shows a series of PSI-Fd crystals that are attached to one another.

thorough exploration of all of parameter space impractical.

Crystallization experiments of the PSI-Fd complex were also conducted using PEG2000 as a precipitant. The experiments showed that the PSI-Fd complex can be precipitated when PEG2000 is used as a replacement for PEG400, with all other solutes at commensurate values to the PEG400 screens. However, due to time constraints, the precipitates were not tested for crystallinity using the SONICC technique or X-ray diffraction. If enough ferredoxin is available, a thorough comparison of many different PEG sizes should be performed, in order to optimize the co-crystallization with respect to PEG chain length. Additionally, the effects of the differing HEPES concentration on the quality of the crystals of the PSI-Fd complex should be revisited in order to ascertain whether the invariability of the result on HEPES pH 7.5 concentration in the vapor-diffusion experiments extends to the batch crystallization experiments.

Photosystem I-Ferredoxin batch crystallization experiments

PSI-Fd batch crystallization experiments were done to grow nanocrystals of the PSI-Fd complex using crystallization conditions similar to conditions that produced showers of microcrystals in the previously done vapor-diffusion experiments. The batch crystallization experiments were conducted, as they opened the possibility for scaling the experiments to larger volumes. The batch crystallization experiments were established as the primary method of producing a bulk quantity of the PSI-Fd co-crystallites for the LCLS experiments.

The PSI-Fd batch crystallization experiments were started by mixing the protein crystallization solution, as described in Section 3.3, with concentrated solutions of PEG400, $CaCl_2$, and HEPES pH 7.5 buffer, such that the final solute and protein concentration was similar to those encountered in the low-protein, high-PEG400, vapor-diffusion experiments (upon full equilibration). In the optimized conditions for bulk co-crystallite production, the final P_{700} concentration was 35 μM, the $CaCl_2$ and HEPES

pH 7.5 concentrations were both set to 150 mM, and the β-DDM concentration was set to 0.013% (m/v).

The batch crystallization experiments were never able to completely reproduce the observed PSI-Fd microcrystal size and morphology from the vapor-diffusion experiments, but this is to be expected; the two crystallization experiments approach the supersaturation zones differently, as shown using a schematic of the PSI-Fd with PEG400 phase diagram in Fig. 4.17. Additionally, the quality and size of the produced microcrystals of the PSI-Fd complex were dependent on the storage time of the PSI sample batches used in the preparation. The "fresher" the PSI preparation was, the larger the resultant crystals grew in the crystallization experiments.

The conditions optimized to produce PSI-Fd co-crystallites that were smaller than 5 μm in size and were positive for second-harmonic generation were established as 35 μM P_{700}, 150 mM $CaCl_2$, 150 mM HEPES pH 7.5, and 0.013% β-DDM, with PEG400 concentrations ranging from 16%-25%. SONICC results are shown in Fig. 4.19. Batch experiments were set up in order to achieve final solution conditions with the above precipitant concentrations; ultimately, 22.5% PEG400 was chosen for the subsequent LCLS experiments. Fig. 4.19b shows the SONICC results from a batch experiment, and although the image does not have much signal, the maximum signal for the entire image is 7492 counts, with a mean count of 5 counts. Such a high count without any evidence of broadband fluorescence is indicative of the presence of crystallites, just at a very low density.

Other difficulties that arose when using the batch crystallization technique for PSI-Fd were that the co-crystallites appeared to have duller edges in comparison to the co-crystallites produced in vapor-diffusion experiments; additionally, up to 10-20% of the crystals could not be removed from the walls of the glass capillary. The co-crystallites that were stuck to the glass may have been stuck due to capillary forces only, but this is unlikely because the glass capillaries had much larger inner diame-

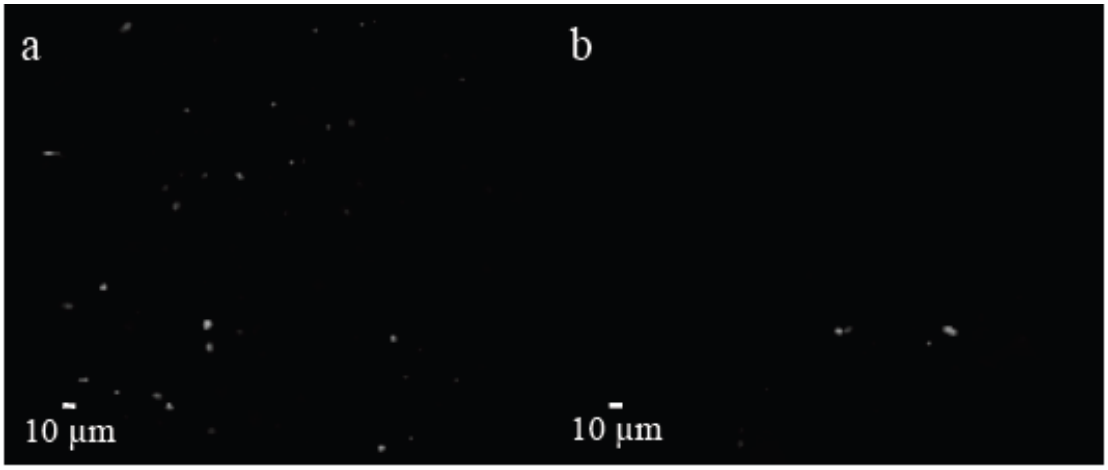

Figure 4.19: **SONICC results for PSI-Fd co-crystallites** SONICC images from two different PSI-Fd crystallization trials. (a) vapor-diffusion experiment with final conditions: 100 mM HEPES pH 7.5, 150 mM $CaCl_2$, 0.013% (m/v) β-DDM, 19.0% (v/v) PEG400, 5 μM P_{700}, and 5.6 μM Fd. (b) batch crystallization experiments with final conditions: 150 mM $CaCl_2$, 150 mM HEPES pH 7.5, 0.013% β-DDM, 165% PEG400, 35 μM P_{700}, and 38.9 μM Fd.

ters than the crystal sizes grown in the crystallization experiment. However, a different explanation may provide insight into the microscopic cause of the deviation in the behavior of the vapor-diffusion and batch crystallization experiments. The poorer-quality co-crystallites and glass-stuck co-crystallites may be connected, as the glass walls of the capillaries could have allowed for heterogeneous nucleation, causing some co-crystallites to be physically stuck to the glass. Removal of the microcrystals of the PSI-Fd complex by force might lead to defects in the crystal. However, heterogeneous nucleation is not a universally bad event, as the heterogeneous nuclei may allow crystal growth at lower supersaturations, possibly leading to more well-order crystals.

Whether heterogeneous nucleation on the glass walls is actually occurring could be tested by thorough comparison of the quality of the co-crystallites grown in batch using different containment vessel types with varying surface-area-to-volume ratios, which is the reason that the batch experiments were done in eppendorf tubes, as well. It is unlikely that both the glass and plastic would allow heterogeneous nucleation in the same manner, and possibly some insight could be gained from the differences. In ad-

dition, the volume of the eppendorf tube could be changed, allowing different surface-area-to-volume ratios to be compared. The heterogeneous nucleation would occur at the surface, therefore more homogeneous nucleation products would be expected for the lower surface-area-to-volume vessels, assuming the homogeneous nucleation pathway has a similar rate, which is an assumption with no experimental evidence in support of, or against, it. However, as described in Section 4.8, the co-crystallites grown in batch using capillaries, as well as the samples prepared at Arizona State University in eppendorf tubes, were all destroyed before the LCLS experiments utilizing the sample were completed; therefore no comparison was possible. However, comparing the co-crystallites grown by batch crystallization in different vessels should be pursued before higher-energy pump-probe experiments on PSI-Fd are conducted.

Summary of PSI-Fd crystallization experiments

The experiments involving the crystallization of the PSI-Fd complex were conducted in order to produce microcrystals of the PSI-Fd complex for time-resolved X-ray crystallography experiments at the LCLS, as detailed in Section 4.8. Vapor-diffusion crystallization experiments were performed to initially explore the phase diagram of the PSI-Fd complex in varying PEG400 concentrations, and subsequent batch crystallization experiments were utilized to produce similar conditions as the vapor-diffusion experiments, but allowed the crystallization to be scaled to larger quantities and volumes.

The vapor-diffusion crystallization experiments and batch crystallization experiments produced nanocrystal samples of the PSI-Fd complex. Although the initial crystallization conditions were known (Fromme et al., 2002), six vapor-diffusion crystallization trays were needed to converge on a region in the phase diagram that was deemed suitable for batch crystallization experiments. The criteria used to determine suitability were crystal sizes between 1 and 5 μm as well as the reproducibility of the crystallization experiments. SONICC results indicated the crystallinity of nanocrystal

and microcrystal samples produced under a variety of conditions, showing the robustness of crystallization experiments in the production of nanocrystals and microcrystals.

Crystallization of protein complexes is notoriously difficult. Surprisingly, the crystallization experiments with the goal of producing nanocrystals of the PSI-Fd complex proved relatively straight forward, and rational crystallization was shown to be very effective. This result bodes well for the generalization of the growth of nanocrystals and microcrystals. When combined with the success of the nanocrystallization of PSI, itself a membrane protein complex, nanocrystallization experiments successfully produced nanocrystals of a membrane protein complex *and* a membrane protein complex that had an additional non-covalently bound protein. As a result, crystallization experiments aimed at the growth of protein nanocrystals and microcrystals may truly be easier in principle and in practice, which combined with the results of Section 4.7, may indicate a new pathway for structural biology that involves optimizing crystallization *in order to* produce nanocrystals and microcrystals.

4.5 Preparation of Photosystem II for crystallization

Motivation for the PSII work

Experiments were carried out in order to produce microcrystals of photosystem II for use in femtosecond nanocrystallography, as described in Section 3.7. The goal was to use PSII microcrystals in a structure determination experiment that avoids the X-ray-induced radiation damage that causes the disassembly of the Mn_4CaCl cluster. The purpose of the PSII work was to extract and purify PSII with the final, future goal to grow microcrystals to be used in femtosecond nanocrystallography experiments to solve the structure of the undamaged Mn_4CaCl cluster in its different S states, S_{1-4}.

Experimental overview of the PSII preparation

In order to grow microcrystals of PSII, the PSII was isolated from *T. elongatus* by thylakoid isolation, solubilization of the protein, weak AEC purification of the PSII from the other solubilized membrane proteins, and final purification by precipitation/crystallization. Results will be presented of the work from the AEC steps through crystallization; for the details of the full preparation, see Section 3.4.

The initial production of PSII microcrystals suitable for LCLS studies was performed by Dr. Ingo Grotjohann of Arizona State University. Separate preparations were done to verify the results.

Overview of the PSII crystallization process

As in the case for the PSI microcrystals, the quality of the PSII microcrystals depended on multiple factors. The wavelength of maximum visible absorbance (in the range of 672 - 681 nm) for the cells will have a major effect, as higher quality PSII is extracted from cells with absorbance maxima at 681 nm. Although counterintuitive at first glance, the conditions that produce higher ratios of PSII:PSI are less desirable even in the case of the PSII preparation. The reason is that the shift in absorbance maximum (between 672 and 681 nm) to 673-674 nm indicates the cells are in a stressed state. In the stressed state, the cells degrade the PSI_T, increasing the ratio of PSII:PSI, but not through the production of more PSII. Consequently, although more PSII is present with respect to PSI, the ideal absorbance maximum for cells used in a PSII preparation was 681 nm.

The amount of light that the protein is exposed to during the PSII purification will have a direct impact on the protein quality (Aro et al., 1993); therefore, the light exposure should be minimized and the light intensity should never be more than the dimmest of green light, as this can provoke photoinduced damage of the protein, which

would be a deleterious effect for the high-resolution structure determination experiments.

Anion-exchange chromatography purification of PSII

AEC was used to purify the PSII from the other proteins in the detergent extract in order to obtain high-purity protein for crystallization experiments. The protein was eluted through the use of a step gradient in $MgSO_4$.

A chromatogram showing the separation of the proteins during the AEC is shown in Fig. 4.20. The PSII dimer eluted in the second band, whereas the PSI and PSII monomers eluted in the first band, and the PSI trimer eluted in the third band. The $MgSO_4$ concentration and the step gradient are shown in the green line of Fig. 4.20. The fractions corresponding to the second peak of the AEC purification were pooled and concentrated using ultrafiltration, as described in Section 3.4.

Final purification through precipitation/crystallization experiments

The concentrated PSII sample was set to a chlorophyll concentration of 0.75 mM by dilution with buffer C. Subsequently, PSII was further purified by three precipitation/crystallization experiments.

The first precipitate was produced by the dropwise addition of a 7.5% (m/v) PEG2000-containing solution ($D_{15.0}$, where D_x contains x % (m/v) PEG2000) to a PSII mixture at 0.75 mM chlorophyll concentration. After incubation for an hour on ice, the supernatant of the sample was removed, and the precipitate was redissolved using buffer C until the chlorophyll concentration was 0.75 mM. The second and third precipitates were formed analogously to the first precipitate, with the difference being that $D_{13.0}$ was used in the second precipitation step and $D_{11.0}$ was used in the third precipitation step, in place of $D_{15.0}$. After the dropwise addition of $D_{11.0}$, the sample was allowed to incubate overnight, and an image of the 2-μm-filtered nanocrystals and

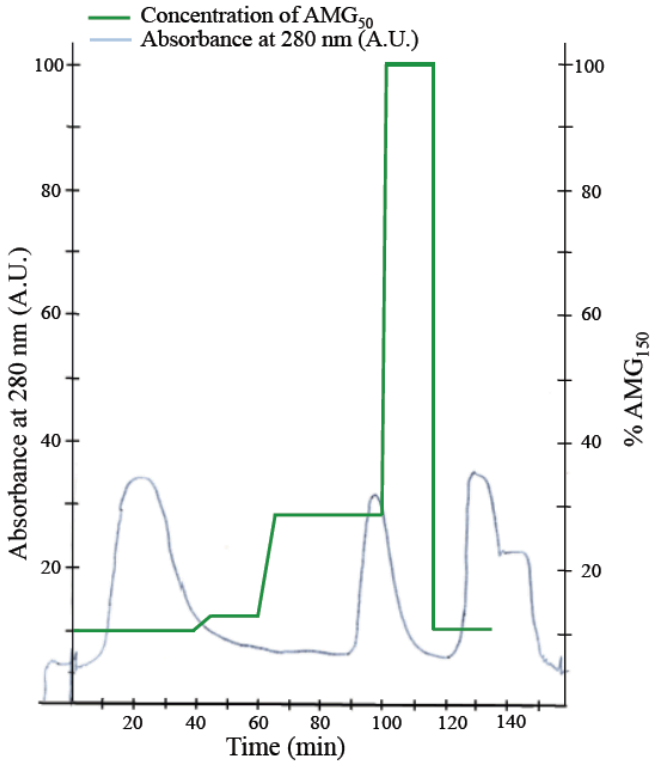

Figure 4.20: **Chromatogram of FPLC purification run for PSII** Chromatogram showing the separation of the PSII sample during an FPLC purification step, using anion-exchange chromatography, in which the 1st peak corresponds to physcobiliproteins, most notably phycocyanin, the 2nd peak corresponds to dimers of PSII, and the 3rd peak corresponds to the PSI monomer and trimer peak.

microcrystals produced by the $D_{13.0}$ crystallization is shown in Fig. 4.21.

The observation was that upon addition of the buffer D_x to the sample, where x = 11.0-15.0 % PEG2000, a turbid drop would appear and fall to the bottom of the container, containing precipitates that were verified to be crystalline in the case of buffer $D_{11.0}$, using both SONICC—see Fig. 4.22b—as well as X-ray diffraction, and $D_{13.0}$, using SONICC. The dropwise addition of buffer $D_{15.0}$ produced the harshest conditions, but the resultant precipitant was not checked for crystallinity by SONICC or X-ray diffraction. Very likely the $D_{15.0}$ produced the smallest PSII crystallites.

The dropwise addition of the D buffers technique to crystallize the PSII represents a classical batch crystallization experiment. Immediately upon addition of the

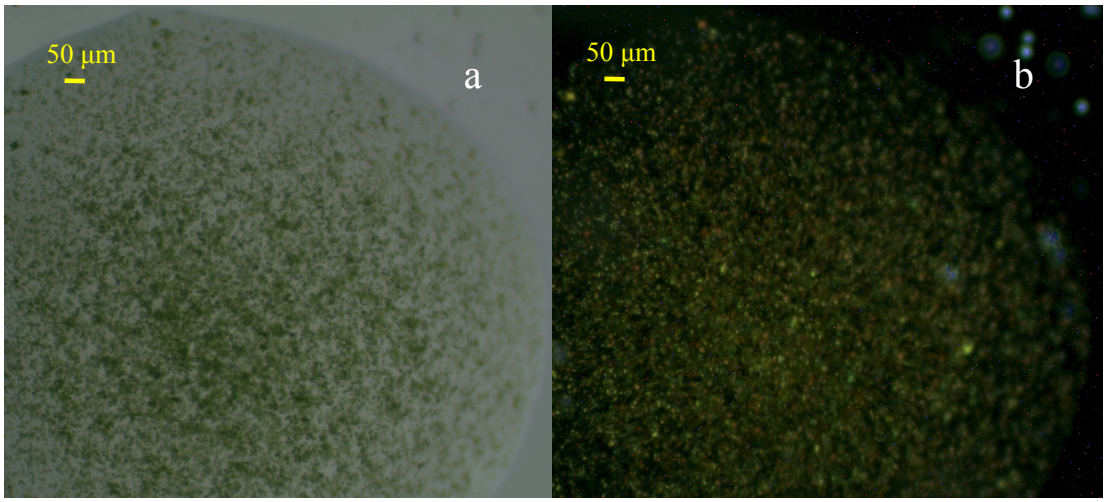

Figure 4.21: **Crystal images of Photosystem II crystallites** Images of Photosystem II grown by the dropwise addition of buffer $D_{11.0}$ to a Photosystem II mixture set to approximately 0.75 mM chlorophyll. (a) Photosystem II crystals viewed in normal light and (b) viewed with crossed polarizers, revealing that the sample is birefringent.

buffer solutions with high PEG2000 concentrations to the solution, the local concentration of PSII and PEG2000 is sufficient to reach the nucleation/precipitation zone, causing the sample to precipitate. The precipitated sample immediately sank to the bottom of the vessel, indicating that the sample was of much higher density than the surrounding solution. The nanocrystals and microcrystals would likely grow slowly at the bottom from any nuclei that sank to the bottom. However, the size of precipitates created by the immediate precipitation is unknown. Additionally, whether the sample is truly crystalline in the immediate aftermath of the dropwise additions of the D buffers, or whether a crystalline state grows from an amorphous state (the amorphous state is higher in density than solution and would likewise sink to the bottom of the vessel). One possibility to infer the pathway of the crystallization is to immediately monitor the precipitate for crystallinity using SONICC, although the timescale that could be monitored with SONICC is only seconds or minutes after crystallization. However, a SONICC instrument at the LCLS would be the ideal case.

The rate of PSII nanocrystal growth by the dropwise addition of the D buffers

may offer important insights into the production of nanocrystals. As mentioned, the procedure utilized to make PSII nanocrystals produces a harsh local environment from which precipitation of the protein occurs. Similar procedures are used in the isolation of other proteins from protein mixtures, such as with $(NH_4)_2SO_4$ precipitation steps. More studies are needed to ascertain whether the initial precipitation conditions for purifying proteins actually induce the growth of nanocrystals for certain conditions. Therefore, the possibility exists that additional studies may allow the generalization of the production of crystallites of a given protein using precipitation steps during a purification procedure. The development of a generalized nanocrystallization protocol, in which precipitation steps normally used for purification are also used for crystallization, would be quite powerful in combination with X-ray crystallography, and therefore should be pursued.

PSII nanocrystal characterization

Fig. 4.22 shows a comparison of the results of UV-microscopy as well as SONICC on PSII crystallites produced in the same crystallization experiment, using $D_{11.0}$, as the crystals shown in Fig. 4.21. The signal-to-background ratio in the UV-microscopy image is 1.6, whereas the signal-to-background ratio in the SONICC image is 1435. The UV-microscopy data clearly show the larger objects are not salt crystals and may contain protein. The SONICC results clearly indicate the crystallinity of the PSII sample produced by the drop-wise addition of $D_{11.0}$ to a 0.75 mM chlorophyll concentration PSII solution. Moreover, the SONICC technique also was shown to be a non-destructive imaging method for PSII crystallites.

UV microscopy can in principle be used to detect protein crystals, or to distinguish precipitated protein (crystalline or amorphous) from soluble protein, due to the higher density of the crystalline/precipitated state, but only if the protein contains a sufficient number of tryptophan residues to allow signal to be measured above the noise.

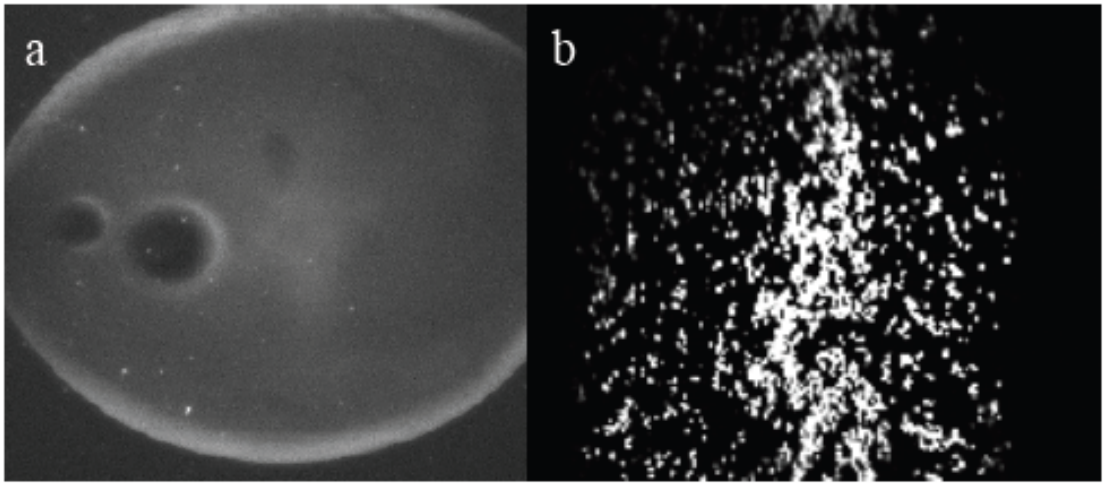

Figure 4.22: **UV microscopy and SONICC results for PSII crystallites** Comparison of imaging techniques used to verify the crystallinity of the Photosystem II nanocrystals. (a) UV-microscopy image and (b) SONICC image of Photosystem II crystals produced by the drop-wise addition of $D_{11.0}$ to a 0.75 mM chlorophyll Photosystem II solution. The sample in (a) is diluted 100-fold with respect to (b).

However, the major purpose of UV-microscopy was to distinguish between protein crystals and detergent or salt crystals. UV microscopy does not allow for differentiation between the various protein-rich phases such as amorphous aggregates, crystals, and oils. UV microscopy data could be used to differentiate protein crystals from salt and detergent crystals, which could prove very useful for quick analysis of nanocrystallization and microcrystallization experiments of non-colored proteins.

4.6 Serial powder diffraction experiments

Motivation

Microcrystals are encountered in membrane protein crystallization screens more frequently than large membrane protein crystals. Despite this relatively higher frequency of occurrence, the microcrystals of membrane proteins and membrane protein complexes were not thoroughly studied using X-ray diffraction methods prior to this. For instance, it was not clear whether there were internal order issues with membrane protein microcrystals.

One hypothesis of the femtosecond nanocrystallography project was that since showers of microcrystals are often produced in crystallization screens, nanocrystals may also be produced, possibly with greater frequency. Additionally, the idea existed that microcrystals and nanocrystals, if nanocrystals of membrane protein even exist, may be more perfect than the macrocrystals, due to a decrease in long-range disorder, leading to lower mosaicity. Therefore, one major aim of the serial nanocrystallography experiment was to test whether membrane-protein crystals smaller than 1 μm /side exist and how well these nanocrystals may diffract.

Serial powder diffraction overview

The serial powder diffraction experiment was used to establish the existence of membrane-protein crystals that are only a few hundred nanometers/side (Hunter et al., 2011). In addition, the scattering power of these small crystals was measured in order to characterize the membrane protein nanocrystals. The experiments were also aimed to provide a test for the liquid injector and prove its suitability for the injection of fully-hydrated protein nanocrystals and microcrystals into an X-ray beam. The experiments were done to test the membrane protein nanocrystals and liquid injector as pre-characterization in anticipation of the femtosecond nanocrystallography experiments using X-ray free-electron lasers, as detailed in Section 4.7.

The serial powder diffraction experiment were done by using a liquid injector—developed at Arizona State University by Dan DePonte, Uwe Weiestall, and R. Bruce Doak—to introduce a sample of fully-hydrated PSI microcrystals to the 520 or 1560 eV X-rays of beamline 9.0.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Diffraction patterns were recorded with both an on-axis and off-axis detector, with maximum resolutions in the corners of the detector for the 1560 eV (520 eV) X-rays being 28 Å (84 Å) and 14 Å (42 Å), respectively.

The PSI samples were prepared by crystallization of the protein in G_6 and G_8

buffers. The samples were stored at 4°C until immediately before use and were both pre-filtered and inline filtered with 100-nm, 220-nm, 500-nm, 2-μm, or 10μm pore sizes. For the 1560 eV experiments with an off-axis detector, as many 30-s exposures as possible were taken and averaged together to produce the powder diffraction pattern.

Serial powder diffraction results

As a comparison to nano- and microcrystals, large single PSI crystals of dimensions 0.5 mm *x* 0.5 mm *x* 2 mm, which have been used to obtain X-ray diffraction to 2.5 Å resolution (Jordan et al., 2001), have a volume of $5\text{x}10^{17}$ nm^3, equaling $4\text{x}10^{13}$ total unit cells within the crystal containing $8\text{x}10^{13}$ PSI trimers (within a beam focus of 0.3 mm *x* 0.3 mm and depth of 0.5 mm yielding $9\text{x}10^{12}$ unit cells in the beam). These single-crystal patterns have been collected using 1 Å wavelength at 3rd generation synchrotrons, and a single-crystal pattern is shown in Fig. 1.7 for reference.

In order to test whether sub-micron PSI crystallites exist, inline filters were used to restrict the maximum size allowed to proceed into the interaction region. After filtering a PSI-crystal suspension with a 500 nm filter and placing the filtrate under an optical microscope, the solution appeared turbid, but no crystals were visible. A completely illuminated (500 nm)^3 crystal would contain 8 *x* 8 *x* 30 (*x* 3) unit cells (the *x* 3 term arises from the hexagonal crystal shape with respect to the PSI unit cell) corresponding to 5800 unit cells containing only 11600 PSI trimers within the crystal, allowing for 5% variation in the filter-pore size, see calculation in Section 3.7. Fig. 4.23 shows the powder diffraction pattern obtained from the $\leqslant$(500 nm)^3 PSI crystals using 1560 eV X-rays. The crystals diffract to a maximum resolution of 2.8 nm at the corner of the detector. In these preliminary experiments, diffraction was observed to the edges of the detector, so that the maximum resolution was instrument limited. The ratio of the number of scattering unit cells in a 0.5 mm *x* 0.5 mm*x* 2 mm single crystal to the number of unit cells that were present in the single 500 nm crystals used for these

powder diffraction experiments is $8x10^8$.

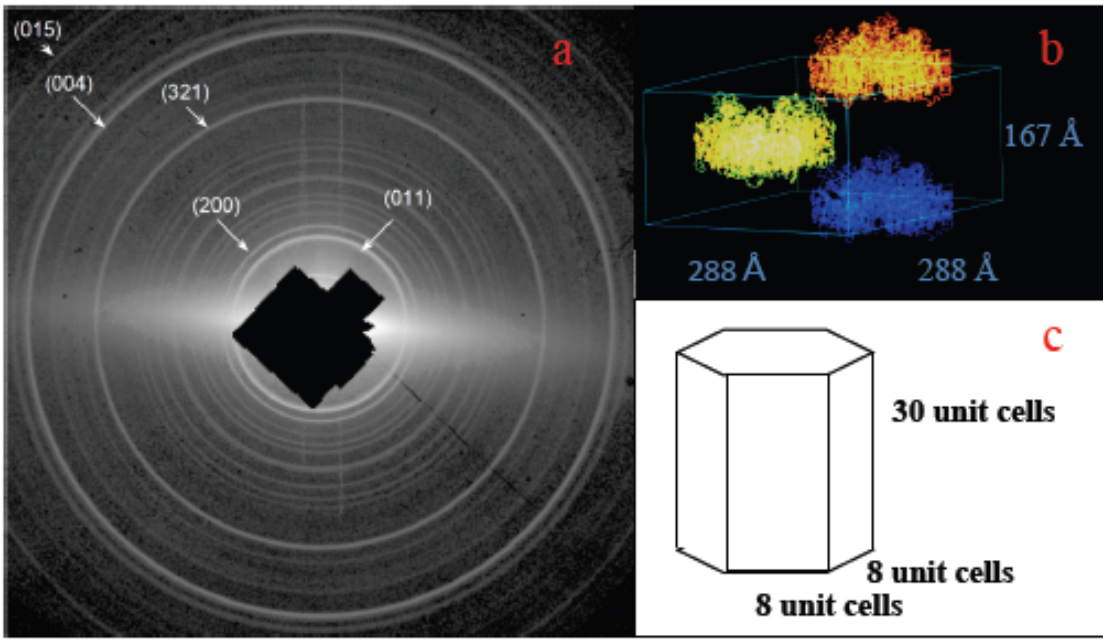

Figure 4.23: **Indexed 500-nm PSI powder pattern** (a) Diffraction pattern from ⩽500 nm PSI crystals with an on-axis detector and X-ray energy of 1560 eV (λ = 8 Å). The resolution at the corner of the detector is 28 Å. A maximum of 11520 PSI trimers are in each crystal. The pattern is wavelength limited. (b) A schematic view of the determined unit cell for PSI showing the unit cell dimensions. (c) A drawing showing the number of unit cells in a (500 nm)^3 crystal. Figure and caption adapted from (Hunter et al., 2011).

Additional diffraction data were collected with the detector moved off-axis diagonally, so that the center of the diffraction pattern was in one corner of the detector. The data shown in Fig. 4.24a are from ⩽(500 nm)^3 crystals using 1560 eV X-rays. At this energy, the maximum attainable resolution is 14 Å; however, the observable diffraction pattern extends only to about 28 Å (after adding ten 30-s exposures), indicating that the resolution was limited by the background scattering and peak overlap. At higher angles, the powder lines have smaller spacing; with the large unit cell of Photosystem I, it would have been impossible to separate the individual lines at 14 Å due to peak broadening from the instrument as well as the shape transform of the crystallites. As a result, the beamline was optimized for the third harmonic of the undulator, at 520 eV, allowing higher flux and scattering strength though a lower resolution of 40 Å in

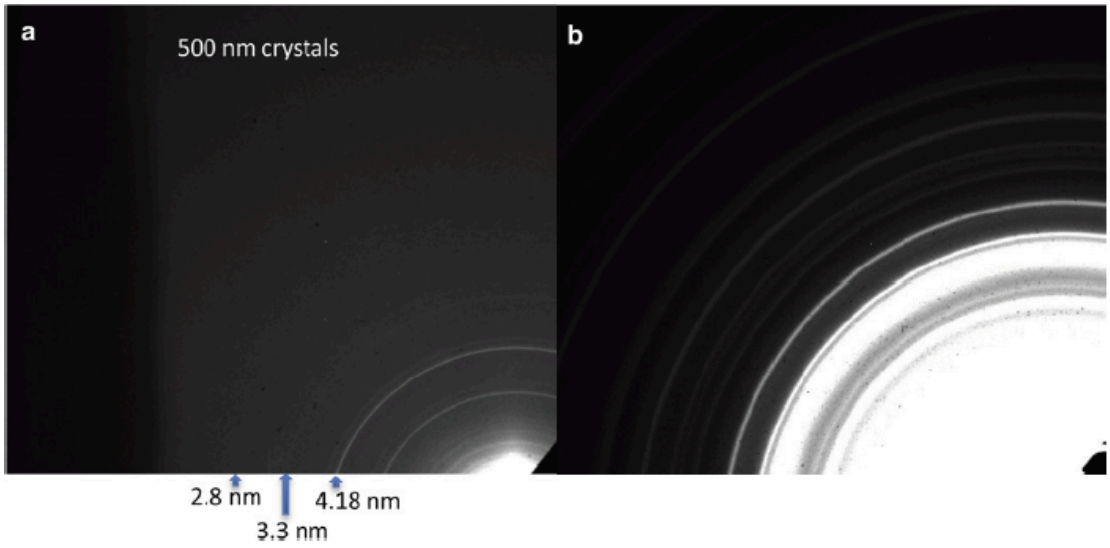

Figure 4.24: **Off-axis PSI crystallite powder diffraction patterns** Diffraction patterns from $\leqslant$ 500 nm PSI crystals using an off-axis detector to increase measurable scattering angle. (*a*) The crystals were irradiated with 1560 eV X-rays (λ = 0.8 nm) and had a resolution of 1.3 nm in the upper-left corner. (*b*) The crystals were irradiated with 520 eV (λ = 2.4 nm) X-rays and had a resolution of 4.0 nm in the upper-left corner. Arrows and associated numbers indicate the d-spacing of the given powder peak. Figure and caption taken from (Hunter et al., 2011).

the corner of the detector. The diffraction pattern recorded from the $\leqslant (500\text{ nm})^3$ crystals is shown in Fig. 4.24b. The data extends out to the highest measurable scattering angle corresponding to a resolution of 42 Å. Using longer wavelength X-rays results in a greater separation between diffraction rings and broader peaks due to particle size effects (*i.e.* more samples per ring and fewer ring overlaps). Though the resolution is limited due to the shorter wavelength, use of this energy can provide more complete information regarding the quality of the crystals studied.

The X-ray powder diffraction patterns shown in Figs. 4.23 and 4.24 are collected from crystals that are maximally 500 nm/side. However, the patterns could have contributions from crystallites of any size smaller than 500 nm/side. In order to elucidate the amount of scattering that was contributed from crystals of various sizes, X-ray diffraction data from crystals with a maximum linear dimension of 220 nm and 100 nm, limited by respective filter size, were collected with the off-axis detector and 520

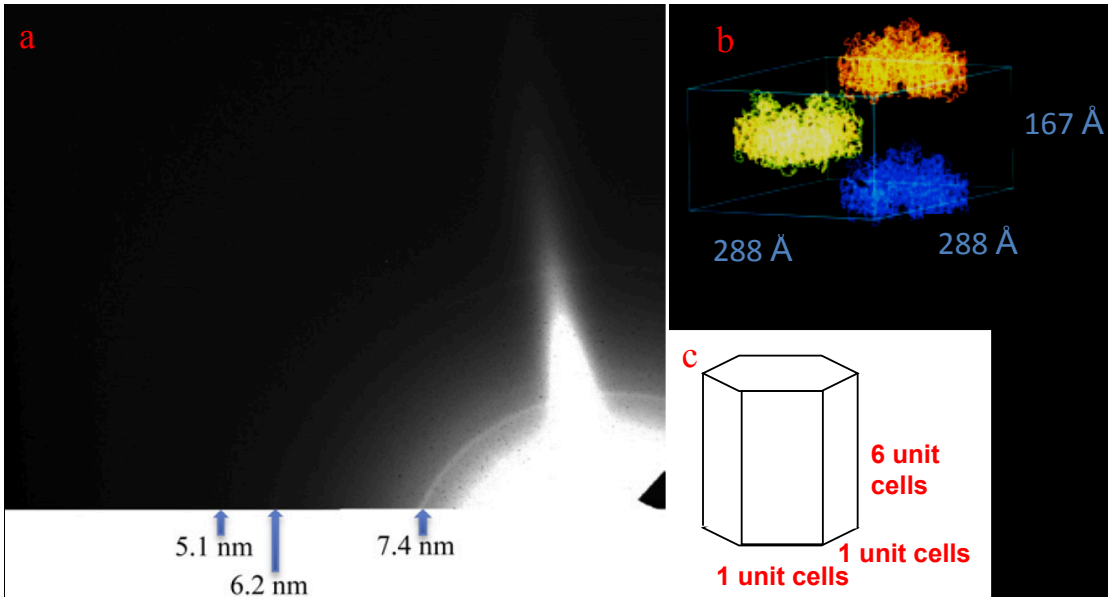

Figure 4.25: **100-nm PSI crystallite powder diffraction pattern** (a) Diffraction pattern from 100 nm PSI crystals with an off-axis detector and an X-ray energy of 520 eV (λ = 2.4 nm). The resolution at the corner of the detector is 4.0 nm. A total of 36 PSI trimers are in each crystal. (b) A schematic view of the determined unit cell for PSI showing the unit cell dimensions. (c) A drawing showing the number of unit cells in a $(100\ \text{nm})^3$ crystal, assuming only integer values of unit cells. Figure and caption modified from (Hunter et al., 2011).

eV X-rays. It should be noted that when placing the filtrate of the 100-nm filter under an optical microscope, the solution was light green but possessed no turbidity. Three powder rings were observed, and the data are shown in Fig. 4.25.

Although a hexagon with a diagonal of 3a (a = 28.8 nm) would fit through a 100 nm filter, this would require the edges to be 1.5a in length, which is 1.5 unit cells. If only integer numbers of unit cells are allowed along the edges of the hexagon, the total number of unit cells contained within the $(100\ \text{nm})^3$ crystals of Fig. 4.25 is 1 x 1 x 6 (x 3) equaling 18 unit cells and 36 total PSI trimers. Since a unit cell of PSI contains two PSI-trimer molecules, multiples of one half of the unit cell are possible. Therefore, if non-integer values of unit cells are permitted along the crystallite edges, the number of unit cells in a PSI crystal capable of passing through a 100-nm filter is 1.5 x 1.5x 6 (x 3), which is 40.5 unit cells containing 91 PSI trimers. The ratio

PSI sample type	Maximum single- crystal size (μm x μm x μm)	Total number of unit cells*	Total number of PSI trimers	Attained resolution (nm)
Typical large, single crystal	500 x 500 x 2000	$4x10^{13}$	$8x10^{14}$	0.24
The experiment with 500 nm filter	0.500 x 0.500 x 0.500	5800	11600	2.8
The experiment with 100 nm filter	0.100 x 0.100 x 0.100	18	36	5.1

Table 4.4: **Crystal size and diffraction resolution comparison** A comparison of the crystal dimensions, number of unit cells, number of PSI trimers, and attained resolution for the crystals used in the serial powder diffraction experiment with that typically used for single crystal X-ray crystallography of PSI.
*Calculation was done considering only integer values for the number of unit cells of PSI along an edge.

of the number of scattering unit cells in a 0.5 mm x 0.5 mm$x$ 2 mm single-crystal to the number of unit cells in a single (100 nm)^3 crystal used for this powder diffraction experiment is minimally $2.3x10^{11}$, and it is remarkable that diffraction was observed with a maximum resolution of 51 Å. One limitation of the 100-nm crystal experiment was that only a small volume of sample remained and had to be diluted 5-fold in order to be passed through the void volume of the filter. The dilution would have caused the intensity of the powder rings to be diminished, making the higher resolution rings more difficult to observe above the background and noise.

A summary of the maximum crystal size, number of unit cells, and number of PSI trimers used as samples in this experiment, as well as a typical sample used for single-crystal diffraction, is given in Table 4.4. The corresponding diffraction patterns were reduced to one dimension, through radial integration, and plotted simultaneously in Fig. 4.26.

An examination of Fig. 4.26 shows that the half maximum peak widths for the

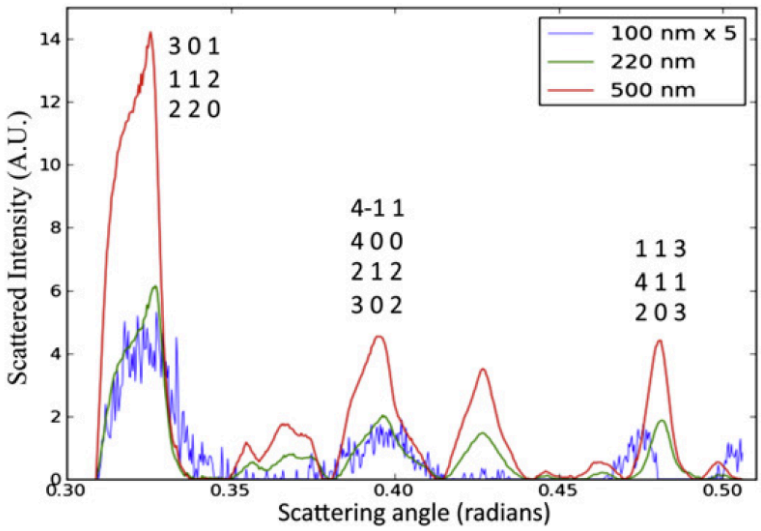

Figure 4.26: **Radially integrated PSI powder diffraction patterns** A comparison of the scattering power of the 100 nm, 220 nm, and 500 nm crystals at specific Bragg reflections using the scattered intensity as a function of scattering angle. Figure and caption taken from (Hunter et al., 2011).

powder rings are approximately the same for the different filter sizes, with a FWHM of 0.020 (+/-) 0.0025 rad. The peak widths will depend on the crystallite sizes and distribution and could vary significantly between crystal batch preparations. The Scherrer equation is used to determine the average particle size by relating the full width at half max (FWHM) of a peak in a powder diffraction pattern, h, to the crystal size as follows (Scherrer, 1918):

$$h = 2\sqrt{\frac{\ln(2)}{\pi}}\frac{\lambda}{\Lambda}\frac{1}{\cos\theta/2} \tag{4.1}$$

where λ is the wavelength of the X-rays, Λ is the length of the side of the nanocrystal, assuming the crystal is a cube, and $\theta/2$ is the Bragg angle. The Scherrer equation considers broadening caused by crystal size; however, the measured peak width is also a function of the beam divergence and bandwidth, as well as characteristics of the sample, including the internal disorder of the molecules within the crystal, as expressed by the term mosaicity. Therefore, the FWHM of the peak is determined by the convolution of Gaussian peak profiles from the crystal size, instrument broadening, and the mosaicity

Mosaicity (°)	Calculated peak width for 100 nm crystallites (rad)	Calculated peak width for 220 nm crystallites (rad)	Calculated peak width for 500 nm crystallites (rad)
0.001	0.024	0.011	0.0052
0.1	0.024	0.011	0.0055
0.5	0.025	0.014	0.010
1.0	0.029	0.021	0.018

Table 4.5: **Effect of mosaicity on peak widths in serial powder diffraction** Calculations of the expected peak FWHM for the powder diffraction patterns of PSI nano- and microcrystals with various values of mosaicity using Eqs. 1 and 2. Note that the experimental determined line width is (0.02 $\pm$ 0.0025) rad.

of the crystal. The following formula can be used to calculate the expected FWHM of the diffraction peaks for the crystallites used in the experiment, $\delta\theta_{tot}$, for comparison with the given line widths in Figure 4.26:

$$\delta\theta_{tot} = \sqrt{(\delta\theta_{size})^2 + (\delta\theta_{inst})^2 + (\delta\theta_{mos})^2} \tag{4.2}$$

where $\delta\theta_{size} = h$ is the broadening of the peak due to the crystal size, $\delta\theta_{inst}$ is the instrumental broadening, and $\delta\theta_{mos}$ is the broadening caused by mosaic spread in the crystal. Using Eqs. 4.1 and 4.2, the predicted diffraction peak broadening for the various nanocrystal sizes used in the experiment were calculated and are listed in Table 4.5, using the known instrument broadening of beamline 9.0.1 of 2.5 mrad (Shapiro et al., 2008) and using putative values of mosaicity of 0.001° to 1.0°, where 0.5° to 1.0° are typical mosaicity values for large PSI crystals. As can be seen in Table 4.5, the calculations show that the experimental line width of 0.02 $\pm$ 0.0025 is in good agreement with nanocrystal sizes of 100 to about 250 nm but could also be explained by larger crystals with higher mosaicity. However, as discussed in Section 4.7, femtosecond nanocrystallography experiments, in which diffraction patterns were collected from individual PSI nanocrystals and microcrystals, show low mosaicity (below 0.1°).

After seeing the possible effects of the mosaicity on the diffraction peak widths, as shown in Table 4.5, a discussion of mosaicity is warranted. Practically, mosaicity is the misalignment of repeating units of a crystal that is manifest in the intensity of a Bragg reflection being subtended over a larger solid angle. Mosaicity is often presented through the mosaic block domain, in which a crystal is composed of domains that are extremely well ordered in the short-range, but become less ordered with respect to long-range considerations. This is an incomplete picture in the case of membrane proteins. If PSI crystals are considered, only four salt bridges make up the crystal contacts, which could be a primary cause of misalignment. However, misalignment of the unit cells could also occur due to excess surface area with respect to volume because the molecules at the interface will have a reduced number of crystal contacts. In the case of a $(500\ \text{nm})^3$ PSI crystal, 30% of the unit cells and 15% of the PSI molecules will be exposed to the surface. Therefore, two effects can be seen as influencing the mosaicity of the crystal: For large crystals, it is the propagation of slight misalignments in different domains, whereas for small crystals, the increased surface area with respect to volume may cause misalignment of a large number of molecules. As such, there may be a minimum in the mosaicity versus size of crystals, but currently, the mosaicity of the nanocrystals is unknown. Data collection using the LCLS—Section 4.7 is currently being evaluated to correlate the nanocrystal size to maximum resolution and mosaicity. However, the Scherrer broadening may dominate the spot size, making it difficult to determine the mosaicity of the crystallites with any confidence, because the broadening adds in quadrature.

Although the crystals were filtered to limit the maximum size, the protein crystals in the solution after filtration would be in equilibrium with any protein in solution. This would, in principle, allow for re-growth of larger crystals after filtration over a long time period. However, this will not occur in the time frame of our experiment for several reasons. The buffer used to transport the PSI crystals to the interaction

region contained no salt, and the Photosystem I crystals are formed through electrostatic interactions between the proteins as the ionic strength of the solution is reduced. The solubility of PSI under these conditions is less than 0.1 μg/mL, and therefore, the number of free PSI-detergent micelles would be very low while the supersaturation is high. Therefore, at this high super-saturation condition, dissolving and regrowing of the crystals is extremely slow. Crystal growth occurs through diffusion of the large protein-detergent complexes (MW = 1,300,000 Da) through the solution to the surface of the crystal, but the diffusion constant for such a large complex is small.

The diffusion constant of the PSI trimer can be calculated using the Einstein-Stokes equation, in which case the PSI trimer molecule will be treated as an oblate spheroid ($a = b \geqslant c$), with radius of 13 nm (11 nm diameter of PSI trimer molecule plus a 2 nm length from the β-DDM molecules) and a height of 3.2 nm. The Einstein-Stokes equation is modified to accommodate an ellipsoidal molecule as follows (Perrin, 1936; Edward, 1970):

$$D^0 = \frac{\mathrm{k}T}{6\pi\eta r(f/f_0)} \tag{4.3}$$

where k = $1.381x10^{-23}$ J/K is the Boltzmann constant, T is the temperature in Kelvin, η is the dynamic viscosity in N·s/m^2, r is the radius of a sphere with the same volume as the ellipsoidal particle, and (f/f_0) is the ratio of the frictional coefficient of the ellipsoidal particle to the spherical particle of identical volume and is unitless. Calculating the axial ratio for the PSI trimer to be 12.5 nm/ 3.2 nm = 3.9, yields a (f/f_0) of 1.16 (Edward, 1970). A sphere of equal volume to the PSI trimer would have a radius of 7.1 nm. The dynamic viscosity of water at 20°C is 1.0016 μN·s/m^2 (Wagner and Pruss, 2002), and therefore, the diffusion constant of the PSI trimer molecule at 20°C is calculated to be $2.55x10^{-9}$ m^2/s. This diffusion constant is so small and the concentration of free PSI-trimer molecules in solution is so low that growth of crystals in the sort period of time after filtration can be excluded.

Furthermore, the solution is not static; rather it flows towards the nozzle. There-

fore, the convection term should dominate the motion of the particles. The crystals were filtered one meter before the interaction region of the X-rays, and were flowing within a 50-μm ID fiber optic at 10 μL/min, which corresponds to a 12-s transit time. It is highly unlikely that significant dissolution and re-crystallization could occur in such a short time frame.

One additional finding of the serial powder diffraction work is that once large, $\leqslant$ (10-μm)^3, PSI crystals were used in the jet, weak flow alignment of the crystals becomes measureable. The flow alignment would result in arcs in the powder diffraction pattern, instead of rings, indicating the preferential alignment of the crystals. For PSI, once 10-μm crystals were used, flow alignment was noticed, as shown in Fig. 4.27. The diffraction pattern shown represents the difference of the 10-μm PSI crystal data and the 2-μm PSI crystallite data for the same batch of protein. The reason for the flow alignment is that at the length scales used in the nozzle, the flow must be laminar; *i.e.* there cannot be turbulence (small Reynolds numbers). The flowing crystals are aligning along to the streamlines, and upon reaching the necking down region, the crystals will be forced to align because of the immobility with respect to the streamline. However, the flow alignment will only occur when the crystals have a direction of preferential growth, *i.e.* crystals that are elongated in one direction. The exact aspect ratio needed for flow alignment is currently unknown, but as larger PSI crystals may be imperative for higher-energy LCLS experiments, the flow alignment as a function of size may require the collection of large data sets due to the preferential orientation of the crystals in the stream.

Challenges Associated with Serial Powder Diffraction

The main challenge associated with the serial powder diffraction data experiments was that powder diffraction was used to evaluate the diffraction power of the PSI nanocrystals. Powder diffraction is a suitable method for structure determination for small

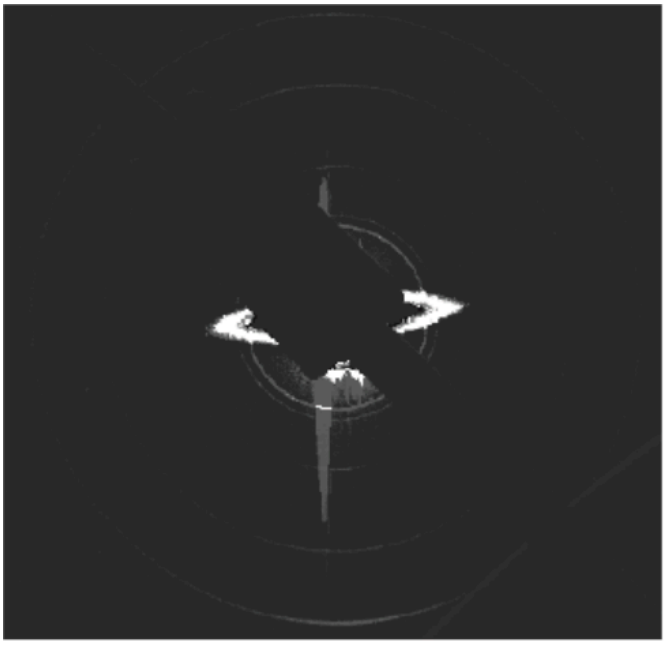

Figure 4.27: **Photosystem I powder pattern with arcs** Powder diffraction pattern of 10-μm crystals of photosystem I generated by subtracting the 2-μm crystallite data from 10-μm crystal data for the sample photosystem I crystal batch. The anisotropy in the pattern is indicative of preferential alignment of the photosystem I crystals in the X-ray beam, caused by flow alignment in the liquid jet.

molecules and peptides, but becomes less effective as the unit-cell volume of the crystal of interest increases (Margiolaki and Wright, 2008), due to a loss of information in powder diffraction. This is due, in part, to the reduction of three-dimensional reciprocal space information into one-dimension, *i.e.*, the diffraction data is treated as a function of d alone. In addition to the distinct possibility of peak overlap caused by the large unit-cell volume of typical protein crystals, peak broadening due to instrumental factors, such as beam divergence, energy spread, etc., will add to peak broadening caused by finite size effects and mosaicity effects. As the protein of interest transitions from small, soluble proteins, to membrane proteins and protein complexes, the broadening of the peaks can become an insurmountable factor in the analysis of powder diffraction data. Furthermore, the small spacing of the powder diffraction lines leads to a "continuum" at higher diffraction angles. Therefore, the serial powder diffraction experiments could not be a stand alone experiment used as a general method to determine the struc-

ture of membrane proteins using membrane protein nanocrystals and microcrystals.

Summary of Serial Powder Diffraction

Serial powder diffraction showed the presence of PSI crystals that contained only a few hundred unit cells and provided a proof of concept that these PSI crystals could be used to collect X-ray diffraction data to low resolution. The PSI powder diffraction patterns showed Bragg spacings that were commensurate with the values obtained from large PSI crystals, indicating that the PSI nanocrystals and microcrystals have the same unit cell constants and could be useful for structure determination experiments. Additonally, the gas-dynamic virtual nozzle developed at Arizona State University for delivering the PSI nanocrystals and microcrystals (DePonte et al., 2008) was shown to be an effective way of delivering the PSI-nanocrystal sample with full hydration.

More generally, the serial powder diffraction experiments resulted in the understanding that very small crystals of membrane proteins can be made, with less than one-hundred unit cells, and that the nanocrystals were still ordered enough to produce measurable diffraction. The serial powder diffraction experiments indicate that the technique will be useful for characterization of membrane protein nanocrystals and microcrystals in preparation for femtosecond nanocrystallography experiments.

4.7 Initial femtosecond nanocrystallography experiments

Motivation for the femtosecond nanocrystallography experiments

The necessity for the use of large, well-ordered protein crystals in X-ray protein crystallography is due to X-ray-induced radiation damage. A large amount of the time during structure determination of a protein is occupied with finding appropriate crystallization conditions for the growth of large, well-ordered, single crystals.

Membrane protein crystallization is notoriously difficult, and has become a major bottleneck in the determination of membrane protein structures by X-ray protein

crystallography, as discussed in Section 1.4. During membrane protein crystallization experiments, microcrystals are found much more frequently than the desired macrocrystals, and numerous hours and resources get devoted to optimizing the crystallization conditions to produce macrocrystals from conditions that originally produced microcrystals, often unsuccessfully. However, due to the importance of membrane proteins for cellular function, such as cellular respiration and signal transduction, to name only two, a method that can be used to facilitate the determination of membrane protein structures is greatly sought.

At this point, the serial powder diffraction experiments have shown that nanocrystals and microcrystals of PSI, as small as a few hundred total unit cell, exist and exhibit enough order to diffract X-rays with measurable Bragg peaks. The preparation of PSI nanocrystals and microcrystals was shown to be reproducible and predictable. The growth of PSI nanocrystals likely indicates that nanocrystals of other membrane proteins can be grown and may even be produced more commonly than microcrystals of membrane proteins.

The serial powder diffraction experiments were shown to be an excellent method to characterize the membrane protein nanocrystals and microcrystals. Unfortunately, powder diffraction experiments are not practical for structure determination of proteins with large unit cells, as the overlap of powder diffraction rings makes data analysis very challenging, if not impossible. Single-crystal X-ray diffraction experiments are necessary to allow for the structure determination of a protein with as large of a unit cell as PSI.

The need for single-crystal techniques warranted a great deal of experimentation for the development and implementation of microfocused beamlines that irradiate smaller volumes of crystals and can have higher flux-densities than regular X-ray protein crystallography beamlines (Riekel et al., 2005). The microfocused beamlines compensate for smaller protein crystals by offering a higher flux-density, although X-

ray-induced radiation damage sets a limit to the size of crystals available for investigation. As an example, the PSI nanocrystals and microcrystals are too small to be useful at a microfocused beamline because X-ray-induced radiation damage becomes an insurmountable obstacle. For these nanocrystals and microcrystals of PSI to be useful in a structure determination method, a single-crystal experiment that overcomes the problem of X-ray-induced radiation damage accumulating in the structure during exposure is necessary. Luckily, the new X-ray free-electron laser offers the potential to break the nexus between X-ray dose and damage.

The X-ray free-electron lasers (XFELs) are a new type of X-ray source—discussed in Sections 2.2 and 2.4—that produce a pulsed X-ray beam with a pulse duration between tens and a few hundreds of femtoseconds. The XFELs can be 10 orders of magnitude more brilliant than the third-generation synchrotrons currently utilized for X-ray protein crystallography, and can have 10^{12} higher photon-flux than third-generation synchrotrons (Altarelli, 2010). Due to the high brialliance and short pulse durations of the X-ray laser, an opportunity has arisen to break the nexus between radiation dose and spatial resolution. It has been proposed that femtosecond X-ray pulses can be used to out-run damage processes, when using single pulses so brief that they terminate before the manifestation of damage of the sample (Neutze et al., 2000). Experiments at the FLASH VUV free-electron laser (FEL) confirmed this idea at resolutions lengths of 200–6Å (Chapman et al., 2007). Theoretical studies suggested that a focused pulse from an FEL might be so intense to produce a diffraction pattern at near-atomic resolution (Neutze et al., 2000). The successful operation of the AMO beamline of the LCLS (Bozek, 2009) offered the first X-ray free-electron laser, with wavelengths initially down to 6.9 Å, for use in biological imaging and crystallography experiments. The wavelength was finally short enough to allow for significant numbers of Bragg reflections to be measured for a protein crystal with a large unit cell, such as PSI.

The motivation behind the initial femtosecond X-ray protein nanocrystallog-

raphy experiments was to provide a proof of principle of the diffract-before-destroy concept for protein crystallography using membrane protein crystals. If the diffract-before-destroy principle is shown to be successful for PSI nanocrystals and microcrystals, the femtosecond X-ray protein nanocrystallography experiment would allow for the results of many crystallization screens that produce microcrystals to be used directly for structure determination experiments, without the need for subsequent optimization experiments. Success with the technique of femtosecond X-ray protein nanocrystallography would provide a new avenue for the structure determination of membrane proteins, offering the ability to use membrane protein crystals of any size for structure determination experiments.

Crystals of the cyanobacterial Photosystem I complex were chosen as a model system. PSI is a large membrane protein complex (1 million Dalton molecular weight, 36 proteins, 388 cofactors) that acts as a bio-solar energy converter in the process of oxygenic photosynthesis. PSI is one of the most complex membrane proteins for which an X-ray structure has been determined (PDB code 1JB0). With a large hexagonal unit cell (a = b = 281 Å, c = 165 Å) and high solvent content (78%), it presents a most stringent test for the analysis of nanocrystals by femtosecond X-ray crystallography.

Three major experiments were done during these femtosecond nanocrystallography experiments. The first experiment was to record diffraction data to the highest allowable resolution, of 8.5 Å, and use the data to produce merged intensities, calculate structure factors, and produce an electron-density map of PSI. Complementary experiments were conducted to assess the internal consistency of the LCLS data as well as to compare the LCLS data to conventional X-ray crystallography data for PSI. The second major experiment done was to compare the diffracted intensity for PSI crystallites at several different X-ray pulse durations, in order to gain insight into the X-ray-induced radiation damage to the PSI crystallites. The third major experiment was to record the interference patterns around the low-resolution Bragg reflections that are caused by the

finite crystallite size.

Acknowledgments regarding the PSI femtosecond nanocrystallography

A huge international collaboration was necessary to establish the first proof of principle of the new technique of femtosecond nanocrystallography work. My role was to find the necessary conditions to produce nanocrystal and microcrystal samples of PSI, produce the sample, work on the design of the experiments, assist with the sample delivery, and collect the X-ray-diffraction data.

The majority of the analysis of the PSI data was done by Rick Kirian of Arizona State University, Tom White of the Center for Free Electron Laser science at DESY labs, with major consulting by James Holton of Lawrence Berkeley National Laboratory. Raimund Fromme contributed by collecting conventional crystallography PSI data sets on large PSI crystals that I grew, while also producing electron-density maps of the LCLS and ALS data. Raimund Fromme and James Holton were also involved in the refinement of the PSI electron-density maps.

The experiments were done at the AMO beamline (Bozek, 2009) of the LCLS, and I would like to thank the hard work of the entire LCLS team. The experiments were completed in the CAMP instrument (Struder et al., 2010), and I would like to thank the entire CAMP team for their hard work.

A complete list of the collaborators and acknowledgement of the contribution of each collaborator can be found in (Chapman et al., 2011).

Femtosecond nanocrystallography experimental overview

Preparation of the sample for femtosecond nanocrystallography

The PSI nanocrystal samples used for the first femtosecond X-ray protein nanocrystallography experiments in Dec. 2009 were obtained from two different large-scale PSI preparations, which were completed directly before the LCLS experiments. A to-

tal of three large-scale PSI preparations were completed in anticipation of the LCLS experiments. The PSI crystallite samples were stored at 4°C, and were only filtered directly before use in the experiments. The PSI nanocrystals from the 40-min settling steps that were used in the experiments were restricted in size by using a 2-μm inline filter; therefore, the PSI nanocrystals contained between 100 and (maximally) 500,000 unit cells for the 2-μm microcrystals. However, there were no visible crystals in the re-suspended sample, and most of the crystals were $\leqslant$ 500 nm in size. It should be noted that even the 2-μm PSI crystallites were between 10^7 and 10^8 smaller than the PSI crystals used in conventional crystallography.

Once filtered, a 25-μL aliquot of the sample was used to do a (635 mg/195 μL / 5 μL) chlorophyll assay. The PSI-crystallite density was approximated using the calculated chlorophyll concentration, and the crystallite density was set such that there would be one PSI crystallite occupying every fourth X-ray interaction volume, as discussed in Section 3.7. Based on crystal-density calculations, the sample was diluted to 1 mg/mL PSI and loaded into the sample line of the dual-line system, shown in Fig. 3.10. The liquid injector was run with a flow rate between 10 and 12 μL/min, and the jet was centered in the X-ray beam by observing a streak on the back detector, due to scattering of the X-rays by the column of the jet. Data for the PSI samples were collected over the course of 48 continuous hours of experiments. Millions of diffraction patterns were recorded, corresponding to 25 TB (25,000 GB) of data.

Diffraction experiments

The soft X-ray femtosecond nanocrystallography experiment was done by introducing a stream of fully-hydrated PSI crystallites to the LCLS X-ray pulses with pulse durations between 10 and 250 fs, operating at a repetition rate of 30 Hz, through a liquid jet at the AMO beamine of the LCLS utilizing the CAMP chamber. The X-ray energy used for the initial experiments was 1.8 keV (λ = 6.9 Å), and diffraction was recorded

on two sets of detector modules: a "front" detector, able to record reflections to maximum resolution of 8.5 Å at the corner, and a "back" detector, able to record reflections between 400 and 10 nm (Chapman et al., 2011). During the femtosecond nanocrystallography experiment, one pulse of the LCLS would ideally inundate one PSI crystallite, producing a diffraction pattern recorded on both the front and back detectors that is read out and digitized before the subsequent LCLS pulse arrives.

The front detector allowed for the recording of the maximum allowable diffraction angles, while the back detector allowed for recorded data with high angular resolution. The LCLS generates an X-ray pulse with high transverse coherence, and the high angular resolution of the back detectors allows for the interference fringes expected from the finite crystal size in a coherent source, analogous in idea to the fringes in a slit experiment in diffraction physics, to be resolved (Woolfson, 1997).

Data processing and analysis

A detailed description of the data analysis of the PSI data from the LCLS is provided in (Kirian et al., 2011), but a brief description is given here. The diffraction patterns were processed by removal of any known bad detector pixels, followed by application of a flat-field correction, normalization of the pixel values with respect to the dark current and gain of the panel, and subtraction of the background. The background was defined as a moving-window selection of 50 frames, with the background for a given pixel defined as the median value over the range (Kirian et al., 2011). However, the moving-average method of background subtraction is predicated on a well-behaved jet.

Once the diffraction data were fully processed, indexing and merging of the data were performed. However, for each X-ray pulse, the Ewald sphere would slice through the reciprocal lattice of the PSI crystal in a random position. Due to the sparse population of reciprocal points in the reciprocal space of the crystal system, most scattering would not occur due to the intersection of the Ewald sphere with a reciprocal lattice

point. Instead, much of the scattering is due to intersection of the Ewald sphere with the shape transform laid down upon the reciprocal lattice point. A peak finder written in Matlab was used to establish peaks, and the peak locations were given to MOSFLM (Leslie, 2006) using the DPS algorithm (Steller et al., 1997) or DirAx (Duisenberg, 1992) for indexing. Indexing the patterns would give the orientation of the crystal relative to the lab frame, and could be used for further processing. Each orientation matrix was further refined by a global optimization of the unit-cell constants and three Euler angles (Kirian et al., 2011).

The LCLS data collected was tested for internal consistency by generating random subsets of the data. The subsets used to determine the internal consistency were defined such that all patterns were counted, and the patterns with an even number were placed in one subset, whereas he patterns with an odd number were placed into a separate subset. The two subsets were compared with linear correlation coefficients and an R factor, as defined in (Kirian et al., 2011). The linear correlation coefficient is defined as follows:

$$C_{\text{int}} = \frac{\sum_{hkl}(F_{\text{even}} - < F_{\text{even}} >)(F_{\text{odd}} - < F_{\text{odd}} >)}{\sum_{hkl}[(F_{\text{even}} - < F_{\text{even}} >)^2]^{1/2}[(F_{\text{odd}} - < F_{\text{odd}} >)^2]^{1/2}} \tag{4.4}$$

where F_{even} are the structure factors associated with even frames and F_{even} are the structure factors associated with odd frames. Perfectly correlated data between the even and odd subsets would yield a correlation coefficient of 1.

The internal-consistency R-value, R_{int} is defined as follows:

$$R_{\text{int}} = \frac{\sum_{hkl} ||F_{\text{even}} - F_{\text{odd}}||}{\sum_{hkl} |(|F_{\text{even}}| + |F_{\text{odd}}|)/2|} \tag{4.5}$$

In addition to internal consistency, the structure factors extracted from the LCLS data were compared to structure factors extracted from a conventional crystallographic

data set of PSI collected at beamline 8.2.2. of the Advanced Light Source of Lawrence Berkeley National Laboratory. The conventional data and LCLS data were collected on samples prepared quite differently, as the conventional data originated from a large single crystal of PSI that had been incubated in 2 M sucrose and cryogenically frozen using liquid propane. The conventional data were collected specifically with large working distance and a 100x attenuated beam, so that the low-resolution data were not from saturated detector pixels.

Femtosecond nanocrystallography results for PSI

A total of 1.85 million "snap-shot" diffraction patterns from PSI were collected at the 70-fs pulse duration. Bragg peaks from the PSI crystallites were recorded to resolution of 8.5 Å on the front detector, as shown in Fig. 4.28. The diffraction of the nanocrystals was so strong that even saturated peaks appeared occasionally in the diffraction patterns. Unlike in cryo-EM or traditional crystallography, the X-ray diffraction patterns were collected at room temperature, without any crystal mounting, freezing, or cryo-protectants. The diffraction patterns showed well-resolved peaks, and some orientations provided many Bragg peaks per pattern.

A 20% hit rate was achieved by crystal-density calculations, as detailed in Section 3.7. Most of the recorded crystallite diffraction patterns were from one X-ray pulse hitting one nanocrystal, with the patterns on the front and back detectors reading out before the arrival of the next pulse. Fine adjustment and optimization of the crystal density was done based upon the perceived hit rate as determined from the data stream at the LCLS control room.

Of the 1.85 million diffraction patterns of PSI crystallites that were collected at the 70-fs pulse duration, over 112,000 contained ten or more identifiable peaks, of which a total of 28,192 were indexed with unit cell parameters within 5% of the mean values, with a representative image shown in Fig. 4.29a. Indexing success was

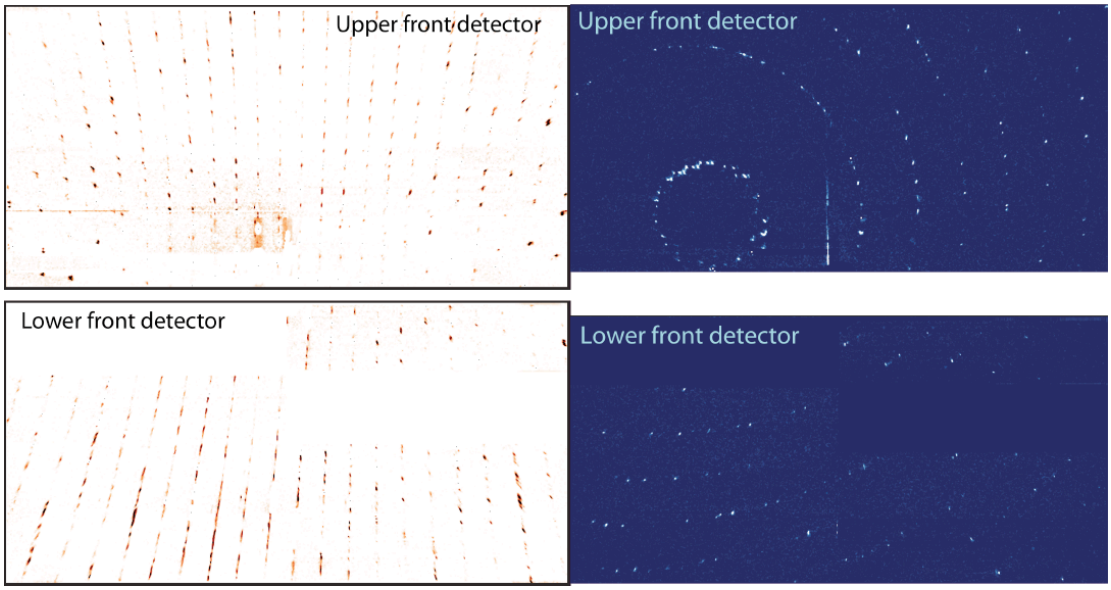

Figure 4.28: **Single-shot crystal diffraction patterns of PSI nanocrystals** (a, Left) A pattern from a crystal oriented with its *c* axis almost horizontal, perpendicular to the X-ray beam. (b, Right) Reciprocal lattice planes are seen as arcs, due to the intersection of the Ewald sphere with the lattice. Cuts through the shape transform, due to the finite crystal size, are evident when the intersection of the Ewald sphere with the reciprocal lattice planes is close to tangential, such as the almost continuous lines of intensity in the bottom left of (a) and the spots in the complete circle of reflection in the top left of (b).

strongly correlated to the number of peaks contained within the pattern, as 60% of the patterns containing 75 or more peaks were indexed. However, approximately 42% of indexed patterns were rejected during the integration steps, due to disagreement between the predicted and observed peak positions. The reason for the disagreement is still unknown, but deviations between the predicted and observed peak positions likely resulted from slight indexing errors that produced large effects at the higher-resolution reflections. 16,374 patterns were integrated and contributed to the structure-factor amplitudes used to produce the PSI structure (Chapman et al., 2011).

In order to further evaluate the nanocrystallography data to 8.5 Å , rigid body refinement of the published PSI structure (PDB code 1JB0) against the 70-fs pulse-duration nanocrystal structure factors was performed using the program *REFMAC* (Murshudov et al., 1997), yielding $R/R_{free} = 0.284/0.327$. In addition, rigid body refinement

of the published PSI structure was performed against the structure factors extracted from a higher-resolution single-crystal PSI data set, in which good low-resolution data were measured and collected at beamline 8.2.2 of the ALS. The ALS data set was truncated to 8.5-Å resolution, yielding R/R_{free} of 0.285/0.298. Figures 4.29c-d show representative regions of the $2mF_0$-$2DF_c$ averaged-kick-map electron density at 8.5 Å from the LCLS and ALS data sets, respectively. The electron density maps clearly show the transmembrane helices, as well as electron density from the membrane extrinsic subunits, and, remarkably, the loop regions are also visible at 8.5-Å resolution. The reference electron-density map in Fig 4.29d from a large, single crystal of PSI was produced using data collected using 12.4-keV X-rays, with a single crystal of Photosystem I that had been cryogenically cooled to 100 K and had 2.00 M sucrose as a cryo-protectant (Chapman et al., 2011).

The refinement statistics for the PSI electron-density map calculated from the LCLS data is shown in Table 4.6. The data show good statistics for all categories until the highest resolution shell, in which case the data were measured in the corner of the detectors, causing a reduced number of total measurements, likely contributing to the poorer statistics.

Internal consistency of the data

In order to ascertain whether the LCLS data were internally consistent, appropriate error metrics, such as R_{int} and C_{int}, were defined (Kirian et al., 2011) and shown as a function of resolution in Fig. 4.30. R_{int} and C_{int} compared the data from odd and even number patterns, assigned arbitrarily as described in Section 4.7. The error metrics depended on the value of the integration radius δ, with R_{int} decreasing with increasing integration volume. For δ = 0.0087 nm^{-1}, which is 1/(4·a), R_{int} asymptotically approached a value of less than 10% when thousands of patterns were included in the calculation, indicating good internal agreement within the LCLS data. Additionally,

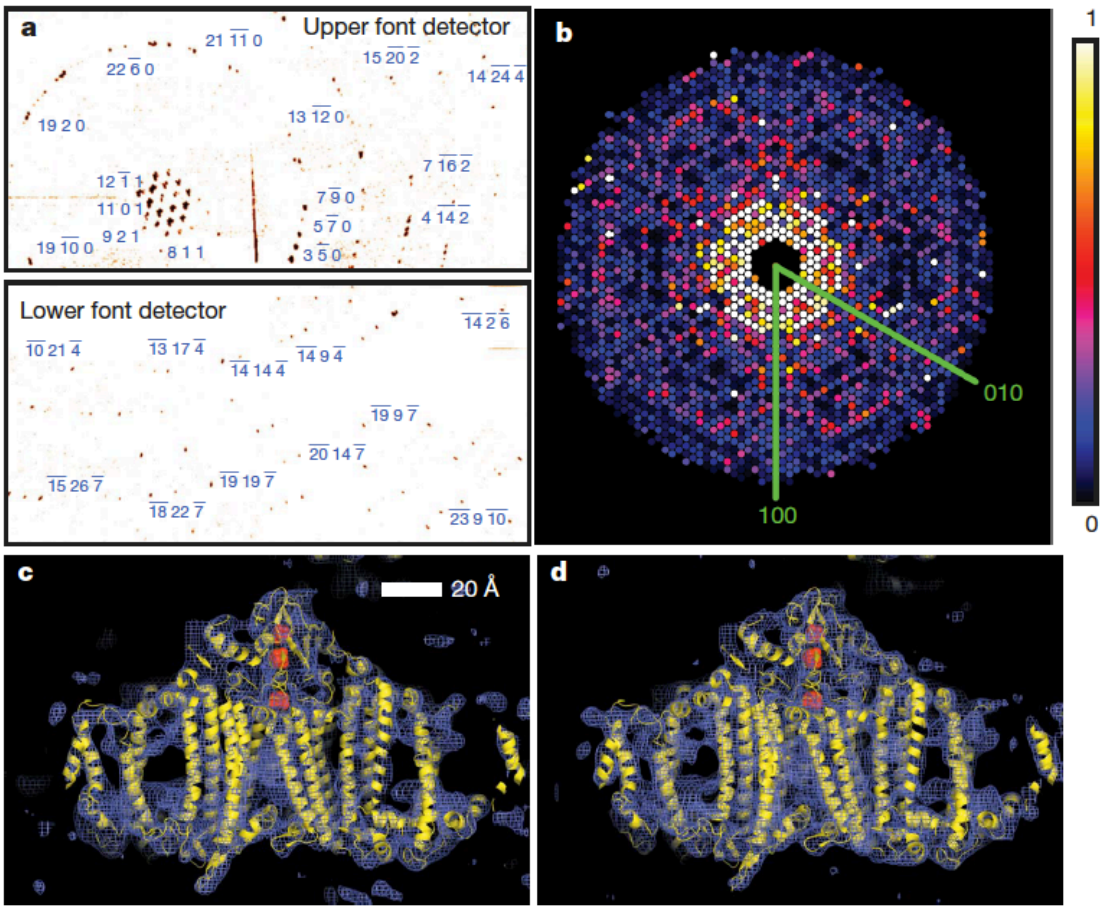

Figure 4.29: **Indexed diffraction patterns and electron density of PSI from the LCLS data** (a) Diffraction pattern recorded on the front pnCCDs with a single 70-fs pulse after background subtraction and correction of saturated pixels. Some peaks are labeled with their Miller indices. The resolution in the lower detector corner is 8.5 Å , (b) Precision-style pattern of the [001] zone for Photosystem I, obtained from merging femtosecond nanocrystal data from over 15,000 nanocrystal patterns, displayed on the linear color scale shown on the right. (c) Region of the the $2mF_0 2DF_c$ electron density map at 1.0σ (purple mesh), calculated from the 70-fs data and (d) from conventional synchrotron data truncated at a resolution of 8.5 Å and collected at a temperature of 100 K. The refined model of PSI is depicted in yellow. Figure and caption modified from (Chapman et al., 2011).

Resolution shell (Å)	Number of unique reflections	Multiplicity	Completeness (%)	¡Peak height/ Noise¿	Merged I/σ	$R_{iso}(F)$ (%)
30.0	416	1369	100	889	16.8	37.6
16.2	382	1278	100	29.3	15.9	15.1
13.6	362	1195	100	19.2	16.2	12.3
12.1	366	1009	100	20.3	14.7	12.9
11.1	357	678	100	19.5	12.7	14.6
10.4	355	492	100	15.4	10.6	28.7
9.85	349	283	100	14.0	9.20	22.2
9.39	352	111	100	11.8	7.12	28.1
9.01	352	28.6	99.4	9.50	4.99	24.8
8.68	88	3.92	25.1	6.70	3.29	30.8

Table 4.6: **Statistics for the merged dataset of PSI at 70-fs pulse duration** $R_{iso}(I)$ compares the LCLS structure factors to a conventional single-crystal dataset collected at the ALS beamline 8.2.2. The bottom center edge of the bottom detector occurs at a resolution length of 10.0 Å.

the linear correlation coefficient C_{int} was over 0.9 until a resolution of approximately 12.5 Å . Beyond 12.5 Å, C_{int} quickly fell to a value of less than 0.2 at 9-Å resolution. One possible reason for the apparently reduced internal consistency of the data at the highest resolution recorded in the experiment is that the unit cell constants determined during indexing could have had small errors that would produce larger effects at higher resolution. Additionally, because the highest resolution data were collected in the corners of a square detector, the highest resolution data were sampled with the lowest frequency, which may be alleviated by collecting more data.

Comparison of the LCLS data to conventional data

One important comparison required for the new PSI femtosecond X-ray protein crystallography data is with conventional X-ray protein crystallography data of PSI collected at a conventional beamline. The comparison will allow the LCLS data to be compared against the standard for X-ray crystallography.

The LCLS structure factors were compared to structure factors determined from PSI data collected conventionally at beamline 8.2.2 of the ALS, on large crystals at cryogenic temperatures, as a measure of the relative accuracy of the LCLS structure factors. The ALS data were collected using a large, single crystal of PSI that was cryogenically frozen in liquid propane, in which the data were collected to a resolution of 2.3 Å, but the data were truncated at 8.5 Å . The overall R_{iso} between the LCLS and ALS structure factors was 23.5%, indicating the structure factors from the LCLS and ALS data are indistinguishable. Although R_{iso} approached 60% for the highest resolution shell (containing the 8.5-9.0Å data), many factors, such as limited numbers of reflections to the highest resolution of the LCLS data, could have contributed to the divergence. The linear correlation coefficient, C_{iso}, comparing the LCLS and ALS structure factors for PSI, approached 0.9 around a resolution of 15 Å, and quickly fell off at higher resolutions, again indicating the similarity in the data sets for the resolutions in which a large number of reflections were recorded and processed (for the LCLS data) (Kirian et al., 2011).

Deviations between the LCLS and ALS data sets for PSI were expected. The ALS data were collected on a crystal that was cryogenically frozen and contained 2.00 M sucrose within the unit cell, whereas the LCLS data were collected at room temperature and had no cryo-protectants. The unit cells had slightly different until cell parameters, and in addition, the very different wavelengths would produce different scattering cross-sections and absorption cross-sections. The LCLS data were not processed to take any absorption into account, although at 1.8 keV, the absorbance correction for the protein crystal and liquid stream may allow for a better comparison between the data sets.

After showing the similarity of the LCLS data and reconstructed PSI electron-density map to the conventional data and resultant electron-density map, it is clear that the LCLS data, to 8.5-Å resolution, is the same—if not better than—the conventional

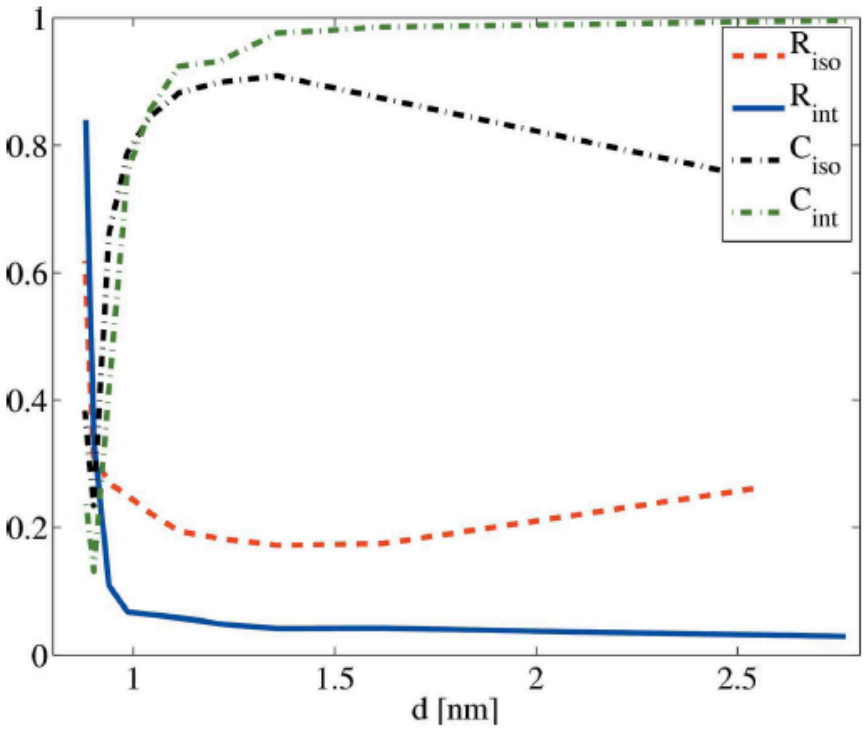

Figure 4.30: **Error metrics for LCLS data of PSI** R factors and linear correlations C (for integration domain δ = 7.3 μm^{-1}) plotted against resolution. Approximately 350 structure factors fall within each resolution shell. The subscript 'iso' refers to a comparison of conventional single-crystal diffraction data to fs nanocrystal diffraction data. The subscript 'int' refers to comparison of even to odd frames from fs nanocrystal diffraction data. Figure and caption taken from (Kirian et al., 2011).

case. That the LCLS data were collected using samples in which the majority of the crystallites could not be seen in a visible microscope is remarkable. This is in stark contrast to the large PSI crystal at a conventional beamline, where crystals larger than 1 mm are used. Yet, the LCLS data is indistinguishable from the conventional data to this resolution, indicating that the femtosecond X-ray protein crystallography could help to solve the structures of difficult-to-crystallize membrane proteins. However, X-ray crystallography is a mature technique, and the data set collected at the LCLS needed to be tested for internal consistency and accuracy in comparison to a "standard" data set in order to fully show that the femtosecond X-ray protein nanocrystallography technique was equivalent to the conventional crystallography.

The diffract-before-destroy principle and radiation damage

The major reason that small crystals of proteins cannot be easily used in conventional X-ray crystallography experiments is that X-ray-induced radiation damage will ultimately limit the exposure time, thereby limiting the true resolution of the structure. A major premise of the femtosecond nanocrystallography project is that the femtosecond X-ray pulses are as fast as, or faster, than the time scale of the conventional X-ray damage processes, such as disulfide bond reduction, decarboxylation of amino acids, etc. Diffraction patterns were collected with X-ray pulse durations of 10 fs, 70 fs, and 250 fs, in order to compare the effects of the different pulse durations on the quality of the diffraction data.

Examining the electron density maps produced using the structure factors from the LCLS data sets, no bulk or long-range damage was evident when using the 70 fs pulse durations. However, the onset of damage, when the secondary electrons were spawned, occurs on the 10-100-fs timescale (Hau-Riege et al., 2004). To assess the damage induced by the XFEL beam, data from 10-fs and 250-fs-duration pulses were collected to directly compare the decay in scattering as a function of time. The 10-fs pulses were operated at a fluence about 10% of the fluence for the 70-fs pulses, corresponding to an absorbed dose of approximately 70 MGy. The integrated structure factors obtained from the three different pulse durations, 10 fs, 70 fs, and 250 fs, using 97,883, 805,311, and 66,063 patterns respectively, and normalized with respect to fluence, are plotted and shown in Fig. 4.31. The plots of the scattering strength of the crystallites versus resolution were generated by selecting and summing Bragg spots from the patterns.

The thickness of the lines in Fig. 4.31 indicates the uncertainty of each plot. The 10- and 70-fs data sets are indistinguishable, indicating that the 70-fs pulse durations were short enough to avoid damage to 8.5-Å resolution, relative to the 10-fs pulses.

The decrease in integrated scattering intensity for the 250-fs pulses, beyond 25-Å resolution, indicates that the 250-fs pulse duration caused significant radiation damage. The data clearly show the onset of global radiation damage below 8.5-Å resolution occurs at time scales greater than 70 fs.

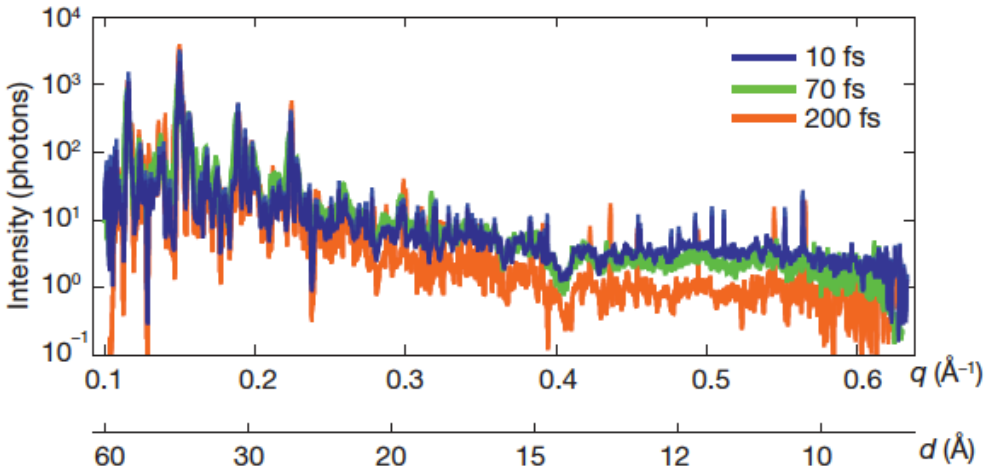

Figure 4.31: **Pulse-duration dependence of diffraction intensities** Plot of the integrated Bragg intensities of Photosystem I nanocrystal diffraction as a function of photon momentum transfer, $q = (4\pi/\lambda)\sin(\theta) = 2\pi/d$, (wavelength, λ; scattering angle, 2θ; resolution, d) for pulse durations of 10, 70 and 200 fs. Averages were obtained by isolating Bragg spots from 97,883, 805,311, and 66,063 patterns, respectively, normalized to pulse fluence. The error in each plot is indicated by the thickness of the line. The decrease in irradiance for 200-fs pulses and $d < 25$ Å indicates radiation damage for these long pulses, which is not apparent for 70-fs pulses and shorter. Figure and caption taken from (Chapman et al., 2011).

Global radiation damage is the result of the radiation physics and chemistry that occur when the sample is inundated with an incident X-ray beam. Global damage is described using metrics such as decreases in the intensity of Bragg reflections (Holton, 2009) and Fig. 4.31 shows a plot of the tradition metric (A plot of I vs. q).

The dose received by the PSI crystallites was calculated to be approximately 700 MGy using RADDOSE (Paithankar et al., 2009) for a fluence of 900 J/cm^2. The dose of 700 MGy is more than a twenty-fold increase over the damage threshold of 30 MGy from conventional X-ray crystallography (Owen et al., 2006). Owen *et al.* determined the $D_{1/2}$ limit of 43 MGy based upon empirical evidence of the totaled

scaled-scattered-intensity ΣI_j over all reflections j, which is a measure of global damage. Although 43 MGy corresponded to $D_{1/2}$, the authors still concluded that the 30 MGy limit was the safer choice to ensure quality data. 30 MGy corresponded to a decrease in total scaled-scattered-intensity by a factor of $\ln(2) \approx 69.3\%$ of the initial total scaled-scattered-intensity in the study (Owen et al., 2006). It should be noted that these damage numbers are for cryogenically cooled protein crystals, and that cryogenic cooling increases the allowed dose by many orders of magnitude (Hope, 1988). The LCLS experiments were carried out at room temperature; therefore, the stability of the crystals at room temperature, despite the increase in the absorbed dose by a factor of 23 when compared to conventional data collection, is even more remarkable.

The dose of 700 MGy in the LCLS experiment corresponds to K-shell ionization in 3% of all carbon atoms present in the sample. The photoionization process can be considered an instantaneous process and can only be outrun with the shortest pulse durations. The loss of the K-shell electrons (whether completely removed from the atom, or in the process of leaving the atom) would requisitely cause a decrease in scattered intensity, as the X-rays are scattered from electrons in the sample. Consequently, the lack of radiation damage for the 70-fs-duration pulses is only relative to the shorter pulses, because a decrease in scattered intensity must occur due to the initial photoionization events, even *if the electron remains in the atom*. However, photoionization of K-shell electrons is a problem that conventional crystallography must also deal with, but relaxation of the system after the initial excitation by the electrons contributes a much larger effect to the damage than the initial photoionization by itself.

The energy absorbed during the X-ray exposure is released through photoionization and Auger decay, followed by a cascade of lower-energy electrons caused by the secondary electrons on the 10-100-fs time scale (Hau-Riege et al., 2004). Additionally, simulations using plasma dynamics indicate that each atom of the PSI crystallite was ionized once during a 100-fs interval after initial exposure and that nuclear motion had

begun on the 100-fs timescale. A photoelectron produced in these experiments from the K-shell ionization of carbon would have a kinetic energy of approximately (1800 eV - 285 eV) 1500 eV, which could produce hundreds to thousands of low-energy electrons. The secondary electrons and low-energy electrons cause the radiation chemistry to initiate, which leads to specific damage at sensitive amino acid residues/sites. However, redox reactions occur in the $\leqslant$ 100-fs timescale and the loss of specific groups from amino acid residues occurs once nuclear motion initiates, which is greater than the 100-fs timescale. It is clearly noted that in conventional crystallography, the specific sample needs to be considered when determining the allowed dose, as the decrease in spot intensity does not give insight into the radiation chemistry occurring. Specific amino acids, such as cysteine, glutamate, and aspartate, are more susceptible to radiation damage than others, and if these radiation-sensitive residues appear in the active site of an enzyme or at the crystal contacts, this could lead to an inability to interpret the electron-density map (Paithankar et al., 2009). Additionally, metal-containing cofactors are acutely prone to local X-ray-induced radiation damage. Using the femtosecond nanocrystallography experiment, consideration of the specific biochemical makeup of the protein will not be necessary, as the experiments are done on a timescale at or faster than the onset of nuclear motion.

The results of the simulations presented in (Neutze et al., 2000) created an expectation that the 70-fs pulse duration data would exhibit some damage effects relative to the 10-fs pulse duration data. The increased time could allow for more Auger relaxation, and secondary ionization events, as well as the Coulomb explosion, which is estimated to begin after just a few tens of femtoseconds (Hau-Riege et al., 2004). There are multiple possible reasons for the lack of visible radiation damage in Fig. 4.31 for the 10-fs and 70-fs data. The most immediate reason is that the displacement of atoms in 70 fs is in the sub-Å range and is not visible at 8.5-Å resolution. data does not attain a high enough resolution for the increased damage to manifest in the intensity

plots of the 70-fs pulses. Another explanation is that the liquid jet surrounding the PSI crystallites acts as a sacrificial tamper for the nanocrystals, in which photoelectrons produced in the stream can neutralize some of the positive charge building up in the sample (Hau-Riege et al., 2010). One additional explanation is that the nominal pulse durations given are from the length of the electron bunch and that the X-ray pulses are shorter in duration than the electron bunches. Indeed, the highly collective nature of the lasing in a free-electron laser allows X-ray pulse shapes that need not match the electron-bunch shapes; simulations of transparency effects in neon using data from the LCLS indicate that nominal 80-fs pulse duration data were more consistent with pulse durations between 20 and 40 fs (Young et al., 2010). Thereby, the lack of damage in the 70-fs data set relative to the 10-fs data set could be attributed to the resolution limit of the data, sacrificial tamper effect of the solvent, or shorter pulses than the nominally reported pulses.

An absorbed dose of 20 MGy will generate a concentration of 12 M in free radicals that cause the radiation damage (Holroyd, 1968). If damage were only a function of fluence and not dose rate (which has been shown in conventional crystallography, but nonlinear effects could become prominent in the higher fields in experiments conducted with XFELs such as that shown in (Young et al., 2010)), the 700 mGy would create a concentration of 420 M in free radicals that are the primary cause of radiation damage. As a result, sample recycling is not possible for femtosecond X-ray protein nanocrystallography experiments.

The LCLS experiments collected crystallographic data on PSI nanocrystals at room temperature. The room temperature exposure creates both benefits and potential problems. There was no need to cryogenically cool the sample, which can damage the crystals (increase in mosaicity), and can be a very time-consuming process as suitable cryo-protectants must be identified and screened. Improper cryo-protectants can lead to the crystal of interest being physically destroyed through cracking, or other

mechanical-strain related problems. The Photosystem I crystals used for conventional crystallography have a complicated freezing procedure in which the crystals must be transferred in small steps into buffers with increasing concentration of sucrose. Finally, the crystals are incubated in 2 M sucrose before being flash-frozen in liquid propane. During the exchange of the sucrose solutions, PSI crystals are often broken or get mechanically attached to the containment vessel and broken upon removal. The PSI nanocrystals prepared for the LCLS experiments did not require any freezing of crystals or handling of dangerous liquid propane, which greatly increased the ease of the sample handling.

Cryogenic cooling of proteins can create complicated changes to the scattered intensity in crystallography experiments. Slow, cryogenic cooling could decrease the temperature factor, potentially leading to stronger scattering at higher resolution than at room temperature. However, more crystals are flash frozen, thereby different protein conformations are locked in time, reducing the potential decrease in the temperature factor. Furthermore, cryogenic cooling can simultaneously lead to an increase in mosaicity, broadening the Bragg peaks over a larger solid angle, which decreases the peak height, as seen in Fig 1.6. Therefore, data collection of PSI crystallites at room temperature is highly favored, although in the case of the large PSI crystals, flash freezing of the crystals has not been shown to appreciably increase the mosiacity spread. Whether the mosaicity of the nanocrystals is really lower than the mosaicity of the large crystals cannot be ascertained yet, but analysis is ongoing. The current resolution does not allow the effects of the temperature factor to be explored, as a Wilson plot is difficult to interpret for low-resolution data.

Shape transform results and discussion

Although the diffraction patterns collected on the front detector were used for the integration of the majority of the data used to reconstruct the PSI electron-density at 8.5-

Å resolution, the back detector contained a wealth of information as well. In at least one regard, the information collected with the back detector was as equally exciting as the front-detector data.

The back detector measured high angular-resolution data and could be used to resolve interference fringes around Bragg reflections that are caused by the small number of unit cells of the PSI crystallites used in the experiment being irradiated by a transversely coherent X-ray beam. The result of the interference effects around the Bragg peaks would be to observe measurable intensities away from the Bragg peaks, which cannot be measured in conventional X-ray protein crystallography experiments using a protein macrocrystal. The "added" intensity contains a large amount of information that is normally lost in conventional crystallography, so the back detector results were greatly anticipated.

Over 1.85 million diffraction patterns were recorded on the back-detector. The back detector recorded diffraction patterns with clearly visible, well-separated shape transforms around the Bragg peaks for the small PSI crystallites. The large working distance of the back detector provided the necessary angular resolution to well-resolve subsidiary maxima, as shown in Fig. 4.32. Thereby, the crystal size could in principle be determined by counting the number of fringes between Bragg peaks (for instance the number of Bragg peaks between the (100) and (200) reflections would give the number of unit cells along $\mathbf{a}^*$). Patterns from crystals containing less than 10 unit cells per edge were recorded, which shows the PSII crystallites were in the size range of 200 nm – 1000 nm, consistent with the serial powder diffraction results shown in Section 4.6. The size range is in agreement with the serial powder diffraction experiments. However, the smallest crystals may not have produced diffraction patterns that were identified by the hit-finding algorithm, as they may produce weaker Bragg reflections. Thereby, patterns from smaller crystals may also be present in the data sets. Re-examining the data to search for smaller crystals is still a possibility.

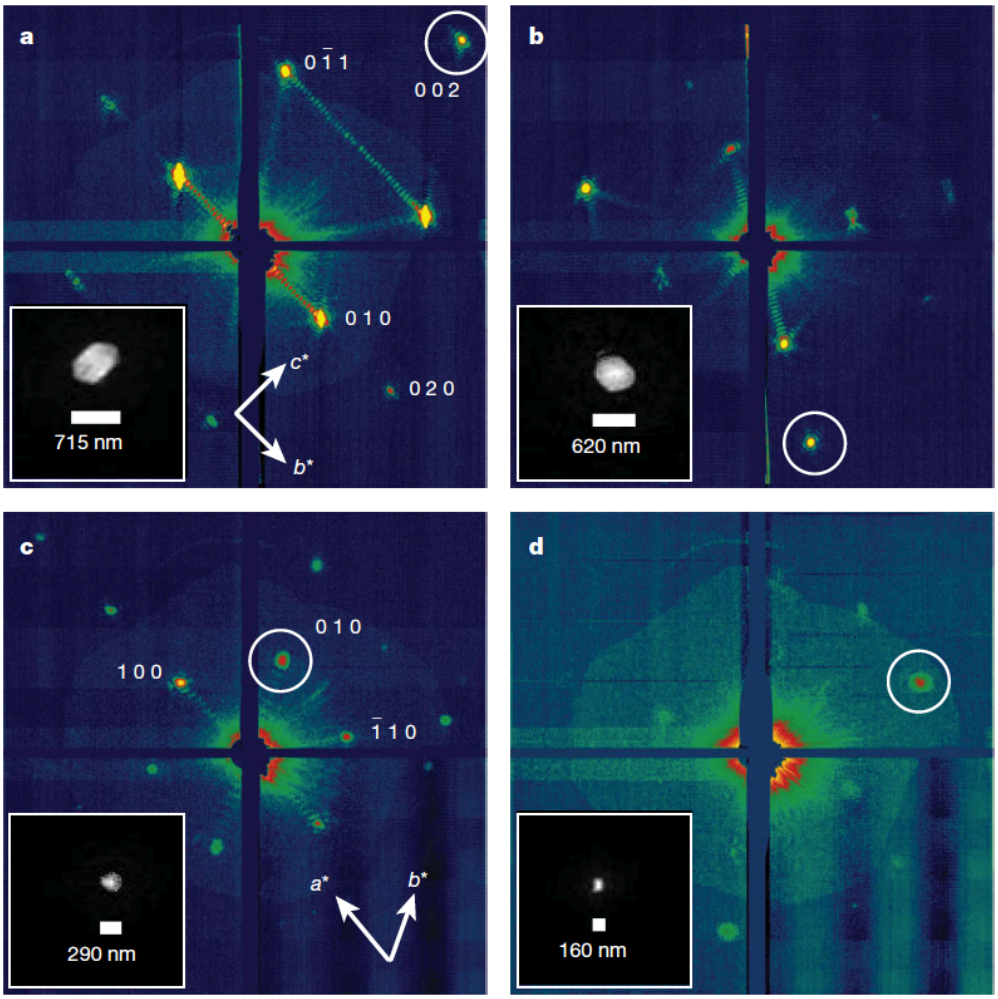

Figure 4.32: **Coherent crystal diffraction** Low-angle diffraction patterns recorded on the rear pnCCDs, revealing coherent diffraction from the structure of the PSI nanocrystals, shown using a logarithmic, false-color scale. The Miller indices of the peaks in (a) were identified from the corresponding high-resolution pattern. In (c) we count seven fringes in the $\mathbf{b}^*$ direction, corresponding to nine unit cells, or 250 nm. Insets, real-space images of the nanocrystal, determined by phase retrieval (using the Shrinkwrap algorithm (Marchesini et al., 2003) of the circled coherent Bragg shape transform. Figure and caption taken from (Chapman et al., 2011).

The interference fringes surrounding the Bragg peaks allow the use of image reconstruction techniques from the coherent diffractive imaging field. Using phase retrieval methods (Marchesini et al., 2003; Robinson and Harder, 2009), a cross section of the crystal size and morphology can be determined from the interference pattern surrounding a Bragg peak, as shown in Figure 4.32a-d (Chapman et al., 2011). The information contained within the shape transform may facilitate a direct solution to the phase problem, according to Shannon's theorem (Sayre, 1952), because the intensity between Bragg peaks allows for oversampling of the diffraction pattern (Spence et al., 2011).

Although the maximum size of the crystals was restricted by inline filtering, the shape transforms were observed because the minimum size could not be strictly chosen, so the minimum size was the soluble PSI_T. The size distribution of the PSI sample was not known in advance; however, the size of the crystal was observed to affect the produced diffraction pattern substantially, as shown in Figure 4.32a-d, with the observation of the shape transforms.

A diffraction grating with N slits will give rise to diffraction features that are finer than the Bragg peaks by a factor of $1/N$. Therefore, a crystal with N unit cells in a given direction will produce diffraction patterns with N - 2 subsidiary maxima between the Bragg peaks, providing a method to determine the size of the crystal projection. The central maxima will have a width of approximately $2/N$, whereas the width of all subsidiary maxima will be $1/N$ for a general parallelepiped crystal. An identical shape transform will be laid down upon each reciprocal lattice site, and the diffracted intensity will have the following form for a general parallelepiped (Woolfson, 1997):

$$I \propto \frac{\sin^2(\pi N_a \mathbf{a}\cdot\mathbf{s})}{\sin^2(\pi \mathbf{a}\cdot\mathbf{s})}\frac{\sin^2(\pi N_b \mathbf{b}\cdot\mathbf{s})}{\sin^2(\pi \mathbf{b}\cdot\mathbf{s})}\frac{\sin^2(\pi N_c \mathbf{c}\cdot\mathbf{s})}{\sin^2(\pi \mathbf{c}\cdot\mathbf{s})} \tag{4.6}$$

where **a**, **b**, and **c** are the unit cell vectors, *i.e.* $\mathbf{a}\cdot\mathbf{s} = \mathrm{h}\cdot\mathrm{x}_j$, N_a represents the number

of unit cells in the **a** direction, etc. The larger the number of unit cells, the smaller the shape transforms until the limit of the conventional crystallography case is reached, in which the inter-Bragg peak intensities are not observable. In the case of a crystal with 10 unit cells in the **a** direction, there will be 8 subsidiary maxima in that direction about the Bragg reflection, and the peak height will be proportional to 10^2, as calculated by Eq. 4.6.

Challenges associated with femtosecond nanocrystallography

Crystallite size distribution

One issue that arose during the femtoseond X-ray protein nanocrystallography experiments was that the PSI nanocrystal and microcrystal size distribution was unknown, which caused some difficulties with adjusting the crystal density for accurate hit-rate calculations. Additionally, knowing the size distribution of the PSI nanocrystals and microcrystals will allow a comparison with the shape-transform data, possibly allowing more information to be gained about which crystals are producing the strongest diffraction patterns, etc.

Knowledge of the size distribution *a priori* would allow the crystal density to be set much more accurately than currently available; therefore, techniques to determine the size distribution of the crystallites will be highly desired for future femtosecond nanocrystallography experiments. Techniques that could be applicable to determining the size distribution of nanocrystals are negative-stain EM, cryo-EM, atomic force microscopy (AFM), and light scattering.

In the case of PSI, the salting-in nature of the crystallization makes negative-stain EM impossible, but the technique could potentially work for crystals that have been salted out; unfortunately, questions about changes to the crystallite size during application of the negative stain (a common stain is ammonium uranyl acetate) will be difficult to answer without the use of an additional technique to determine the crystal-

lite size. Cryo-EM would be a possibility, although this technique was not available before the experiments. However, PSI nanocrystals could not be imaged using cryo-EM, because the evaporation of the solvent before flash freezing would cause the ionic strength of the solution to increase, dissolving the crystals. However, cryo-EM should be applicable to "salting-out" crystals, although large crystals may not be transparent to EM.

AFM has been used to image large protein crystals (Malkin et al., 1995; Malkin and Thorne, 2004), but initial attempts with PSI proved ineffective because the substrates were made of glass. Unfortunately, glass adsorbs many monovalent cations (especially sodium), which causes the PSI crystals to dissolve. Additionally, the PSI crystallites would be moved by the action of the cantilever and tip, indicating that some method of restraining the mobility of the crystallites is needed. In spite of the difficulty with PSI, AFM should be agreeable with most crystals, and functional substrates could be attempted in the future for PSI.

Light scattering should be the most available technique, but cannot discriminate between amorphous aggregates and crystals. However, care should be taken when using light scattering with crystallites being used in optical-pump-probe experiments.

Hit rates and sample settling

When examining the LCLS data for PSI, the low hit rate of 0.004% for the PSI data became a concern. There are a few possible reasons for the low hit rate of patterns with more than 10 identifiable peaks. For one, the crystal density within the solution was set in a coarse manner, using the chlorophyll concentration and assuming an average size for the crystals that was in the midpoint of the size distribution. This would allow for the crystallite density of the PSI crystallites to be coarsely determined; however, the actual crystallite density depends on the size distribution. If the majority of the crystallites were very small, the crystallites would not diffract very strongly, but still contribute

chlorophyll to the crystal number density calculation, providing a larger number of nanocrystals than expected, but with weaker average scattering. Additionally, many diffraction patterns may have been produced from crystallites that were at the "tail" portions of the X-ray spot, *i.e*. only part of the crystal is irradiated by the X-ray pulse, creating a weaker incident intensity and a smaller number of unit cells. However, these two issues were not likely the major issue leading to a lower than expected hit rate.

The major issue for the low hit rate of the LCLS experiments is shown in Fig. 4.33. The plot shows the decrease in sample concentration as a function of time, as measured by absorbance at 280 nm (for peptide bonds) and 680 nm (for chlorophyll *a*). The plot was made by using a HPLC to run a suspension of PSI crystallites through a 15-m long sample line. Settling of the larger crystallites in the sample reservoir lines was likely the major cause of the decreased protein concentration shown in Fig. 4.33.

The crystallite settling that arose in the femtosecond nanocrystallography experiments produced additional deleterious effects. The $\geqslant$ 10-mL sample loops, acting as sample reservoirs, were loaded with several milliliters of re-crystallite suspension. In order to load the sample, the hutch containing the experimental apparatus had to be accessed, which caused the beam to be stopped until all work inside the hutch was completed. Due to the small allotment of beamtime, the number of times the hutch was accessed was minimized. However, due to the settling of the PSI crystallites, large volumes of the sample couldn't be loaded, and the constant sample replenishment led to a loss of precious beamtime. Additionally, the settled crystallites were dried out and could not be recovered, causing a loss of sample that increased the total amount of protein needed.

A few methods could be available to prevent the larger crystallites settling from solution at a faster rate than the smaller crystallites. One method would be to keep the crystallites in a neutral-buoyancy solution, which prevents any settling of the crystallites. In the case of PSI, a neutral buoyancy solution can be prepared using 1.4 M

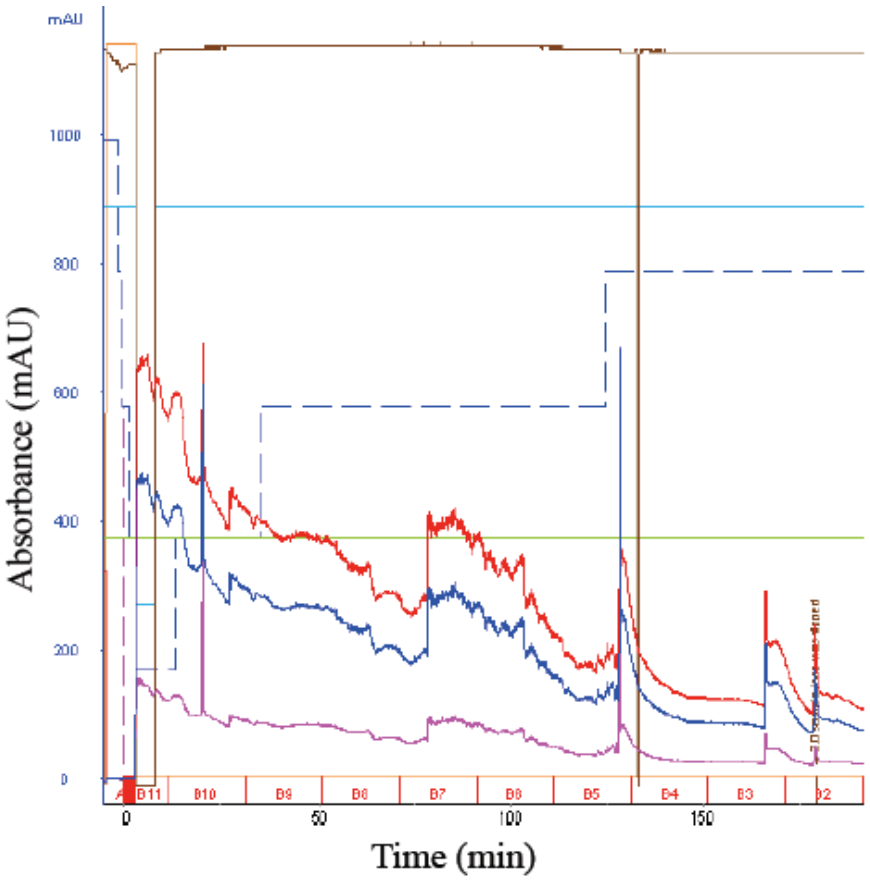

Figure 4.33: **Photosystem I crystallite settling with time** Plot of the absorbance at 680 nm (red), 600 nm (fushia), and 280 nm (blue) of suspension of 2-μm-filtered Photosystem I crystallites as monitored using an HPLC flowing at 10 mL/min. Sudden increases in absorbance at around 75 min and 130 min correspond to the sample loop being inverted.

sucrose in G_0 buffer. A difficulty that arises with the sucrose-containing solution is that the viscosity increases with the amount of sucrose, and consequently, the liquid pressure in the jet needs to be increased. This can lead to an unstable jet or even a jet that will not run with a given inner-diameter fiber optic for the jet. Consequently, the inner diameter of the fiber optic would need to be increased, making for a thicker, higher flow-rate jet. The thickness of the jet is a problem when higher-energy X-rays are used, as a strong solvent background will be recorded, which may reduce the dynamic range, or worse, fully saturate the detector pixels. One possibility to overcome the difficulties listed above is to use lower sucrose concentrations. The lower sucrose concentrations

will allow for a decrease in viscosity while still increasing the settling time of the larger crystallites.

Another method to avoid the settling of the larger crystallites from solution would be to invert the sample reservoir (and preferentially the sample transport lines to the injector). The gas-pressure driven injector would need to be modified by using a syringe pump to drive the liquid flow (inverting the reservoir with a gas line at one end would simply cause the liquid to flow to the opposite side of the reservoir). A setup allowing for the sample reservoir to be inverted was developed by Dr. Robert Shoeman of MPI Heidelberg and was tested in the Feb 2011 femtosecond X-ray protein nanocrystallography experiments. However, difficulties have arisen when trying to operate the jet with a syringe pump, and therefore, detailed testing of a syringe-based system would need to be done before it became a viable option to prevent the settling problem. However, experiments utilizing an anti-settling device were done during the higher-energy LCLS experiments, discussed in Section 4.11.

PSI sample consumption during femtosecond nanocrystallography

An issue that arose during the first femtosecond nanocrystallography experiments related to the low hit rate was that a large amount of protein was used in the experiments, as the PSI data were collected for 24 hours at a flow rate of 10 μL/min. Approximately 14 mg of PSI was used for the first femtosecond X-ray protein nanocrystallography experiments in Dec 2009. Operating at 30 Hz, the X-rays only interacted with 0.004% of the sample, hitting one out of every 25,000 nanocrystals at a concentration of one crystallite per four interaction volumes. The efficiency will be increased as the LCLS moves to 120 Hz, but this is only an improvement by a factor of four in the efficiency of sample use. Once the European XFEL becomes operational, with tens of kilohertz repetition rate, the efficiency will be increased by orders of magnitude, so that most nanocrystals within a sample interact with the XFEL beam (Altarelli, 2010). Using 12

kHz with the same setup, one out of every 250 crystallites would be utilized.

Even though the data collection used large amounts of protein samples, the process may not be very "wasteful" in comparison to the conventional crystallographic experiment. In order to show the relative efficiencies of protein use between the conventional experiments and the LCLS experiments, the amount of protein in each crystal, and the number of crystals used to collect a data set in each case, must be compared.

There are approximately 10^7 to 10^8 crystallites with 2-μm edges within a 0.5 mm$x$ 0.5 mm x 2 mm PSI crystal. Conservatively, with a hit rate of one in 25,000 (the hit rate for the first LCLS experiments), $5x10^7$ 2-μm PSI crystallites would produce 2000 diffraction patterns, whereas approximately 10,000 patterns were needed for the completeness of the LCLS data to approach 100%, without any type of post refinement—post-refinement may significantly decrease the number of patterns necessary for completion, to approximately 2000 patterns—and pure Monte Carlo integration of the intensities (Kirian et al., 2011). Once the LCLS operates at 120 Hz, $5x10^7$ 2-μm PSI crystallites will produce 8000 diffraction patterns, generating the necessary amount of diffraction patterns for a full Monte-Carlo integration of the intensities.

When conventional X-ray crystallography experiments are done using large PSI crystals, up to 100 large crystals—that each contain 10_7–10_8 of the 2-μm PSI microcrystals—are grown and shipped to the experiments. Therefore, the total amount of PSI used for the preliminary LCLS experiments, once scaled for the full repetition rate of the LCLS, is less than the amount of PSI used data collection on single, large crystals, and the efficiency of the femtosecond X-ray protein nanocrystallography technique will be increased by at least two orders of magnitude when the European XFEL becomes operational. Additionally, the conventional crystallization experiments will produce PSI crystals that are too small to attain diffraction beyond 2.5-Å resolution, as well as PSI crystals that grow with large internal disorder (mosaicity $\geqslant 1.0°$). As a consequence, only 1-2% of all protein in a conventional PSI crystallization experiment

produces crystals that have data sets recorded. As discussed in 1.3, the optimization of the PSI crystal growth took 13 years, and consequently, the fact that the femtosecond X-ray protein nanocrystallography experiments were successful on the first attempt should not be overlooked due to protein consumption.

One advantage of Photosystem I in the LCLS experiments is the high natural abundance of the protein. The large amount of PSI available can realistically reflect the case for membrane proteins produced through recombinant expression, as $\sim$0.8 mg/L of protein in the *T. elongatus* cultures, which is comparable to the yields of the recombinant expression systems of membrane proteins, and between 10 and 100 times less than the yields for over-expression of soluble proteins, where up to 100 mg/L can be achieved. Furthermore, as the production of a large, well ordered protein crystal is a challenging task, a large amount of protein is utilized in the exploration of suitable crystallization conditions and optimization of crystal quality and growth. If femtosecond nanocrystallography can use microcrystals for structure determination that are so small the nanocrystals cannot be seen using an optical microscope, attention could be turned to producing larger numbers of nanocrystals and microcrystals with the available resources, as opposed to using the resources in the optimization of the growth of large, well-ordered protein crystal.

Decreasing sample consumption is still a major optimization factor for the femtosecond nanocrystallography project. One prominent method to decrease the sample usage would be to decrease the flow rate of the jet. As the jet currently runs at approximately 10 μL/min, slowing the flow rate to a few hundreds of nanoliters per minute would offer a substantial improvement. However, the minimal flow rate is limited by the repetition rate of the X-rays. The linear velocity of the jet, which is related to the flow rate and jet diameter, must be large enough to deliver fresh sample to the X-ray interaction region that could not have received any ill effects of the previous pulse before the subsequent X-ray pulse. In the case of 120 Hz repetition rate, the lower limit of the

flow rate will be ∼10 nL/min, but the exact number will depend on the jet thickness. In the case of high-viscosity anti-settling liquid jets utilizing high sucrose or PEG concentrations, a decrease in flow rate to ∼ 100 nL/min has been observed, and is being further developed.

Resolution limits of the first femtosecond nanocrystallography data

The initial femtosecond nanocrystallography experiments were conducted at 1.8 keV X-rays (λ = 6.9 Å) as determined by the initial operating conditions of the LCLS. The resolution limit of approximately 8.5 Å in the PSI femtosecond nanocrystallography data set was due to the long wavelength used in the experiment, as well as the instrumental geometry (less of an effect than the wavelength, but it still restricted the resolution).

The resolution limit of the data set, to 8.5 Å, also contributed to the small number of peaks in the patterns. The indexing success of MOSFLM or DirAx was correlated to the number of peaks contained within the diffraction pattern. The indexing success being correlated to the number of peaks supports that moving to higher resolution will increase indexing success because the total number of measurable reflections in a given resolution sphere is calculated as follows (Rupp, 2010):

$$N_{\text{reflections}} = V_{\text{unit cell}} \cdot V_{\text{reciprocal sphere}} \tag{4.7}$$

where $V_{\text{reciprocal sphere}}$ is calculated as follows (Rupp, 2010):

$$V_{\text{reciprocal sphere}} = \frac{4}{3}\pi\left(\frac{1}{d_{\text{min}}}\right)^3 \tag{4.8}$$

Using equations 4.7 and 4.8, the total number of reflections to 8.5 Å for the PSI structure is calculated to be approximately 82000 non-unique reflections corresponding to 6800 unique reflections. However, the number of reflections in the 3.0-Å-resolution

sphere is calculated to be 1.9 million non-unique reflections, corresponding to 158,000 unique reflections. Additionally, the size of the PSI crystallites will need to be increased to compensate for the reduced scattering cross-section at higher X-ray energies. This would decrease the measured intensities of the shape transforms, providing another benefit to auto-indexing once higher energy X-rays are utilized. In order to approximate the number of unique reflections, the number of total reflections must be divided by a number that takes into account the number of symmetry operations and can be found in tables containing the Laue groups, and in the case of P_{6_3}, the number of unique reflections is divided by 12.

General discussion

Although many more experiments are necessary to firmly establish the viability of the femtosecond nanocrystallography experiment, the initial results presented in this section indicate the great potential of the technique for structural biology applications. The results shown to this point indicate that not only do very small crystals of membrane proteins exist, but that the small crystals can be used to determine structures to low resolution. However, the strength of the scattering from the PSI nanocrystals indicates that the data extends to higher resolution than recorded, and therefore, the question moving forward is to what resolution can given nanocrystals scatter, *i.e.* will nanocrystals be able to produce diffraction to sufficient resolution to produce molecular- or atomic-resolution electron-density maps?

4.8 PSI-Fd femtosecond nanocrystallography pump-probe results

Motivation for the PSI-Fd pump-probe experiments

Time-resolved X-ray crystallography experiments are used to monitor the evolution of a process with time at a high spatial resolution, with one notable example being the dissociation of carbon monoxide from myoglobin (Teng et al., 1997). Time-resolved crystallography at third-generation synchrotrons can achieve temporal resolution of tens of

nanoseconds to microseconds. The success of the femtosecond X-ray protein nanocrystallography experiments (Chapman et al., 2011) made the extension of time-resolved X-ray crystallography to XFELs highly desirable. At XFEL sources, the temporal resolution could theoretically be in the hundreds of femtoseconds to microseconds, set only by the time delay between the excitation and probe.

One biological system of much interest that would provide light-induced structural changes is the photoinduced charge separation that leads to the reduction of ferredoxin in the PSI-Fd complex. Flash spectroscopy has been used to study the kinetics of the reduction of ferredoxin by PSI (Setif and Bottin, 1994, 1995). The spectroscopic data is interpreted such that the fast first-order phase is the reduction of ferredoxin docked in close proximity to F_A^- or F_B^-, whereas the intermediate first-order phase may correspond to a different, more distal, binding site for ferredoxin. The slowest phase is dependent on the concentration of ferredoxin and PSI, and is thereby attributed to the collision of undocked ferredoxin with PSI (Setif and Bottin, 1995).

Time-resolved X-ray crystallography experiments were designed with the goal to unravel the conformational changes during the light-induced undocking of ferredoxin from PSI; the changes could be monitored using pump-probe experiments at the LCLS. The experiments were conducted in June 2010 during LCLS beamtime. The PSI-Fd crystals contain the photo-active complex of PSI with its natural electron acceptor Ferredoxin (Fromme et al., 2002). Upon light exposure, PSI performs light-induced charge separation whereby electrons are transferred from PSI to Fd. Reduced Fd undocks from its binding site at the stromal hump of PSI. The crystals of the PSI-Fd complex dissolve when exposed to light, indicating that electron transfer from PSI to ferredoxin occurs, and that the ferredoxin is situated at a crystal contact of the crystalline system. The undocking is irreversible, and the irreversible undocking process within the crystals of the PSI-Fd complex, as well as the time-resolution of the LCLS pulses, made femtosecond nanocrystallography an ideal technique to study the undock-

ing process at high spatio-temporal resolution.

Acknowledgments regarding the PSI-Fd time-resolved nanocrystallography

My role was to find the necessary conditions to produce nanocrystal and microcrystal samples of the complex of PSI-Fd, produce the sample, work on the design of the experiments, design and assist with the sample delivery and data collection of the X-ray diffraction snapshots. The majority of the analysis of the PSI data was done by Andrew Aquila and Anton Barty of the Center for Free Electron Laser science at DESY labs.

The experiments were done at the AMO beamline (Bozek, 2009) of the LCLS, and I would like to thank the hard work of the entire LCLS team. The experiments were completed in the CAMP instrument (Struder et al., 2010), and I would like to thank the entire CAMP team for their hard work.

PSI-Fd pump-probe experimental details

Pump-probe experiments were performed at the AMO beamline of the LCLS using an analogous experimental setup as described in (Chapman et al., 2011). The changes to the experiments in June 2010 are as follows: the LCLS was operated at 60 Hz, 2 keV X-rays were used, the fluence was 10x larger, and only the front detector was utilized; due to the parallel orientation of the optical laser and X-ray beam, a "light catcher" was installed between the two panels of the front detector, and the back detector was removed.

The samples were delivered to the X-ray interaction region by a newly re-designed injector—designed and assembled at Arizona State University by R. Bruce Doak and Uwe Weiestall—that contained multiple improvements over the original injector design, as shown in Fig 3.11. A gate valve was incorporated in the new injector to allow the exchange of a nozzle without needing to bring the main chamber to atmospheric pressure, a microscope was installed so that the liquid jet could be visibly

monitored within the control room of the AMO beamline, and a multi-mode fiber optic allowed for the introduction of the optical pump laser to the optical-pump-laser interaction volume. Nanosecond pulses from a frequency-doubled Nd:YAG laser, operating at a wavelength of 532 nm, were delivered through the multi-channel fiber optic that was led through the improved version of the injector to allow the laser light to inundate the liquid jet in the chamber, as shown in Fig. 3.12.

The PSI-Fd samples were made using batch crystallization experiments in glass capillaries and eppendorf tubes, as described in Section 3.3, at Arizona State University and onsite at the LCLS. The PSI-Fd samples were stored in an opaque box that was wrapped in two layers of aluminum foil and kept in a drawer at room temperature. The nanocrystals of the PSI-Fd complex were inline filtered with 2-μm and 5-μm filters, and diffraction patterns were recorded after delays of -10, 5, and 10 μs. The -10-μs data sets were used as the dark data sets, as the visible laser light arrives after the X-ray pulse.

PSI-Fd nanocrystallography results

In excess of 5000 diffraction patterns from nanocrystals of the PSI-Fd complex have been collected, and Bragg peaks were recorded to 10-Å resolution. Only two hours of data were collected for the PSI-Fd samples due to experimental problems that reduced the total amount of beamtime, such as the injector clogging, as discussed in Section 4.8. The majority of the data collected for nanocrystals of the PSI-Fd complex was from nanocrystals of the PSI-Fd complex grown using 22.5% PEG400, 250 mM $CaCl_2$, 250 mM HEPES pH 7.5, and 0.013% β-DDM. The PSI-Fd samples that were used for the experiments were grown onsite at the LCLS because the original crystals of the PSI-Fd complex, which had been pre-characterized by SONICC, were dissolved due to the temperature dropping to $\leqslant 18$°C in the biochemistry lab at the LCLS, where the crystals had been stored.

Approximately 10% of the PSI-Fd patterns from the -10, 5, and 10-μs-delay data sets can be indexed with unit cell constants within 5% of the values reported for the published unit cell (Fromme et al., 2003). The 10-μs-delay data sets had an indexing success of approximately 6%. Three reasons can possibly explain the low indexing success of the three data sets: the small amount of peaks in the patterns due to the limited resolution, too high of a hit rate, as well as polycrystalline samples. The first possibility has been addressed already in Section 4.7, and the second possibility would necessitate a careful examination of the concentrations used in future experiments. However, the third possibility was not encountered when using the PSI crystallites. An examination of the crystals highlighted by the arrow in Fig. 4.18c shows the presence of a series of crystals that are seemingly connected at the corner positions. There are multiple cases of the strung-together crystals, and this could have produced readouts with diffraction patterns from multiple nanocrystals of the PSI-Fd complex, causing the auto-indexing software difficulties. However, if the diffraction pattern from one of the nanocrystals of the PSI-Fd complex is significantly stronger, the auto-indexing software may still be able to index the pattern. The auto-indexing software could be manually changed to choose one lattice, and to exclude the second lattice from the Monte Carlo indexing.

A Wilson plot showing the intensities of the data with respect to resolution is shown in Fig. 4.34 for the three different time points. Although some differences are noticed in the 10-μs-delay data, the error of the plots makes interpretation difficult. A promising feature that was found in the data is that the unit cell parameter in the **b**-direction decreased.

The decrease in indexing success for the 10-μs-delay data sets could be attributed to a change of unit cell constants that may be directly linked to the undocking of the Ferredoxin from the PSI. Fig. 4.35 shows the distribution of the magnitudes of the **b** unit-cell vector, as determined by indexing the PSI-Fd patterns for the three different data sets. The plot clearly shows that the magnitude of the **b**-vector increases after

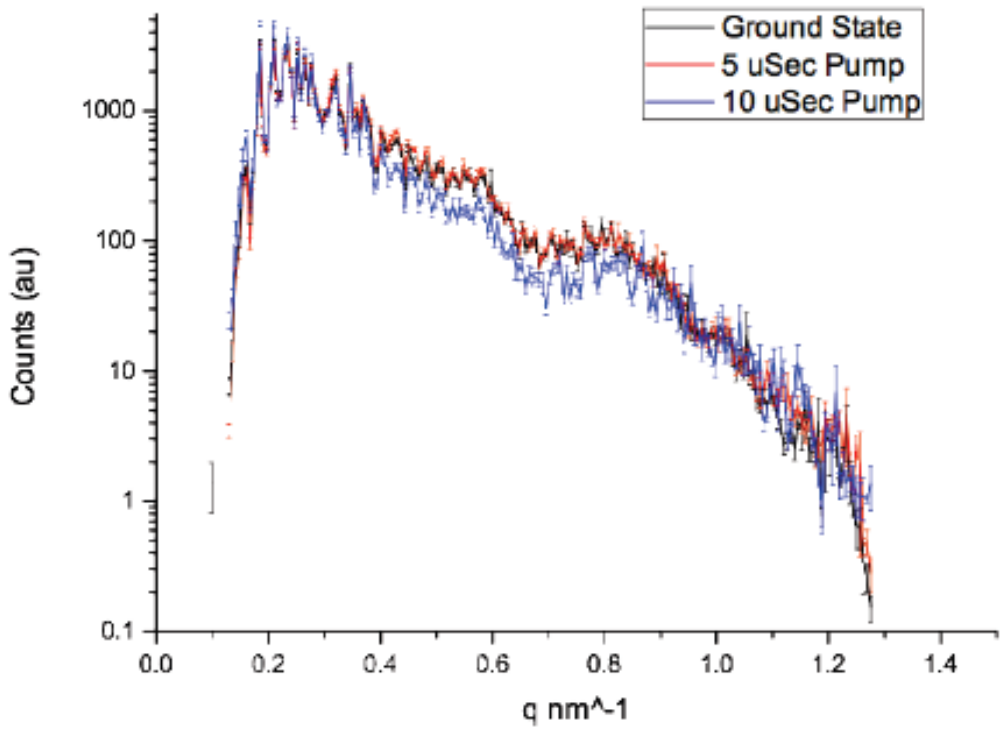

Figure 4.34: **Wilson plot of the PSI-Fd pump-probe data** Wilson plot showing the diffracted intensity as a function of q. Plot made by Dr. Andrew Aquila of the Center for Free Electron Laser studies at DESY.

the optical pump laser and continues to increase in magnitude when the time delay between the optical laser pulse and the X-ray probe laser is increased from 5 μs to 10μs. Although the location of the Ferrdoxin is not known for this space group (Fromme et al., 2002), the increase in the magnitude of the **b** unit-cell vector may be indicative of undocking of Ferredoxin from PSI. The results suggest that molecular motion within the crystal may be occurring, but due to so many variables, drawing a strong conclusion has been difficult to date. Analysis is ongoing, and higher resolution experiments on PSI-Fd co-crystals are already scheduled in Aug. 2011 at the CXI beamline of the LCLS.

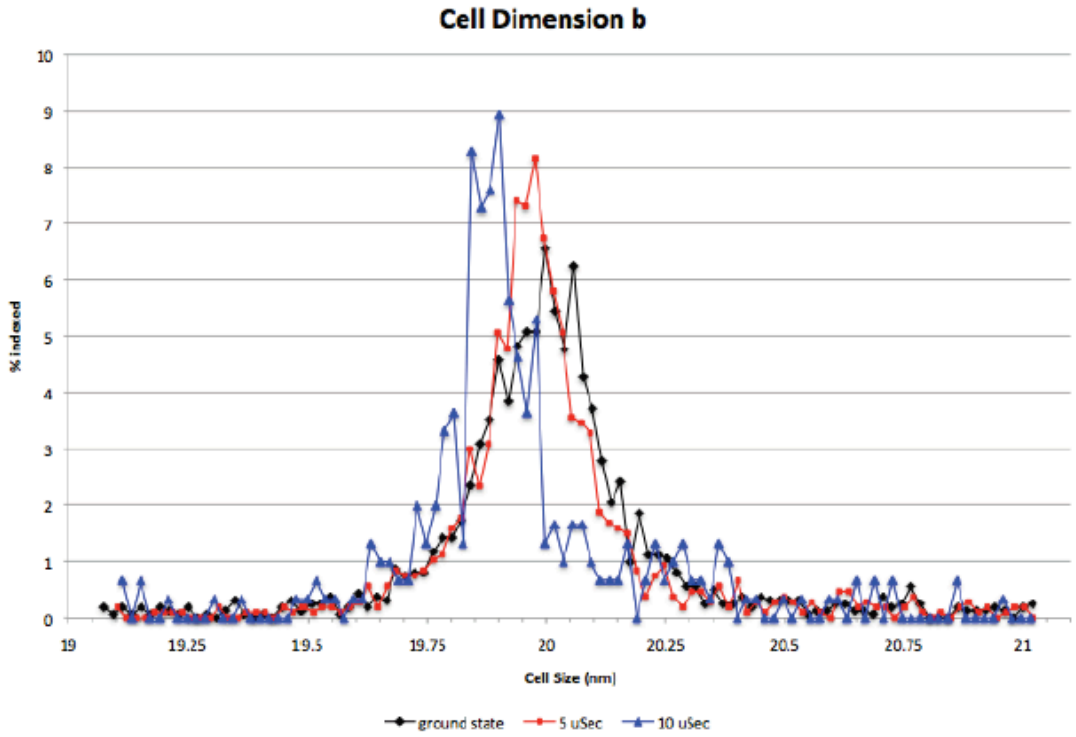

Figure 4.35: **PSI-Fd b axis determined from indexing** Plot showing the calculated unit cell parameters determined from indexed LCLS diffraction patterns of crystals of the PSI-Fd complex. The black, red, and blue curves show the unit-cell parameter results for diffraction patterns generated during PSI-Fd experiments with delays of -10 μs, 5 μs, and 10 μs, respectively. The unit cell parameter in the **b** direction changes once the sample is irradiated with the pump laser. Plot made by Dr. Andrew Aquila of the Center for Free Electron Laser studies at DESY.

Challenges of the PSI-Fd pump-probe experiments

A major difficulty of the pump-probe experiments on the nanocrystals of the PSI-Fd complex was that the temperature of the room used for storage of the nanocrystals dropped below 18°C, and the pre-made nanocrystals were destroyed by a phase transition of the PEG400 in the solution. Therefore, the nanocrystals of the PSI-Fd complex had to be grown on site during the off shifts of the beamtime, and were only allowed to equilibrate for less than 12 hours (often 6 hours) before being directly used in the time-resolved femtosecond nanocrystallography experiments. The short equilibration time could have led to smaller crystals that were poorly formed; the freshly made sam-

ples used in the experiments could not be tested with SONICC or any other technique prior to use. Additionally, all remaining ferredoxin was used in the attempt to make enough nanocrystals of the PSI-Fd complex for the beamtime, and consequently, the crystallization experiments could not be repeated after the beamtime in order to test the samples.

One additional difficulty that arose during the experiments was that the nozzle clogged frequently by other samples (not PSI or PSII) and the shroud became overfilled with the sponge-phase crystallization matrix of a different sample used during the experimental beamtime. The sponge phase, which did not sublime after freezing onto the back wall of the shroud, reduced the total volume of the "catcher" and facilitated the growth of icicles from the back of the shroud to the nozzle. The growing icicles caused a disturbance to the laminar flow of the accelerating gas of the liquid injector, causing the stability of the jet to be decreased. The position of the jet became a function of time, and the jet started flickering, until the icicle grew to the nozzle and froze the entire nozzle. In order to compensate in future experiments, a larger sample shroud may be designed and utilized.

New standard operating procedures need to be developed in preparation for future pump-probe experiments of visible-light driven photochemistry. The temperature in the LCLS hall was not stable, and this would add an additional variable to describing and analyzing the results.

General discussion of the PSI-Fd experiments

Although few data were collected on the nanocrystals of the PSI-Fd complex, the major experimental goal was achieved. The femtosecond X-ray protein nanocrystallography technique, as adapted for pump-probe experiments at the LCLS on systems that allow detection of light-induced changes, was established, the first data sets were collected, and are currently under analysis. The setup still needs optimization, as the injector

had difficulties with the sponge phase, but the framework is in place for time-resolved X-ray crystallography experiments at the LCLS and other higher-energy XFELs.

The experimental setup had several notable accomplishments. Laser excitation of the sample, through an externally operated nanosecond optical laser, was achieved and the microscope of the improved injector allowed for direct visualization of the positioning of the pump laser and the performance of the liquid jet. The optical pump laser was fully synchronized with the X-ray probe laser, making time delays in the range of nanoseconds to microseconds possible. The microscope of the improved injector allowed for the thermal radiation of the plasma, generated by the interaction of the X-rays with the liquid sample, to be visualized and allowed alignment of the jet without the use of a back detector. The Wilson plot of the three different time-delay data sets as well as the variable **b** vector indicate that light-induced changes in the PSI-Fd complex can be detected.

Time-resolved X-ray crystallography has been established for use at XFEL sources by adapting the femtosecond X-ray protein nanocrystallography technique for pump-probe experiments. As the higher-energy X-rays become available for time-resolved X-ray crystallographic experiments, the high temporal-spatial resolution offered by this technique will offer new insights into the dynamics of biomacromolecules that could never before be seen.

4.9 Femtosecond nanocrystallography of PSII

Motivation of the PSII femtosecond nanocrystallography work

Phototsystem II is and has been one of the most studied proteins due to its responsibility for the oxygenic state of the atmosphere and the higher life as seen on the earth today. PSII catalyzes the thermodynamically unfavorable water-splitting reaction, in which two molecules of water are converted to a molecule of oxygen, four protons, and four electrons, using light energy from the sun while transitioning between the S-states

(S_{0-4}), as discussed in Section 1.2.

PSII have been studied extensively using X-ray protein crystallography so that an electron-density map could be determined that would allow for an understanding of the fundamental process of water splitting that occurs in the oxygen-evolving complex of the protein. The structure determination experiments to date are done with the oxygen-evolving complex in the S_1 state. The first high-resolution structure was published at 3.8-Å resolution in 2001 (Zouni et al., 2001), followed by a 3.0-Å structure published in 2005. Although the resolution of the vast majority of the protein subunits and cofactors increased with the resolution of the data, the catalytic center of the oxygen-evolving complex, the Mn_4CaCl, remained "fuzzy" and unresolved. Although initially not entirely understood, the reason for the invariant response of the electron density of the Mn_4CaCl cluster with respect to resolution in the data was suspected to be X-ray-induced radiation damage. In 2005, this suspicion was confirmed by (Yano et al., 2005), when X-ray absorbance spectroscopy was used to determine that a redox reaction occurs at the Mn_4CaCl cluster at an X-ray dose that is over an order of magnitude lower than the doses used before deterioration of the diffraction pattern of PSII crystals is measured. However, in 2011, a 1.9-Å resolution structure of PSII was published that finally showed well-defined electron density in the Mn_4CaCl cluster. Although the electron-density surrounding the Mn_4CaCl cluster is well-resolved, the occupancy of the Mn sites is not discussed in the paper, the X-ray doses are not tabulated, and the Mn-Mn distances still do not agree with EXAFS studies (Pushkar et al., 2008), allowing questions of X-ray-induced radiation damage to linger (although it is clear that the structure has much reduced damage with respect to its predecessors).

However, it is not only the ground state, S_1, that is of interest to the structural biology community. As the S-cycle consists of multiple steps, S_{0-4}, the structural changes that facilitate the catalytic activity of the Mn_4CaCl are of keen interest. Unfortunately, the fast time scales involved in the S-cycle do not allow for conventional

time-resolved crystallographic measurements to be useful in determining the electron-density of any state but the S_1 state. Consequently, the S-cycle and the "ground state" structure of the Mn_4CaCl structure have always been studied using separate techniques; the high temporal-resolution studies accomplished using ultrafast spectroscopy and X-ray absorption spectroscopy, while the high spatial-resolution experiments accomplished by X-ray protein crystallography, but only on the S_1 state.

The successful proof-of-principle experiments of femtosecond X-ray protein nanocrystallography, Section 4.7, has shown the potential of this new style of X-ray crystallography. However, one of the most exciting aspects of the new femtosecond X-ray protein nanocrystallography technique is the use of X-rays produced by an XFEL; the short pulse durations can allow high temporal-resolution in addition to the high spatial-resolution in the electron-density map. Effectively, the femtosecond X-ray protein nanocrystallography technique allows the power of ultrafast spectroscopy for probing matter with high temporal resolution to be combined with the power of X-ray protein crystallography for probing matter with high temporal resolution. Therefore, the femtosecond X-ray protein nanocrystallography technique could be used to perform a completely new type of time-resolved X-ray crystallography experiment, with time scales short enough to reach the limit of atomic and molecular motion itself.

As a consequence, time-resolved femtosecond X-ray protein nanocrystallography experiments were attempted on PSII following the proof-of-principle experiments done using crystals of the PSI-Fd complex. The experiments were designed as a proof-of-principle for using femtosecond X-ray protein nanocrystallography on PSII, in anticipation of the higher-energy LCLS X-rays available approximately one year later. The goals of the experiments were to collect damage-free diffraction patterns of PSII in the S_1 state (and produce an electron-density map), as well as to collect diffraction data from PSII crystallites that had been excited by an optical laser, in order to excite the Mn_4CaCl cluster to the S_2 state. Therefore, the experiments were using the LCLS

to study PSII in both the S_1 and S_2 states. The maximum resolution attainable for these experiments was $\sim$7.3 Å.

Acknowledgments regarding the PSII time-resolved nanocrystallography

My role was to characterize the nanocrystal and microcrystal samples of PSII, work on the design of the experiments, design and assist with the sample delivery, and collect the X-ray-diffraction data. The majority of the analysis of the PSII data was done by Tom White of the Center for Free Electron Laser science at DESY labs. Raimund Fromme and Tom White were involved in the production and refinement of the PSII electron-density maps.

The experiments were done at the AMO beamline (Bozek, 2009) of the LCLS, and I would like to thank the hard work of the entire LCLS team. The experiments were completed in the CAMP instrument (Struder et al., 2010), and I would like to thank the entire CAMP team for their hard work.

PSII time-resolved femtosecond nanocrystallography experimental overview

The time-solved femtosecond nanocrystallography experiments using PSII were carried out using an analogous setup, to the setup described in Section 4.8, using the femtosecond X-ray protein nanocrystallography technique (Chapman et al., 2011). Briefly, the experiments were done at the AMO beamline (Bozek, 2009) of the LCLS (Arthur et al., 1995) in the CAMP instrument (Struder et al., 2010) using 2000 eV X-rays (λ = 6 Å). The repetition rate was 60 Hz, and the sample was irradiated by a 532-nm frequency-doubled Nd:YAG laser. The experiments were done with time delays of -10 μs and 5 μs between the optical pump laser and the X-ray probe laser, and the maximum resolution in the corner of the detector was 7.3 Å.

The PSII crystallites were made 11 days prior to the LCLS beamtime by crystallization through the dropwise addition of a buffer containing 11.0% (m/v) PEG2000, as

detailed in Section 4.5. The PSII nanocrystals and microcrystals had been confirmed as crystalline through SONICC measurements, as shown in Fig. 4.22. After the dropwise addition of the PEG2000-containing solution, the sample was allowed to incubate in a falcon tube that had been wrapped in aluminum foil at 4°C until the LCLS experiments.

The PSII sample was prepared by pre-filtering the re-suspended PSII crystallites with a 5-μm inline filter. The PSII crystallites were stored in opaque boxes at 4°C, and the crystallite suspension was loaded into the sample lines in a room that was completely dark, save for two green LEDs that were used to indirectly illuminate the sample. The data collection was done with all lights covered and the main lights of the LCLS hutch turned off.

PSII femtosecond nanocrystallography results

Almost 10,000 diffraction patterns of PSII in the dark state, using a -10-μs delay between the optical pump laser and the X-ray probe laser, were collected; in addition, a few hundred patterns were collected for PSII nanocrystals and microcrystals that had been excited with the optical laser. The PSII crystallites diffracted very well, with data collected to the corners of the detector at a resolution of 7.3 Å. The Bragg reflections were much stronger than the comparable PSI crystallites of the same (maximum) size, and this is due to the smaller unit-cell volume, the lower solvent content, and the $P_{2_12_12_1}$ space group.

A representative PSII diffraction pattern from the LCLS experiments, as well as the indexing results on the pattern, are shown in Figs 4.36A and B, along with a plot of the extracted structure factors in Fig 4.36c. An 8-Å electron-density map generated using the LCLS structure factors is shown in Fig. 4.37, but data analysis is ongoing, especially in regard to the comparison of the light and dark data sets. The electron-density map shows the transmembrane helices, as well as the luminal hump of the PSII. Although the resolution remains too low to determine whether the structure

shows the undamaged PSII, many striking features of the electron density are visible. The electron density shown in Fig. 4.37 is excellent for 8-Å resolution, as the trans-membrane helices and alpha helices of the oxygen-evolving complex are clearly visible, and to better detail than an conventional X-ray crystallography map of commensurate resolution.

A remarkable feature of the PSII data sets was that PSII crystallites, which had been made 11 days before the experiments, diffracted strongly and showed no degradation. From the large, single-crystal work on PSII, it was established that the PSII crystals used for data collection needed to be frozen directly after growth, or the resolution of the diffraction patterns would progressively decrease; the PSII crystals grown for the conventional case are frozen at most two days after crystal growth.

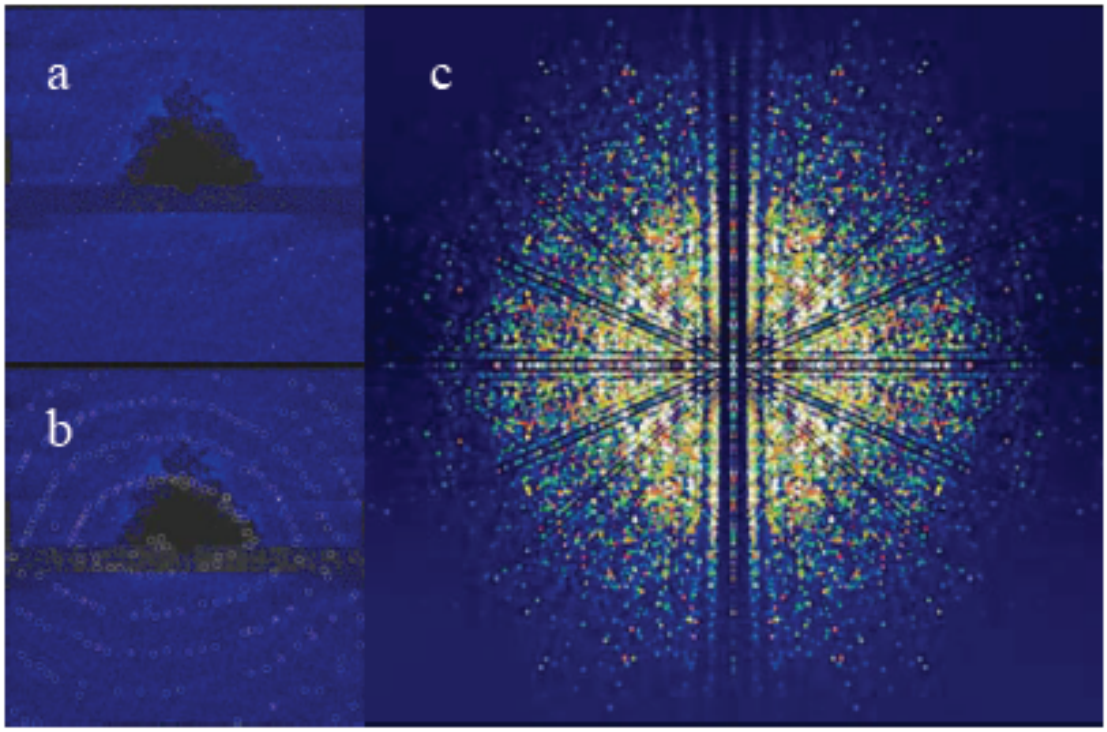

Figure 4.36: **PSII diffraction patterns and structure factors from the LCLS** (a) A recorded diffraction pattern of photosystem II using 2 keV X-rays of the LCLS and 70-fs pulse duration X-rays in June 2011 (b) the indexing result of the pattern in (a) using auto-indexing programs. (c) projection of the three-dimensional structure factors extracted from the "dark" LCLS data sets of PSII collected using 70-fs pulse duration X-rays from the LCLS.

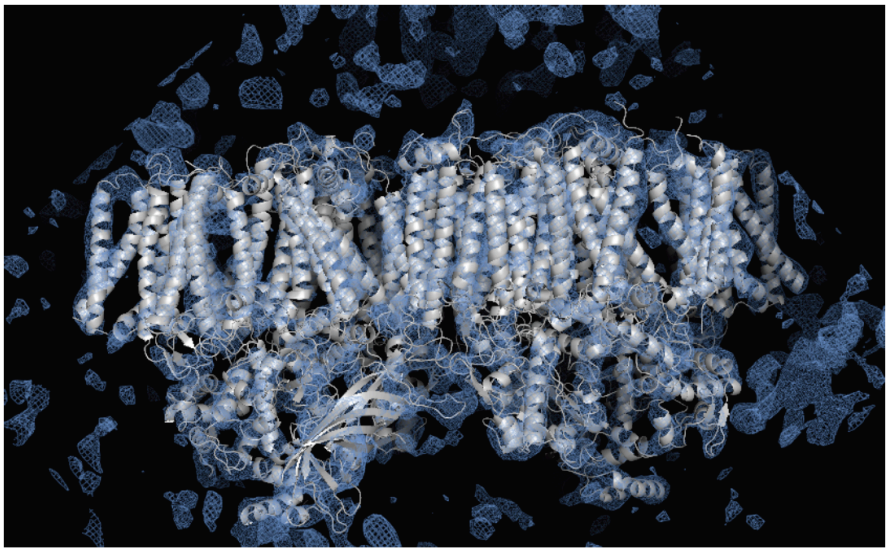

Figure 4.37: **Electron-density map of PSII from the LCLS data** 8-Å electron-density map of photosystem II calculated from the LCLS data set with 70-fs pulse duration and 2-keV X-rays collected without laser excitation, using the phases from the known 3-Å model (pdb code: 2AXT) without refinement, see Section 1.2 for details of the PSII structure.

The PSII nanocrystals and microcrystals used for the pump-probe experiments were larger than the PSI nanocrystals and microcrystals used for the 2009 femtosecond nanocrystallography experiments. The maximum number of unit cells in a (5-μm)^3 PSII crystallite (using the unit cells parameters from PDB code 2-AXT) is approximately 1.4 x 10^7 unit cells, but the vast majority of the sample was of $\leqslant$ 2-μm microcrystals, as can be seen in Fig. 4.21.

The reason a larger filter size was used to filter the PSII samples in comparison to the PSI samples used in 2009 was that the experiments focused on different goals. Whereas the PSI experiments were utilized for technique development, the PSII experiments were done as proof-of-principle experiments for using femtosecond X-ray protein nanocrystallography for time-resolved crystallography on PSII. Consequently, no size restriction is necessary for the PSII crystallites, as we sought to produce damage-free electron-density maps of the S_1 and S_2 states to 7.4-Å resolution.

The goals of the femtosecond nanocrystallography experiments on PSII were

to show a proof of principle that the new technique of femtosecond X-ray protein nanocrystallography could be used to study PSII. In this regard, the experiments were a complete success. Nanocrystal samples of PSII that were 11 days old, much older than samples used for conventional X-ray crystallography, showed strong diffraction to the resolution limit of the experiment and actually produced stronger diffraction than comparably-sized PSI crystals. Although the diffraction was limited in resolution, this result may be the most remarkable result of the PSII work, and may indicate that nanocrystals and microcrystals are generally of higher quality than the larger protein crystals typical of conventional X-ray crystallography.

The electron-density map calculated to 8-Å resolution showed excellent electron density, with clearly visible transmembrane and extra-membrane helices. The 32-kDa protein of the PSII was visible and still intact in the electron-density map; the 32-kDa protein is one of the first proteins to be unresolvable when PSII is photo-damaged.

In total, the PSII results show the great potential of the femtosecond X-ray protein nanocrystallography technique for determining damage-free structures as well as for time-resolved experiments. The structure of the catalytic center has eluded scientists for decades, even though the mechanism of its action is of the utmost relevance. In addition, the unique capabilities of the LCLS, combined with the technique of femtosecond X-ray protein nanocrystallography, may allow for the first damage-free structure of the oxygen-evolving complex of PSII as well as the structures of the excited S states (2,3,4,0) of the catalytic cycle. The results shown here pave the way for time-resolved femtosecond X-ray protein nanocrystallography experiments with higher-energy X-rays at the CXI beamline of the LCLS. In fact, beamtime has been awarded to continue studying PSII using time-resolved femtosecond X-ray protein nanocrystallography, and the next set of experiments begins in Aug. 2011 with 1.33-Å wavelength.

4.10 Damage studies using PSI

Motivation for the PSI damage experiments

The Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory produces individual X-ray pulses of up to 3 mJ that can be varied between 30 and 400 fs duration, corresponding to about 10^{13} photons per pulse at a photon energy of 2 keV (6 Å wavelength) (Emma et al., 2010). This source has enabled a radically new approach to protein structure determination using crystals that are too small (or radiation sensitive) for conventional analysis. In femtosecond nanocrystallography, a suspension of nanocrystals flows across the focused, pulsed beam, with an X-ray diffraction pattern recorded on each pulse (Chapman et al., 2011) at a repetition rate of up to 120 Hz, as shown in Fig. 4.40. If a crystal intersects the focused X-ray pulse, it generates Bragg spots corresponding to the intersection of the Ewald sphere with the reciprocal lattice in the particular random orientation of the crystal (see Fig. 4.40b). Measurements are typically carried out with a peak X-ray fluence in the focused beam of 4 kJ/cm^2, corresponding to a maximum dose to a protein crystal of 3 GGy per pulse, 100 times higher than tolerable doses for cryogenically cooled crystals exposed at synchrotron sources (Owen et al., 2006). The peak irradiance (photons per unit time and area) of each LCLS pulse is about 10^{17} W/cm^2 at 70 fs duration, completely vaporizing the sample.

Acknowledgments regarding the PSI damage studies

My role was to find the necessary conditions to produce nanocrystal and microcrystal samples of PSI, produce the sample, work on the design of the experiments, assist with the sample delivery, and collect the X-ray-diffraction data. A complete list of the collaborators and acknowledgement of the contribution of each collaborator will be found in (Barty et al., 2011).

The experiments were done at the AMO beamline (Bozek, 2009) of the LCLS,

and I would like to thank the hard work of the entire LCLS team. The experiments were completed in the CAMP instrument (Struder et al., 2010), and I would like to thank the entire CAMP team for their hard work.

The work presented in this subchapter is from a paper that is in preparation for submission to Nature Physics (Barty et al., 2011).

PSI damage experiments overview

The initial femtosecond X-ray protein nanocrystallography experiments in Dec. 2009 provided data at 70-fs pulse duration that was indistinguishable from data collected at 10-fs pulse duration. Simulations by molecular dynamics (Neutze et al., 2000; Bergh et al., 2004; Caleman et al., 2011) and hydrodynamic codes (Hau-Riege et al., 2004, 2007; Gnodtke et al., 2009; Bergh et al., 2008) have predicted that motions of the ions created in the intense XFEL beam by 5 Å can occur in less than 100 fs, and that pulses as short as 10 fs may be required to achieve atomic resolution.

On exposure to an X-ray pulse, energy is primarily transferred to matter by atomic photoabsorption. For a 40 fs, 2 keV pulse at an irradiance of 10^{17} W/cm^2, about 10% of carbon atoms in a protein crystal absorb a photon, for example. Energy is initially released via emission of photoelectrons and Auger electrons, followed by a cascade of lower-energy electrons caused by secondary impact or field ionizations taking place on a 10 to 100 fs timescale. The Coulomb repulsion of the ions and the rapid rise in electron temperature of the system causes displacement of both the atoms and ions during the pulse. This heating leads to a high pressure that drives the explosion of the sample.

The difference between the experimental results and simulations indicated that a tampering effect (Hau-Riege et al., 2010) by the liquid jet or PSI crystallite itself may be allowing for longer pulse durations before the Coulomb explosion. However, the initial experiments in Dec. 2009 did not collect diffraction patterns from a sufficient

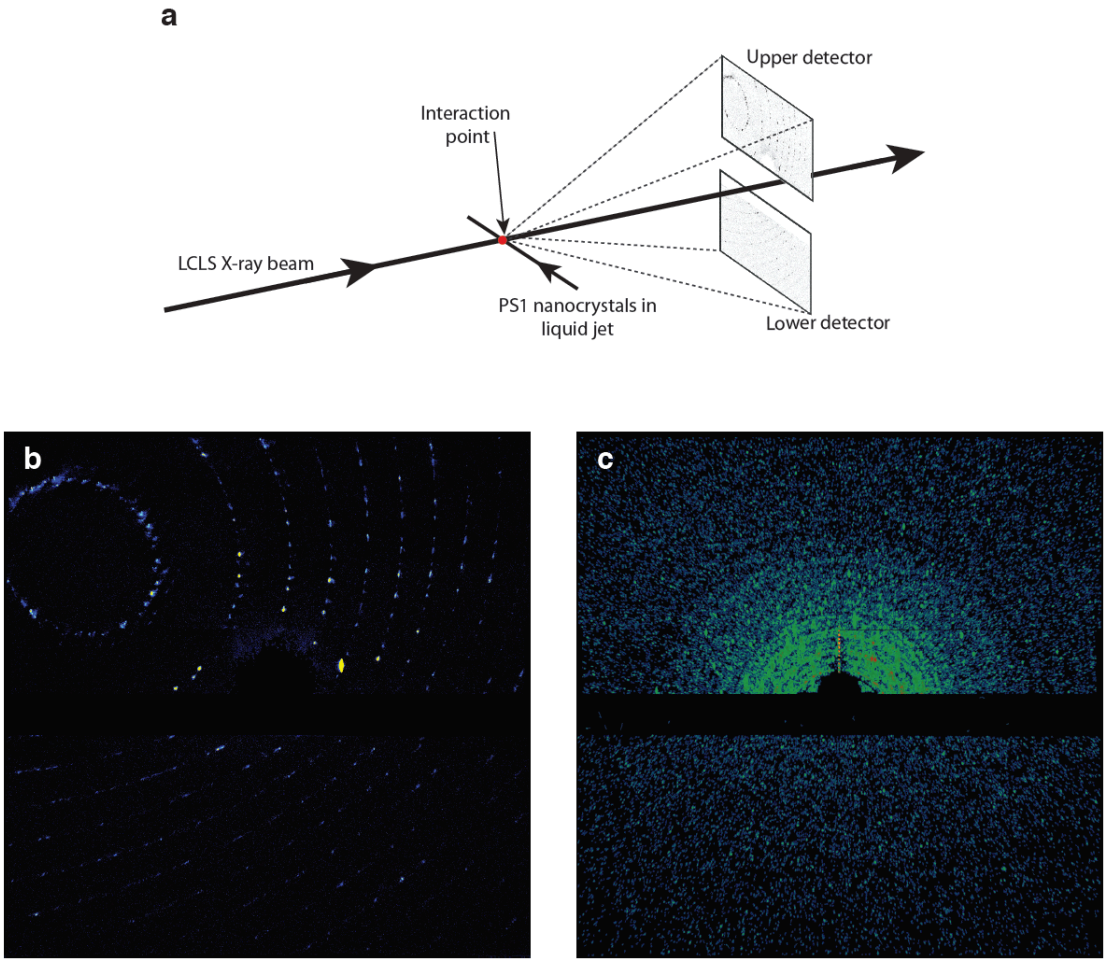

Figure 4.38: **2 keV femtosecond X-ray diffraction from photosystem I nanocrystals.** (a) Schematic of the LCLS diffraction experiment, in which a suspension of nanocrystals flows in a liquid jet across the X-ray beam, and diffraction is recorded using a pair of pnCCD detectors (Struder et al., 2010). (b) Diffraction pattern from a single nanocrystal recorded with a single 200 fs pulse of $2x10^{17}$ W/cm^2 irradiance. (c) "Virtual powder pattern" formed by summing 2639 single-pulse patterns with 200 fs duration pulses. Bragg peaks are visible to the corners of the detector, corresponding to a resolution length of 7.6 Å. Due to the large unit cell size of the crystal, Debye-Scherrer rings overlap and are not resolved at $q \geq 0.5$ nm^{-1}, consistent with the results of the serial powder diffraction work presented in Section 4.6. Figure and caption adapted from (Barty et al., 2011).

number of pulse durations, and a more systematic study was necessary. Consequently, the goal of this experiment, conducted in June 2010, was to study the effect of the short-pulse X ray-matter interaction on diffraction from protein nanocrystals, using PSI as a particular example (Jordan et al., 2001), using 10-fs, 70-fs, 100-fs, 150-fs, 200-fs, 250-fs, and 300-fs pulse durations, while maintaining a constant fluence between the different pulse durations.

PSI-crystallite sample preparation and delivery

Photosystem I crystallites made from one protein preparation were used for the damage experiments in June 2010 at the LCLS. The PSI crystallites were grown in low-salt conditions (10 mM $MgSO_4$) and settling experiments were used to select the PSI crystallites by size. The PSI crystallites in the 40-min settling step were used for the experiments. The PSI crystallites were inline filtered using a 2-μm filter.

The liquid jet was focused to a diameter of 4 μm by coaxial gas flow, and intersected the LCLS beam in the continuous liquid column, upstream of the breakup of the jet into drops due to the Rayleigh instability. The suspension of Photosystem I nanocrystals in their unadulterated mother liquor flowed at a rate of 10 μL/minute at a protein concentration of 1 mg/ml (1 μM PSI). The crystals intersected randomly with the LCLS pulses at a 60-Hz repetition rate. The pnCCD detectors were read out and digitized after each LCLS pulse. The probability of a crystal being in the intersecting volume of the X-ray and fluid beams at any point in time was 20%, in accordance with the observed rate of crystal diffraction patterns.

Data collection

Femtosecond nanocrystal diffraction patterns were collected at the Atomic Molecular and Optical (AMO) beamline (Bozek, 2009) at LCLS, in the CFEL-ASG Multi-Purpose (CAMP) end-station (Struder et al., 2010). Diffraction patterns were recorded

on a pair of X-ray p-n junction charge-coupled device (pnCCD) modules located 64.7 mm from the liquid jet carrying the nanocrystal suspension. The largest scattering angle intercepted by the detector modules was $2\theta = 49°$, yielding a highest resolution of 7.3 Å at a photon energy of 2 keV. The LCLS operated at a repetition rate of 60 Hz. X-rays were focussed into a focal spot of area of 10 μm^2 by a pair of Kirkpatrick-Baez mirrors in the AMO beamline.

Data reduction

Each recorded pnCCD frame was corrected for detector offset and gain, and a time-windowed average background was subtracted for each pixel. The corrected frame was subsequently searched for Bragg spots using a thresholding and morphological analysis algorithm. This algorithm generated a list of peak locations (in detector coordinates) with the X-ray counts equaling the sums of pixel values within a contiguous region assigned to the peak. Detector artifacts and scattering from the liquid jet were identified and excluded from the list of Bragg spots, and only patterns with at least three peaks were accepted for inclusion in the list. The magnitude of the photon momentum transfer, $q = (2\sin(\theta)/\lambda)$, for each peak was determined from the peak location and the shot-to-shot wavelength variation of LCLS (as calculated from the electron beam and undulator parameters (Emma et al., 2010)). Plots of $I(q;T)$ were generated by averaging the peak intensities in the list for each pulse duration. These 1D plots are also corrected for the solid angles of the detector pixels and scaled to the integrated counts at the lowest scattering angles (indicated by the vertical dashed lines in Figs. 4.40a and 4.40b) to account for fluctuations in pulse energy. However, this serial data, collected one crystal at a time, can be assembled into a 3D set of structure factors by first indexing each pattern and summing counts over each Miller index (Chapman et al., 2011; Kirian et al., 2011).

PSI damage experiment results

Strong Bragg diffraction was observed from single nanocrystals of PSI at the highest scattering angles intercepted by the detectors, for pulse durations up to 300 fs. The single-shot diffraction pattern shown in Fig. 4.40b was recorded at 200 fs duration. The observation of strong diffraction for pulses much longer than 100 fs (Chapman et al., 2011) appears at first sight to be at odds with the above mentioned models. The PSI damage experiments were devised to provide a systematic study of the decrease in scattered intensity that accompanied an increase in the pulse duration, if the pulse fluence was kept approximately constant to the different pulse durations. The damage experiments were done to hopefully reconcile the difference in observed PSI crystallite diffraction as a function of pule duration with the predicted behavior.

This apparent contradiction between the experimental results and the simulations can be resolved by acknowledging that damage does indeed occur on timescales shorter than the pulse duration and that Bragg diffraction from the crystal turns off as damage increases. As the explosion progresses over the duration of the pulse, disordering of the atomic positions proceeds to ever-longer length scales. As correlation of structure over the unit cells (i.e. periodic order) is lost on longer length scales, the scattered photon flux accumulating into the corresponding Bragg peaks diminishes and eventually terminates. The pulse-integrated counts in Bragg peaks are proportional to their undisturbed diffraction efficiencies, the pulse irradiance, and the lifetimes. Simultaneously, diffuse scattering will accumulate until the end of the pulse. The strength of the accumulated Bragg signals relative to this background depends proportionally on the number of unit cells in the crystal and inversely on the pulse duration. Although the diffuse background is negligible for the crystal sizes and pulse durations considered here, it could be minimized using the shortest possible pulses.

The pulse-integrated diffraction pattern can be modeled in terms of the distinct

processes of ionization and displacement of the atoms in the crystal. Ionization of atoms in the crystal occurs randomly, modifying atomic scattering factors (Hau-Riege et al., 2007; Quiney and Nugent, 2011), leading to both a constant reduction in Bragg signal and the addition of uniform diffuse scattering (Warren, 1990). For example, single ionization of half the atoms decreases the Bragg signals by about 20%. The measurements are not sensitive to this uniform change.

The atomic displacements caused by the X-ray interaction were calculated using the plasma modeling code Cretin (Scott, 2001) on a homogeneous protein sample in water. As the ion temperature rises during the X-ray pulse, so, too, does the mean square atomic displacement, shown in Fig. 4.39. At any point in time, the displacement of any particular atom in the unit cell is considered to be random with zero mean. As with the familiar analysis of a thermally-disordered crystal (Warren, 1990), the diffraction pattern is modified from the perfect crystal by an addition of a diffuse scattering term that increases with increased scattering angle, and a compensating reduction in the Bragg signal by a term $\exp(-4\pi^2 q^2 \sigma^2(t))$, where $q = (2/\lambda \sin(\theta))$ for a wavelength λ and scattering angle 2θ, and $\sigma^2(t) = < D^2(t) >$ where D is the component of the atomic displacement in the direction of the photon momentum transfer of the Bragg peak at time t during the pulse. For the conditions of this experiment, Cretin simulations show that the mean square displacement $\sigma^2(t)$ increases approximately as t^3 for high irradiance pulses and times longer than 10 fs (see Fig. 4.39a). For a pulse of irradiance I_0 and duration T, the accumulated Bragg signal is therefore given by

$$I(q;T) = I_0 T r_e^2 \, P\Delta\Omega \, |F(q)^2| \, g(q;T) \tag{4.9}$$

with

$$g(q;T) = \frac{1}{T} \int_0^T \exp(-4\pi^2 q^2 \sigma^2(t)) \, dt \tag{4.10}$$

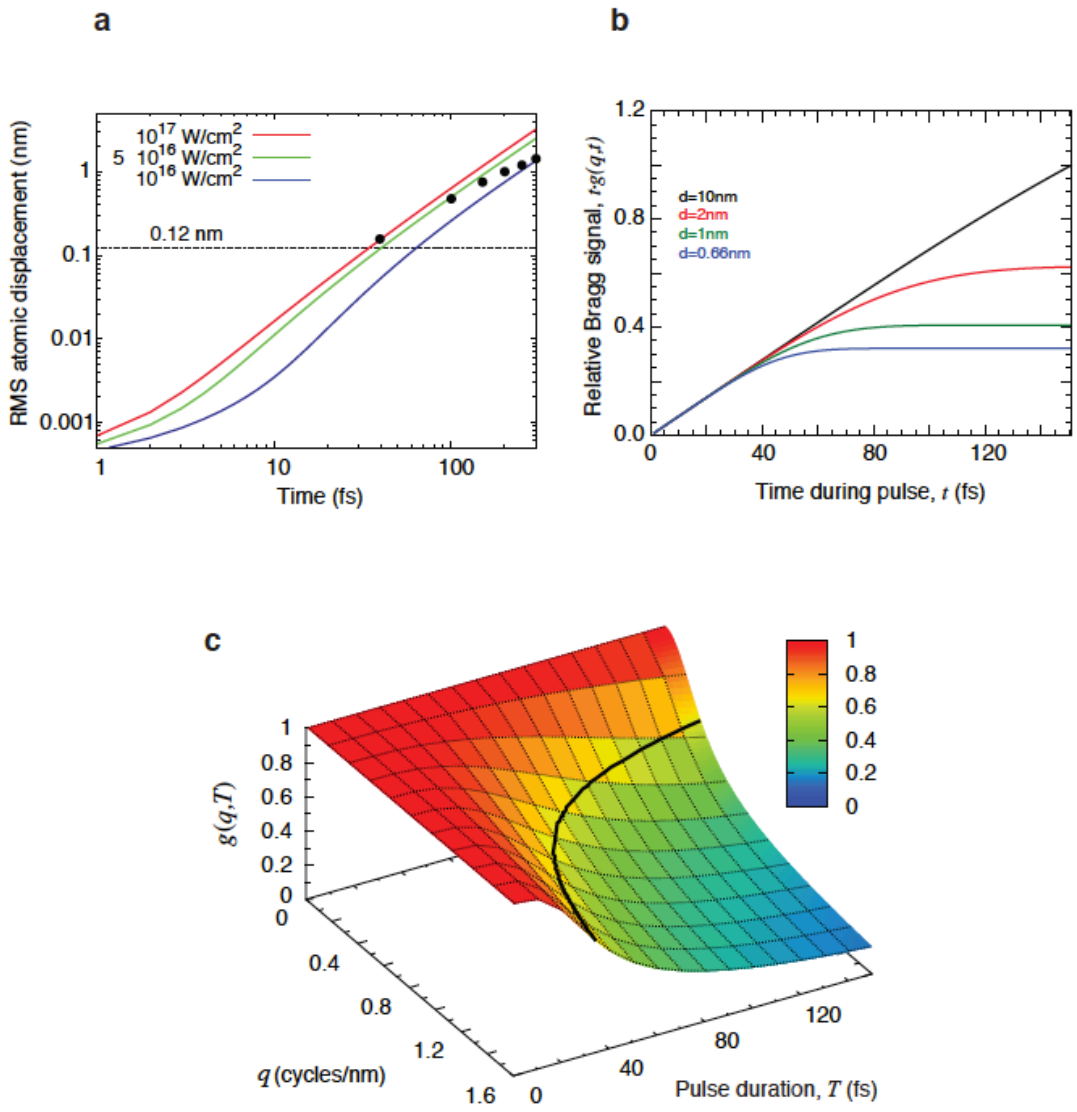

Figure 4.39: **Dynamics of exploding crystals calculated with CRETIN.** (a) Plot of the one-dimensional component of the RMS displacement, $\sigma(t) = \sqrt{< D^2(t) >}$, of atoms in a PSI protein sphere for constant irradiance X-ray pulses of 2 keV photon energy. The turn off time for our highest resolution length d = 0.76 nm occurs when $\sigma(t)$ reaches a value of $d/(2\pi) = 0.12$ nm (dashed horizontal line). Black circles show the predicted component of RMS displacement at the end of pulses of a constant fluence of 4 kJ/cm^2. (b) Plot of the relative accumulation of Bragg signal, t $g(q;T)$, and (c) the dynamic disorder factor $g(q;T)$, given by Eq. 4.10 using simulated $\sigma(t)$ values from (a) at I_0 = 10^{1}7 W cm^{-2} and pulse duration $T = 150$ fs. Bragg peaks accumulate signal (relative to the undisturbed case) at the same rate, dependent on the irradiance I_0. The accumulation of counts into the higher resolution peaks ends sooner, and the black line shows the turn-off time. The zero-frequency signal, which depends only on the total electron mass of the crystal, is unaffected by the crystal explosion. Figure and caption adapted from (Barty et al., 2011).

for a detector pixel solid angle $\Delta\Omega$, and where r_e is the classical electron radius, P the polarisation factor, and $F(q)$ is the structure factor of the room-temperature nanocrystal (with an implicit Debye-Waller term due to the initial displacements of the atoms from perfect lattice positions). The dimensionless dynamic disorder factor $g(q;T)$ gives the change in signal relative to the undisturbed sample for pulses of irradiance I_0 and duration T.

As can be seen from Eqs. 4.9 and 4.10 and Fig. 4.39a, as time progresses during the pulse, the Bragg diffraction effectively terminates when the root mean square displacement exceeds $1/(2\pi q) = d/(2\pi)$ where d is the "resolution length" of that Bragg peak. Higher resolution peaks turn off sooner, leading to lower counts accumulated on the detector (Fig. 4.39b). For the derived t^3 dependence of $\sigma^2(t)$, the turn-off time of a Bragg peak is estimated as $t_{\text{off}} = (2\pi q\sigma_T)^{-2/3}T$ where $\sigma_T = \sigma(T)$ is the RMS atomic displacement at the end of the pulse. Indeed, from Eq. 4.10, the disorder factor $g(q;T)$ tends towards

$$g(q;T) \approx \Gamma(4/3)\frac{t_{\text{off}}}{T} \approx \frac{\Gamma(4/3)}{(2\pi q\sigma_T)^{2/3}} \tag{4.11}$$

for pulse durations $T > t_{\text{off}}$ and where Γ is the Gamma function. Given high enough pulse irradiance I_0, scattering will be observed, even for pulses of longer duration than the explosion dynamics. The Cretin calculations predict a RMS atomic displacement of about 0.25 nm by the end of a 40-fs pulse at an irradiance of 10^{17} W/cm^2 for a homogeneous protein object (Fig. 4.39a). In this case, the accumulated Bragg signal at a resolution of 10 Å will be about 0.6 of the undisturbed signal, as contributed by approximately the first 30 fs of the pulse.

The effect of Bragg termination was measured over a wide resolution range, and as a function of pulse duration, by forming "virtual powder diffraction" patterns. An example of such a powder pattern is shown in Fig. 4.38b, obtained by summing

2639 single-pulse single-crystal patterns acquired with pulses of 200 fs duration. By averaging these patterns over shells of q, measures of $I(q;T)$ are obtained, as shown in Fig. 4.40a. As with conventional powder diffraction, these measurements are averaged over crystal shapes, orientations, wavelength spreads, and beam divergences (0.1% and 0.5 mrad, respectively, in this case). Various pulse durations between nominal values of 70 fs and 300 fs were achieved by varying the compression of the FEL electron pulses, which keeps the pulse fluence I_0T approximately constant. Previous work indicated that pulses below 100 fs may be less than half the nominal duration (Young et al., 2010). In Fig. 4.40a, there is a clear trend that the Bragg signal at higher scattering angles diminished as pulse duration increased.

The effect of Bragg termination can be isolated from other contributions in Eq. 4.9, such as the structure factors $F(q)$ and the initial room-temperature Debye-Waller factor, by taking the ratio of I(q;T) to the shortest, least damaging duration, $I(q;T = 70\text{fs})$. These ratios are plotted versus q in Fig. 4.40b and are compared with simulated plots from the Cretin calculations (dashed lines) where we assume the shortest pulses have 40 fs duration and all others follow their nominal values. Using Eq. 4.11, it is expected that the ratios of $I(q;T)$ are independent of q when $T > t_{\text{off}}$, which is indeed the case for $q > 0.75$ nm^{-1}. In this regime, and for constant pulse fluence I_0T, the ratios are equal to the ratio of relative turn-off times of the pulses, $\frac{t_{\text{off}}}{T}$. At 300 fs the 10 Å resolution Bragg signal therefore originates from about 0.25 of the contribution compared with 40 fs pulses (corresponding to a turn-off time of about 0.25 x 300 fs x 0.6 = 50 fs at this lower irradiance). Highest signals are achieved with the highest irradiance pulses. Future experiments will be carried out at shorter wavelengths to access molecular resolution (2-3 Å). The photoabsorption cross sections of light elements are about 60 times lower at 8 keV photon energy than at 2 keV, giving rise to a corresponding reduction in the dose that drives the explosion dynamics. This is offset somewhat by the requirement for higher X-ray fluence due to lower diffraction

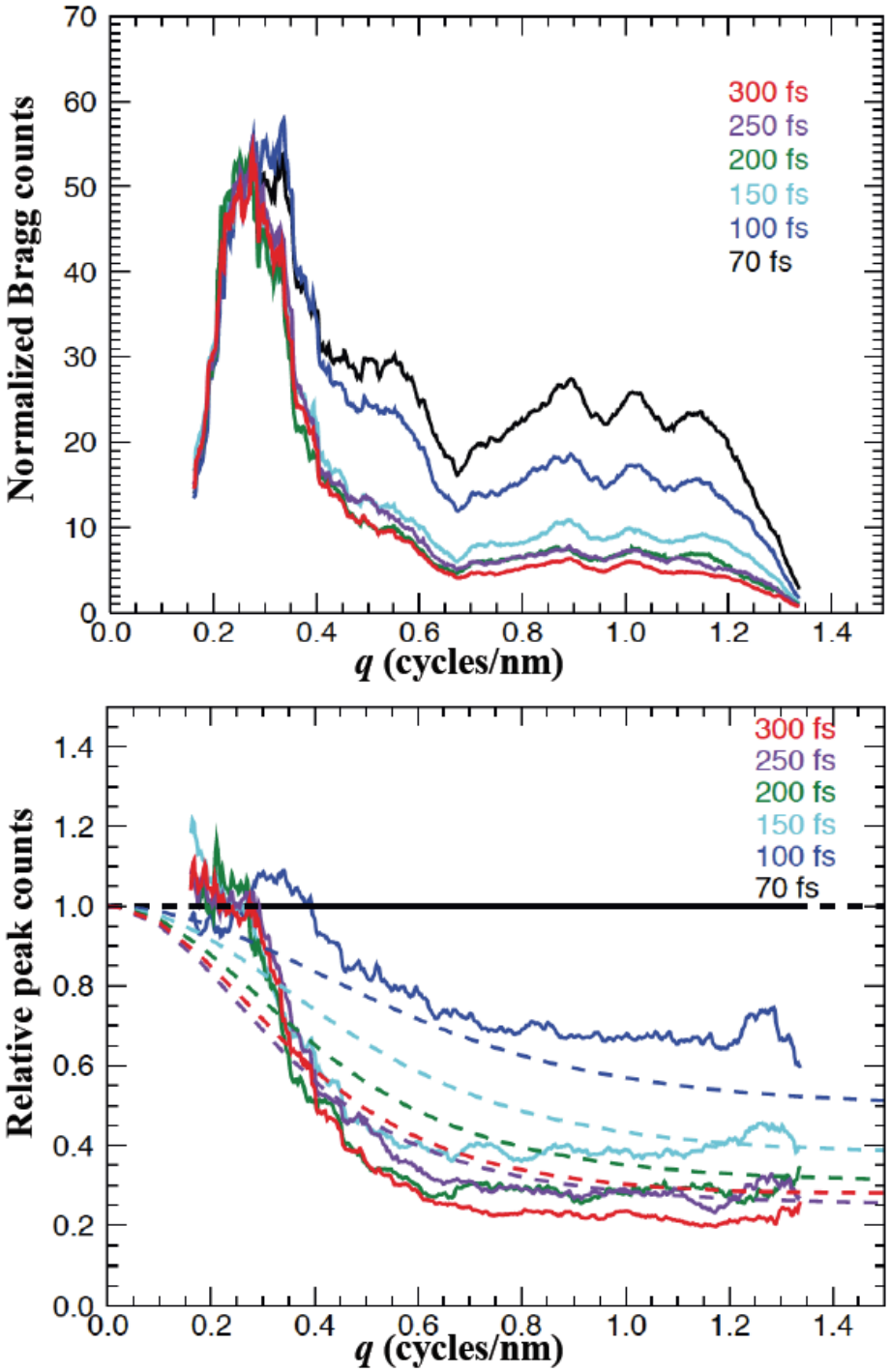

Figure 4.40: **Bragg termination observed at approximately constant X-ray pulse fluence $\mathbf{I_0T}$** (a) Bragg signal $I(q;T)$ of PSI nanocrystals averaged over q shells of virtual powder patterns such as in Fig. 1c, for nominal pulse durations T varying between 70 fs and 300 fs. (b) Bragg signal relative to the shortest pulses, plotted as solid lines. These ratios isolate the disorder dynamics and indicate the initial proportion of the pulse that contributed to Bragg diffraction. Dashed lines give the computed ratios of $I(q;T)$ / $I(q;T=40fs)$ from the CRETIN simulations of Fig. 4.39. Prior experiments at LCLS indicate that the nominal "70 fs" pulses are shorter than indicated (Young et al., 2010). We achieve a best fit assuming these pulses are 40 fs duration. Figure and caption adapted from (Barty et al., 2011).

efficiency. With a 30-fs duration pulse of 10^{18} W/cm^2 (e.g. a 1.5 mJ pulse energy focused to 5-μm^2 focus) we expect an RMS displacement of less than 1 Å by the end of the pulse (see Fig. 4.41a), for a deposited dose of 270 MGy. At 3-Å resolution, Bragg diffraction will self-terminate in about 20 fs, with g = 0.5 (see Fig. 4.41b), giving strong diffraction from crystals of the same size used here, and possibly beating extensive ionization of the sample (Young et al., 2010). The model predicts that pulses of even higher irradiance will give larger diffracted signals.

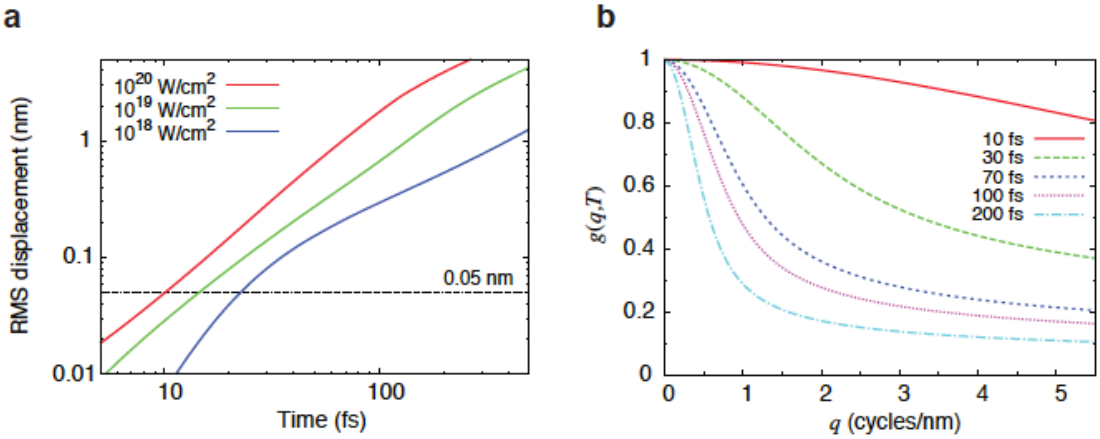

Figure 4.41: **Dynamic disorder factor at atomic resolution.** (a) Plot of the 1D component of atomic displacement in a PSI protein sphere for constant irradiance 8-keV X-ray pulses (1.5 Åwavelength). Higher irradiances than for 2 keV are required to achieve similar diffraction signals. The turn-off time for 3-Å resolution occurs when $\sigma(t)$ reaches 0.5 Å. (b) Plot of $g(q;T)$ for 8-keV pulses at 100 kJ/cm^2 fluence ($8x10^11$ photons μm^{-2}), for different pulse durations. At 100 fs duration, the pulse irradiance is 10^{18} W/cm^2. Highest diffraction efficiency and signal to background is reached with the shortest pulses, but longer pulses do not preclude the observation of Bragg peaks. Figure and caption modified from (Barty et al., 2011).

Summary of the PSI damage experiments

The experiments and models show that, as opposed to single-particle diffraction (Neutze et al., 2000), protein nanocrystallography does not require X-ray pulses to be strictly shorter than the radiation damage timescale. The pulse arriving at a Bragg spot on the detector is shorter than the incident pulse, as dictated by the explosion dynamics. These dynamics depend on the overall atomic composition, not on the unit cell parameters,

and hence should remain indicative of most protein crystals. Measurements at various pulse irradiances offer a way to characterize, and hence correct, the effect of Bragg termination, allowing accurate structure factors for structure determination (Chapman et al., 2011; Kirian et al., 2010). The ability to collect data from longer pulses for X-ray protein crystallography data may allow for the use of LCLS modes that have higher electron bunch currents, and larger fluences, albeit with longer pulse durations. The higher fluences will lead to stronger Bragg reflections, and may help increase the signal-to-noise and signal-to-background ratios.

4.11 First high-energy femtosecond crystallography experiments using PSI as a model system

Motivation for the high-energy LCLS experiments

The success of the femtosecond X-ray protein nanocrystallography experiment (Chapman et al., 2011) led to great excitement in the structural biology community, especially for those working with membrane proteins. However, the results of the initial experiments were limited in resolution by the wavelength of the X-rays provided by the LCLS. Low-resolution electron-density maps can offer valuable insight into the structure of protein complexes and membrane proteins, but often, higher resolution structures of sub-structures or subunits are modeled into the low-resolution envelopes. The goal of the femtosecond nanocrystallography project was ultimately to bridge the gap and make the determination of molecular- or atomic-resolution electron-density maps of membrane proteins and protein complexes using nanocrystals and microcrystals possible.

As a result of the initial femtosecond nanocrystallography success, as well as the PSI damage study experiments, new experiments were planned to utilize the upgrades to the LCLS. The major upgrade was the operation of the Coherent X-ray Imaging (CXI) beamline (Boutet and Williams, 2010) of the LCLS, with X-ray en-

ergies up to 9.4 keV (λ = 1.3 Å). The higher-energy X-rays had sufficiently-small wavelengths to allow for molecular- or atomic-resolution diffraction patterns and reconstructed electron-density maps of nanocrystals and microcrystals of PSI.

The major goal of the higher-energy femtosecond nanocrystallography experiments was to establish the proof of principle that membrane protein nanocrystals and microcrystals can diffract to molecular and atomic resolution and to test the extension of the diffract-before-destroy principle to the achieved resolutions. In order to monitor the radiation damage, data sets of two time points were measured: 20 fs and 42 fs.

The motivation behind the higher-energy femtosecond X-ray protein nanocrystallography experiments was to provide a proof of principle of the diffract-before-destroy concept for protein crystallography, using membrane protein crystals, to molecular or atomic resolution. If the diffract-before-destroy principle is shown to be successful for PSI nanocrystals and microcrystals, the femtosecond X-ray protein nanocrystallography technique would make viable use of the results of many crystallization screens that produce nanocrystals or microcrystals by using the crystals directly for structure determination experiments, without the need for subsequent optimization experiments. Success with the higher-energy femtosecond X-ray protein nanocrystallography technique would provide a new avenue for the structure determination of membrane proteins, making the new purpose of crystallization experiments the pursuit of well-ordered membrane protein crystals *of any size*.

Nanocrystals and microcrystals of Photosystem I were chosen as a model system. PSI is a large membrane protein complex ($\geqslant$1 million Dalton molecular weight, 36 protein subunits, 388 cofactors) and it presents a most stringent test for the analysis of nanocrystals by femtosecond X-ray crystallography. If nanocrystals and microcrystals of Photosystem I can generate diffraction patterns to molecular or atomic resolution, the results would indicate the generalizability of the femtosecond X-ray protein nanocrystallography technique.

Three major experiments were done during the first higher-energy femtosecond nanocrystallography experiments. The first experiment was to record diffraction data to molecular or atomic resolution, and use the data to produce merged intensities, calculate structure factors, and produce an electron-density map of PSI. The second major experiment emphasized the recording of the diffracted intensity for PSI crystallites at several different X-ray pulse durations, in order to gain insight into the X-ray-induced radiation damage to the PSI nanocrystals at higher resolution. The third major experiment was to test new injector types that could allow much lower sample use for future femtosecond X-ray protein nanocrystallography experiments.

Acknowledgments regarding the high-resolution PSI experiments

My role was to find the necessary conditions to produce nanocrystal and microcrystal samples of PSI, produce the sample, work on the design of the experiments, assist with the sample delivery, and collect the X-ray-diffraction data. I was involved in the decision-making process regarding the order of sample testing and the experiments done on the PSI.

The experiments were done at the CXI beamline (Boutet and Williams, 2010) of the LCLS, and I would like to thank the hard work of the entire LCLS team.

Experimental details of the high-energy LCLS experiments

The high-energy X-ray femtosecond nanocrystallography experiment was accomplished by introducing a stream of fully-hydrated PSI nanocrystals and microcrystals to the LCLS X-ray pulses, operating at a repetition rate of 120 Hz, through a liquid jet at the CXI instrument of the LCLS, using the method of femtosecond nanocrystallograhy (Chapman et al., 2011). The X-ray energies used for the initial experiments were between 9.3 and 9.4 keV. The detector was positioned such that the corners could record scattering to a resolution of approximately 1.9 Å.

PSI crystallites produced in G_{10} buffer using the ultrafiltration crystallization method (see Section 4.2), as were used in previous LCLS experiments, as well as re-crystallized PSI grown in G_0 buffer, were used as the PSI samples. The re-crystallized PSI nanocrystals and microcrystals were produced on site at the LCLS.

The PSI nanocrystal and microcrystal samples were pre-filtered using a 5-μm inline filter using the setup shown in Fig. 3.9. The PSI nanocrystal and microcrystal samples were loaded into a 2-mL steel syringe and placed into the anti-settling device developed by Dr. Robert Shoeman of the Max Planck Institute in Heidelberg. The sample was injected at flow rates between 10 and 25 μL/min.

The sucrose injector and LCP injector were tested at the CXI instrument in order to ascertain their viability as injector designs for future femtosecond X-ray protein nanocrystallography experiments. Three different sucrose injectors and two different LCP injectors were used, and diffraction patterns were collected to measure the background produced by the liquid jets of the new injectors.

PSI sample preparation and introduction

Photosystem I samples from eight different preparations were used during the course of the higher-energy LCLS experiments at CXI. The PSI crystals in the 10-min settling step of the ultrafiltration crystallization were used to re-crystallize PSI on site, in the designated biochemistry laboratory. Additionally, nanocrystals produced using the ultrafiltration crystallization method were used to collect data: Although the majority of the PSI sample used was crystallites in the 30-min settling step, PSI crystals from the 20-min settling step were also used for two runs. In all, data collection on PSI lasted 20 hours.

The PSI crystallite samples were filtered using a 5-μm inline filter, restricting the PSI crystallite size to less than 7.82 x 10^6 unit cells. However, the majority of the PSI crystallites were below 1-2 μm in size. In order to mitigate the crystal settling, a

large, 360° rotating setup housing a syringe pump—designed by Dr. Robert Shoeman of the Max Planck Institute in Heidelberg—was used to operate the liquid jet. The improved injector was used during the experiments to assist with jet alignment and to allow rapid replacement of nozzles.

Results of the high-energy LCLS experiments

Over one million diffraction patterns from PSI-nanocrystal samples were collected using 9.3 keV X-rays at the CXI beamline. The data were collected too recently for a large amount of indexing results to be available, but at present, the analysis is focusing on accurately finding the position of the detector tiles, so that a larger percentage of the diffraction patterns can be successfully indexed. Diffraction patterns of PSI were recorded beyond the water ring, to approximate resolutions of 3 Å. A diffraction pattern from a PSI nanocrystal to approximately 3.0 Å is shown in Fig. 4.42. Additionally, a diffraction pattern from a PSI nanocrystal with preliminary indexing results is shown in Fig. 4.43. These patterns are the first example of measurable Bragg reflections for such small membrane protein crystals to molecular resolution.

One quality of the molecular-resolution diffraction patterns collected from PSI nanocrystals is the well-defined and well-separated Bragg reflections. The patterns shown in Fig. 4.42 and Fig. 4.43 appear to be of much higher quality than the conventionally-collected diffraction patterns, as shown in Fig. 4.44. Whereas in the conventional case the overlap of the peaks becomes a limiting factor, as shown in Fig. 4.45a, no such problem exists in the recorded femtosecond X-ray protein nanocrystallography data, as shown in Fig. 4.45b.

The Nanocrystals used to obtain diffraction patterns similar to that in Fig. 4.42 were obtained by filtering the 20-min and 30-min settling steps from the normal PSI crystallization as well as re-crystallized PSI. No morphological differences were noticed in the diffraction patterns of the re-crystallized versus ultrafiltration crystallites,

Figure 4.42: **9-keV diffraction pattern from Photosystem I at LCLS** Femtosecond X-ray diffraction pattern "snap-shot" of PSI using 9.33 keV X-rays and a 42 fs pulse duration at the CXI endstation of the LCLS. The crystal was size selected using a 5 μm inline filter, and the diffraction is to molecular resolution. This is one of the first examples of femtosecond nanocrystallography diffraction patterns extending to molecular resolution. Data analysis is ongoing.

indicating that the re-crystallization did not produce conditions that were too harsh for the growth of comparably ordered PSI nanocrystals. However, the data has yet to be thoroughly analyzed, but comparing the data sets of the PSI prepared using ultrafiltration and pseudo-batch crystallization will be of much interest. The results could be an indication that nanocrystals are intrinsically better ordered than the large protein crystals and that the disorder is of similar magnitude regardless of the method of growth in the case of PSI. However, more analysis is needed to do a thorough comparison.

Analysis of the data is ongoing, but some quick analysis on the data is possible.

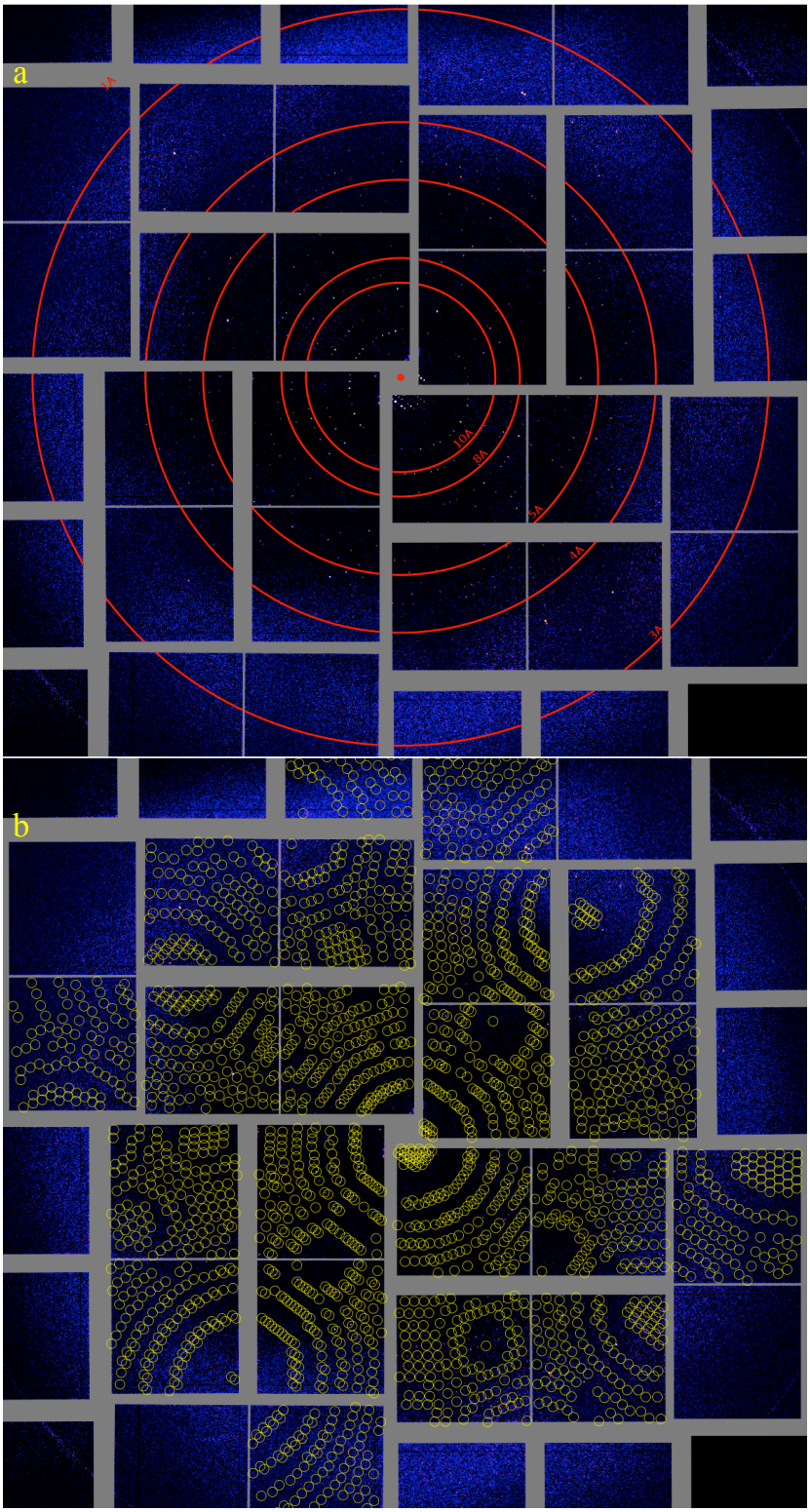

Figure 4.43: **9-keV femtosecond nanocrystallography PSI pattern with indexing results.** A Photosystem I diffraction pattern recorded to $\sim$ 3-Å resolution recorded at the CXI beamline of the LCLS. (a) Pattern with resolution rings indicated and (b) pattern showing the results of indexing using the current parameters for the detector geometry.

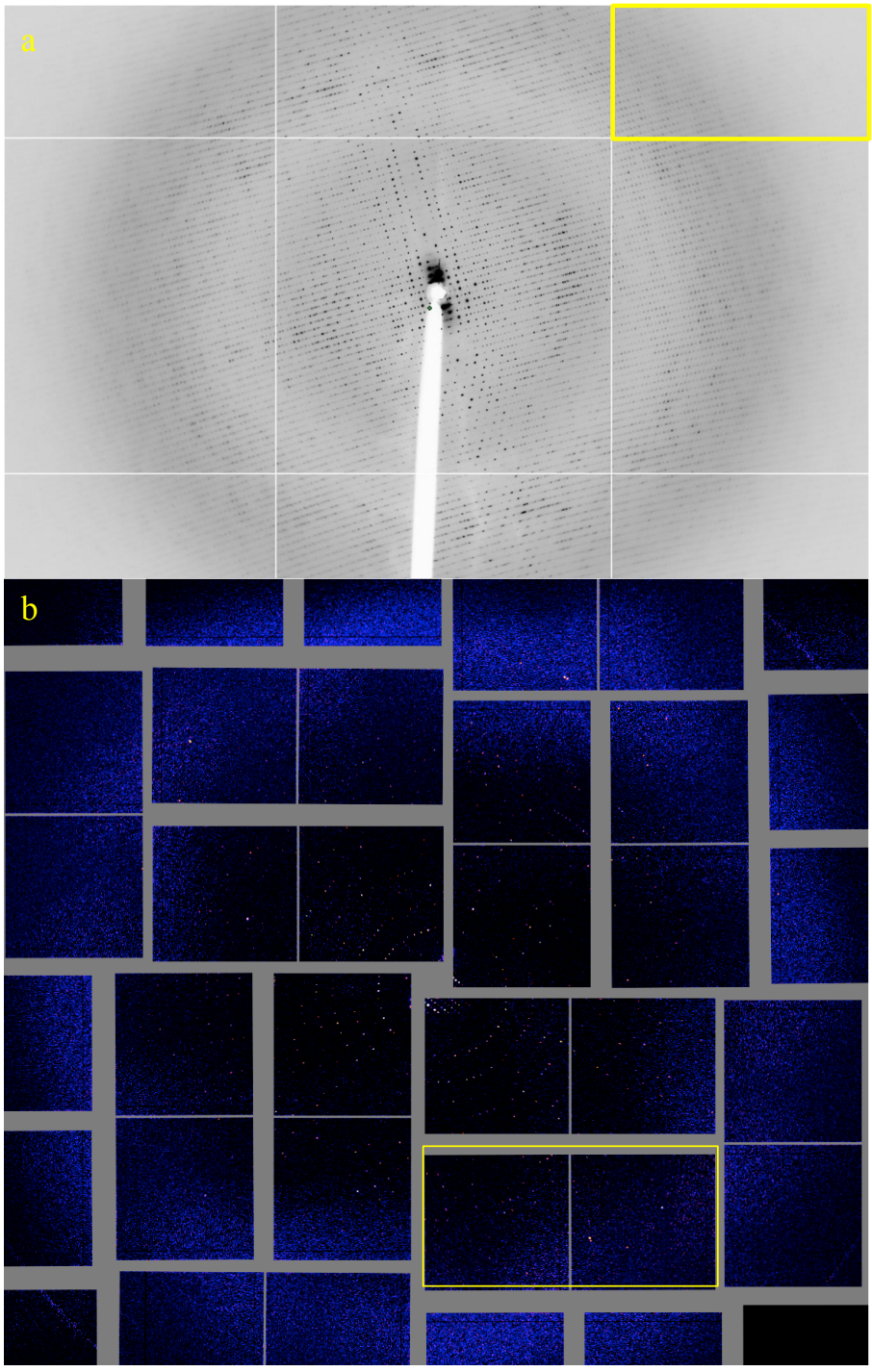

Figure 4.44: **Comparison of femtosecond and conventional data for PSI.** (a) A conventional PSI diffraction pattern of PSI collected at beamline 8.2.2 of the Advanced Light Source of Lawrence Berkeley National Laboratory using 0.2° rotation and a wavelength of 1 Å. (b) a femtosecond X-ray protein nanocrystallography pattern collected at the CXI beamline of the LCLS at SLAC National Laboratory using a wavelength of 1.3 Å. The Bragg reflections are much more dense in the conventional pattern because of the rotation of the crystal during the data set and its higher internal disorder.

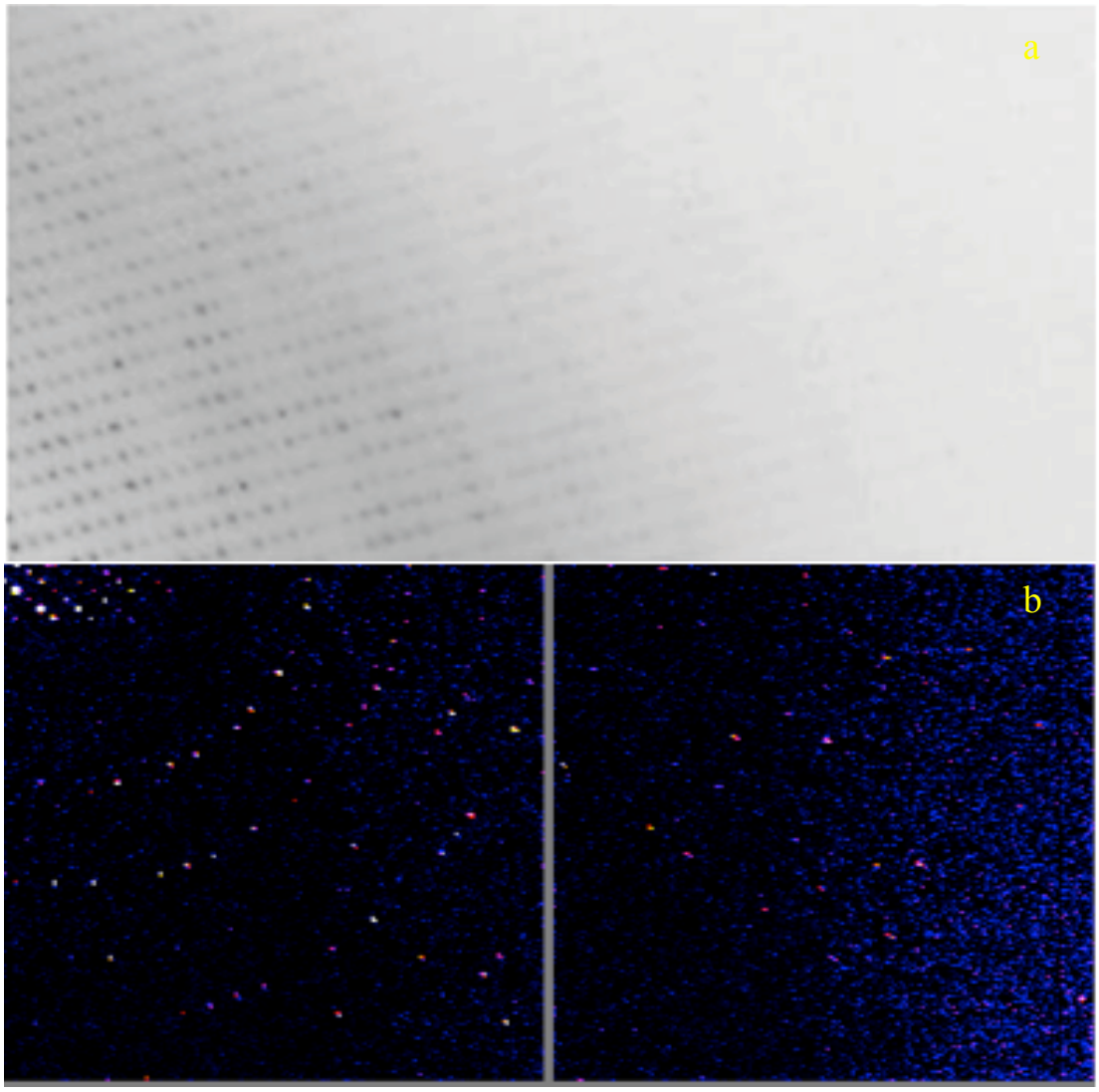

Figure 4.45: **Magnified comparison of femtosecond and conventional data for PSI.** (a) A magnified image of the tile highlighted in yellow of Fig. 4.44a. (b) A magnified image of the tile highlighted in yellow in Fig. 4.44b. The magnified images in (a) and (b) show the difference in patterns between the conventional PSI data (a) and the femtosecond X-ray protein nanocrystallography data (b). The pattern in (b) does not suffer from overlapping peaks, whereas the overlapping peaks in (a) limit the effective resolution of the data set.

The patterns collected from the LCLS showed much sharper diffraction patterns than any conventional pattern, due to the lack of angular integration in the LCLS "stills." There were limited shape transforms apparent in the patterns, which should allow the diffraction patterns to be indexed. Additionally, although the re-crystallized PSI was produced in very harsh conditions, there was no evidence of mosaicity or a change in

morphology to the Bragg spots or patterns.

Diffraction patterns showing internal disorder

Occasionally, a diffraction pattern was recorded that contained streaking of the peaks, as shown in Fig. 4.46. The streaking of the spots indicated that the PSI microcrystal used to collect the diffraction pattern had internal disorder. The patterns with streaks were only collected when PSI crystals from the 20-min settling step were used, and this indicates that the internal disorder was only appreciable once a certain PSI-microcrystal size was reached. Whent he 20-min settled crystals of PSI were used, large crystals—larger than the filter size—were seen post-filtering. Likely, these large crystals were some of the crystals removed from the membrane of the ultrafiltration-crystallization experiments. These larger crystals were often physically bound to the membrane, and could have had an increased internal disorder due to the mechanical disturbance of washing them off the membrane. It should be noted again for emphasis that these patterns were only collected when the large crystals were seen post-filtration using the 20-min settled crystals of PSI. The PSI nanocrystals did not produced diffraction patterns showing any internal disorder.

Currently, the dependence of the measured disorder on the size of the PSI microcrystal is unknown, but this will present an interesting area of research moving forward. As the discussion of the quality of protein nanocrystals with respect to large, single crystals increases, knowledge of the dependence of the measured internal disorder as a function of the size will be of the utmost interest.

Anti-settling results

A further and new achievement of the experiments was the mitigation of the settling of the nanocrystals and microcrystals within the reservoir system. A syringe-pump-driven nozzle was used, which uses coaxial gas flow to focus the liquid jet but uses the syringe

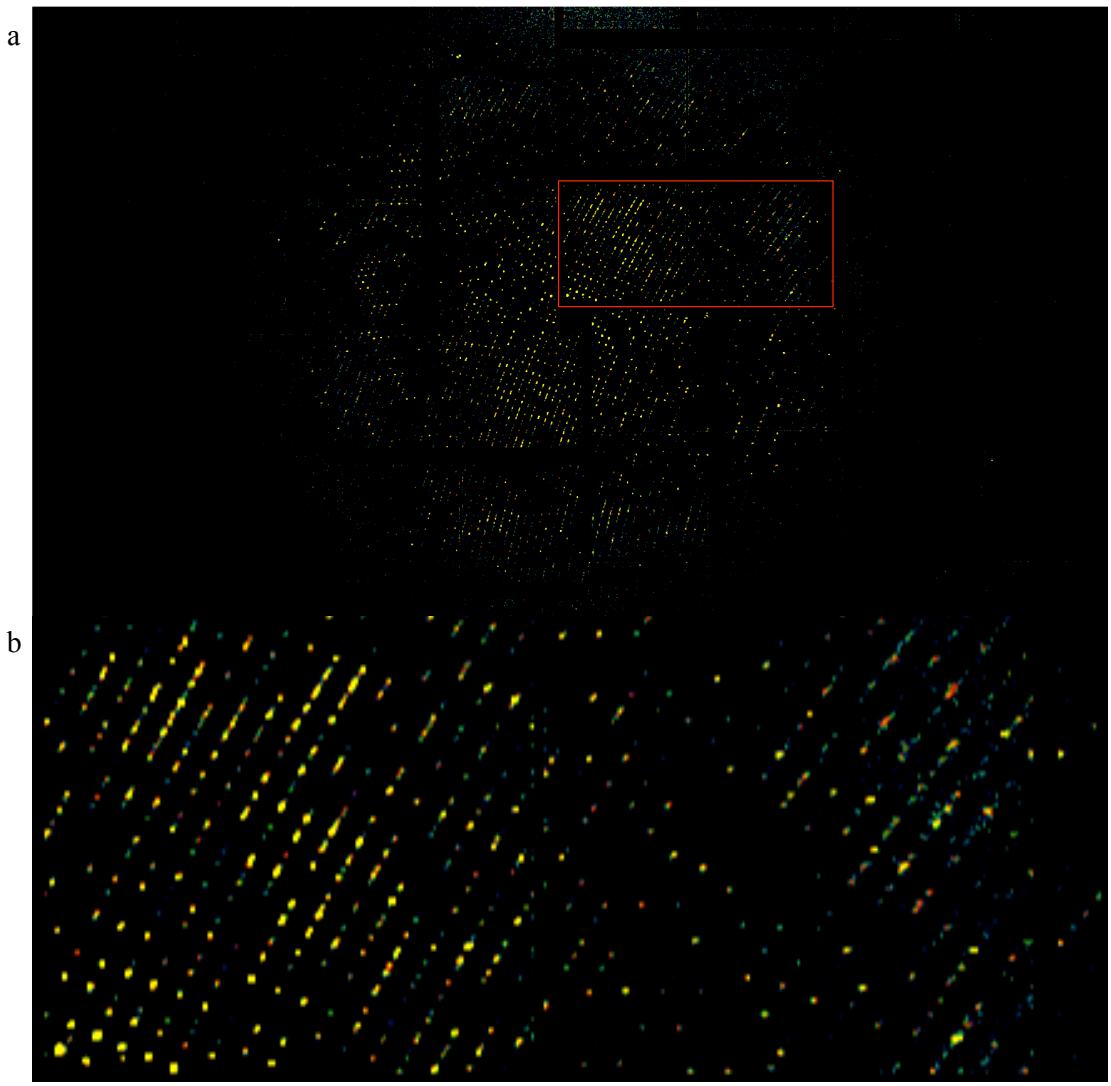

Figure 4.46: **High-resolution femtosecond nanocrystallography pattern with disorder.** Photosystem I femtosecond nanocrystallography pattern showing increase spot widths and streaking of reflections that is commensurate with a crystal containing high mosaicity. Diffraction patterns exhibiting high internal disorder were only measured once PSI crystals in the 20-min settling step were used in the data collection at the LCLS. Approximately 2% of the diffraction patterns collected from the 20-min settled PSI crystals showed high internal disorder. (a) Full pattern showing the reflections and (b) highlighted box from (a) that is magnified.

pump to drive the liquid flow of the sample. The syringe-pump and loaded syringe were mounted on a rotating mixer in order to keep the crystallites re-suspended. The syringe-pump-driven injector appeared to solve the settling problem, from monitoring the number of hits as a function of time within the control room, but further analysis is needed before conclusions can be drawn about the effectiveness of a rotating mixer.

Results of the sucrose and LCP injectors

The sucrose and LCP injectors were tested during the CXI experiments, but difficulties were experienced that need to be overcome before the injectors could be viable for femtosecond X-ray protein nanocrystallography. In the case of the sucrose injector, the injector immediately clogged, and after many attempts to start the flow that produced thick jets, the injector irreversibly clogged.

The LCP injector produced a 50-μm jet that was too thick for the femtosecond nanocrystallography experiments. Even with the incident intensity at 10% of the maximum value, the background scattering from the LCP was so strong that the detector was saturated. Consequently, both injector types require improvement before becoming viable in femtosecond nanocrystallography experiments.

Challenges during the high-energy LCLS experiments

Due to the water jet, the background scattering was very large and could present a problem by limiting the dynamic range of the detector pixels in future femtosecond nanocrystallography work. The jets produced from several nozzles were approximately 2 μm in size, and were noticeably perturbed by the presence of the few larger crystals. Additionally, the jet stability was highly dependent on the crystal density of the sample solution, with lower crystal densities allowing for a much more stable jet. Therefore, studies are necessary to determine the optimal working conditions of the liquid jet when using high crystallite-density suspensions.

One major concern that developed during the experiments was a higher-than-expected flow-rate for some of the experiments. For some of the collected data sets of PSI, the injector was operating at 25 μL/min, which consumed large amounts of the PSI sample. The normal operation of the injector was between 10 and 25 μL/min, but a large amount of sample was wasted in the higher flow rate experiments. Although PSI is made in large enough quantities to handle such large flow rates, most proteins will be excluded from such a technique sheerly by the amount of protein necessary for the experiments. In total, approximately 500 mg of protein was consumed during the coarse of the CXI experiments, which would be a large amount of protein for a recombinantly expressed membrane protein, even when scaled for the different protein masses. Thereby, increased attention will be directed towards new injector designs.

Additionally, the injector clogged more frequently during the CXI experiments than for any of the previous LCLS experiments. The apparent causes were, a too high nanocrystal density, which causes the inline filters to slowly clog, as well as a buildup from the liquid on the outer glass capillary. The deposition of sample onto the outer glass capillary of the nozzle was directly related to the proximity of the LCLS focus to the nozzle. Once buildup started on the surface of the nozzle, the nozzle would ultimately need to be replaced and cleaned, typically by ultrasonication. Consequently, the X-ray focus should remain at the maximum distance possible from the nozzle while still remaining within the stable region of the jet. Moving further along the jet to the droplet region is a possibility, but the hit rate will decrease and a larger background will exist in all of the patterns. Having the X-ray focus at the breakup region of the liquid jet will cause an inconsistent water background as well as jet streak in the diffraction pattern.

The water background may present difficulties in the analysis of the diffraction patterns to molecular resolution. In the case of the PSI diffraction patterns, the water background varied considerably from pattern to pattern. As a result, the moving-

average method of background estimation (Kirian et al., 2011) will likely be insufficient when recording Bragg reflections within the water ring (3.2-4.0 Å). The method most likely to allow an appropriate estimation of the background for each Bragg reflection is to use the local intensities within the pattern surrounding the Bragg peak, but the high spot density of the large PSI unit cell may present a challenge. The difficulty will likely be alleviated by the sparse number of Bragg reflections in a given pattern, which yields well-separated Bragg reflections that avoid overlaps.

Discussion of the femtosecond nanocrystallography project

The diffraction patterns that were recorded to $\leqslant$ 3-Å resolution using PSI nanocrystals and microcrystals that were maximally 5 μm in size indicate that the femtosecond nanocrystallography technique extends to molecular resolution. Although the largest crystals permitted through the filter were 5 μm in size, the majority of the PSI nanocrystals and microcrystals were below 1 μm. The nanocrystals and microcrystals are much too small for use with a microfocused conventional X-ray source—the task of finding an individual crystal that is $\leqslant$ 1 μm in size using a beamline camera system would be a daunting enough challenge—due to the X-ray-induced radiation damage effects. Therefore, the femtosecond X-ray protein nanocrystallography technique has allowed for the use of crystals of a membrane protein that cannot be seen under an optical microscope, which represents an unprecedented milestone for X-ray protein crystallography.

The first results of high-energy femtosecond X-ray protein nanocrystallography recorded Bragg peaks to molecular resolution. New developments will likely produce even better resolutions. The 5-μm PSI crystals could not be used at a microfocused beamline using a conventional X-ray source, and yet, when utilizing the LCLS, the crystallites are more than sufficient to attain molecular resolution diffraction patterns. Once the new optics are installed at the CXI instrument, significant background reduction should be achieved, allowing for even higher-quality data collection.

The 3-Å-resolution PSI data has yet to be analyzed to the point that any discussion regarding X-ray damage to molecular resolution can be initiated. However, the nominal 42-fs pulse duration of the X-rays used at the CXI instrument are shorter than the timescale of nuclear motion; therefore, the damage that could occur in this timescale would be due to the production of Auger electrons and low-energy electrons producing a background scattering to the diffraction pattern. Consequently, no specific damage is expected on this timescale, and the resulting electron-density maps should not suffer from significant X-ray-damage induced alterations. However, the final answer will not be known until the data is fully analyzed.

As can be seen in Figs. 4.42 and 4.43, the molecular resolution diffraction pattern is of higher quality than the comparable conventional diffraction pattern with similar resolution. Many reasons are involved, and may be involved, in the higher quality of the data. The femtosecond X-ray protein nanocrystallography patterns contain much sparser diffraction patterns, *i.e.* because the diffraction patterns are "snap-shots," the data will not contain as many overlaps as in the conventional case. As seen in Fig. 4.44a,b the peak overlap that occurs conventionally can limit the resolution to which the data can be analyzed. Another possible reason for the higher-quality data is that nanocrystals and microcrystals are (likely) more perfect.

Many discussions can be initiated regarding the use of femtosecond nanocrystallography to avoid radiation damage and the growth of large, well-ordered protein crystals. The use of protein nanocrystals and microcrystals for structural studies hinges on the ability to produce protein crystallites. The experience working with large, membrane-intrinsic protein complexes indicates that finding conditions to produce nanocrystals is not as challenging as finding the conditions to produce a large, high-quality single crystal. In the cases of both PSI and PSII, relatively harsh crystallization conditions were utilized and produced nanocrystals and microcrystals that diffracted strongly when placed into the LCLS beam. These results led to the view that the growth of

nanocrystals and microcrystals of membrane proteins and protein complexes will be easier and more prevalent than the large crystal growth that is currently sought.

Statistical analysis of protein crystallization experiments will be necessary in order to ascertain the probability of finding conditions that lead to nanocrystals and microcrystals during crystallization screens, in the general case. Although the large protein structural biology centers would seem to have all the data necessary to conduct a thorough analysis, the production of showers of nanocrystals and microcrystals was never the goal and may not have been reported. However, researchers at the Hauptman-Woodward Medical Research Institute have observed over 50% – possibly as high as 80% – of the proteins used in crystallization experiments (out of >12,000) produce mixtures with precipitate. The researchers sample over 1500 different crystallization conditions during the crystallization experiments for a given protein (Snell, 2011).

Conceptually, it would seem likely that the probability of finding conditions that produce nanocrystals and microcrystals is higher than the probability of establishing conditions that produce large protein crystals. Protein crystallization is a thermodynamically driven process in which phase separation occurs due to the supersaturation of the protein in a given solution. Ideally, when seeking large protein crystals, a crystallization experiment will create conditions that are within the metastable region of the phase diagram (or enter the metastable region), with the nucleation zone being transiently entered to allow for the creation of a single critical nucleus, from which the system will grow a single new crystal. The phase separation that occurs in the nucleation zone is driven by local and global thermodynamic instabilities that force the system to spontaneously decompose into separate phases. In principle, the space of the phase diagram in which spontaneous decomposition of the system occurs, and hence the critical nuclei for a new phase are formed, would be larger than the space separating the metastable and nucleation zones. Consequently, it may be a simpler task to find conditions that produce protein nanocrystals and microcrystals.

The PSI and PSII examples are from proteins with well-characterized phase diagrams. One question undoubtedly to arise is whether the nanocrystals and microcrystals produced from a random crystallization trial of a protein could be directly used for nanocrystallography experiments or if rational design is a prerequisite. Currently, there is not a straightforward answer because the variability in the quality of nanocrystals has never been directly addressed. In conventional crystallization experiments, the conditions used to crystallize a protein will have a large impact on the internal order of the protein crystals. Results using the PSI nanocrystals would indicate that protein crystallites produced using different conditions have surprisingly similar internal order, as the diffraction patterns from the re-crystallized and ultrafiltration crystallized PSI show no morphological differences, interpreted as differences in the spot profiles and in the number of reflections per pattern. However, it is unlikely that "all nanocrystals are created equally," *i.e* the conditions used to produce the nanocrystals do not impact the quality of the protein nanocrystals, in this case. This is one area that should see a lot of experiments moving forward, but it ties in closely with the discussion on whether protein nanocrystals are of the same quality as or superior quality to large protein crystals, as the first high-energy femtosecond X-ray protein nanocrystallography results suggest.

An experiment to possibly allow insight into the variability of the quality of nanocrystals (or lack thereof) would be to compare PSI crystallites grown in G_8 through G_{14} at constant protein concentration, normalizing for differences in the size distribution of the crystallites. The salt concentration used to grow large PSI crystals has a distinct effect on the average internal disorder of the PSI crystals when making large crystals, and the described experiment would analogously test the PSI nanocrystals. However, insight can be gained from the PSI re-crystallization experiments. Although the PSI was re-crystallized in very harsh conditions, i.e. very low ionic strength (essentially zero), the diffraction patterns produced from the re-crystallized sample appear to

show no morphological differences to the nanocrystals grown by batch crystallization. However, both are far superior to large, single crystals of PSI grown over several days, with the slowest growth rates possible and using micro- and macroseeding, with regard to internal disorder. However, the data sets from the PSI crystallites produced in the re-crystallization method and the batch method still need to be thoroughly analyzed and compared before final conclusions can be drawn. Ultimately, comparison of the integrated data sets, structure factors, and electron-density maps will be needed, which is a work in progress.

In the case of PSI, the diffraction patterns alone show the superior order of the nanocrystals and microcrystals compared to large, single crystals of PSI. Using the mosaic domain model of a crystal, one reason for the exceptional order of the nanocrystals is that these avoid the formation of different mosaic domains. This leads to an extremely low mosaicity and less overlap of reflections for protein crystals with large unit cells. Additionally, since the intensity is not subtended over as large a solid angle, improvement in the signal-to-noise ratio could be expected. However, as discussed previously, the mosaic domain model may be slightly too simplistic, and there may also be nanocrystals that diffract weakly due to flexible domains, weak crystal contacts, heterogeneities in the protein, or simply too large of a ratio of surface-exposed unit cells to volume-enclosed unit cells.

A fundamental question moving forward is the nature of the production of nanocrystals. Thermodynamically, nanocrystals are initially formed because a large supersaturation of the protein causes spontaneous decomposition at multiple places on the same timescale. Crystallite growth stops when the saturation line is reached, at which point the crystallites remain in equilibrium with the monodisperse protein in solution. However, microscopic reasons may exist that stop the growth of crystallites, such as the buildup of contaminants on the growth surfaces. Further investigations involving the monitoring of crystallite growth using AFM should be extended (as has been done for

soluble proteins). In general, it is the view of the author that the main crystallization method that will be utilized for crystallography involving protein nanocrystals is batch crystallization. As such, the ideal strategy for conducting nanocrystallography experiments is recommended to be to first use vapor-diffusion experiments and micro-batch experiments to accurately determine the phase diagram of the protein of interest (this includes conditions that produce a "shower" of nano- and microcrystals during a routine crystallization screen and proceeding in optimization from there). The knowledge of the phase space is then used to design batch crystallization experiments targeted to a specific area in the solubility curve. Of course, crystallization methods such as free-interface diffusion can be attempted and scaled appropriately. However, batch crystallization offers the control and scalability that is highly desired for the experiments described below.

The amount of protein consumed during the femtosecond nanocrystallography experiments needs to be minimized in order to make the technique more applicable to less naturally abundant proteins. Besides optimizing the nozzle and injector to allow lower flow rates, the use of proper scaling and auto-indexing of the crystallographic data would allow for a reduction in the total number of patterns needed to properly determine the intensity associated with the peak and calculate accurate structure factors. Proper scaling of the data is currently being worked on, but without success. However, continued experiments with scaling are planned.

The ability of femtosecond nanocrystallography to allow the determination of experimental phases, utilizing techniques such as isomorphous replacement (SIR or MIR) or anomalous dispersion (SAD or MAD), as well as using shape transforms, is another point of interest. Experiments are planned to test the feasibility of determining experimental phases using the information contained within the shape transform as well as using MAD and/or SIR/MIR. MAD will require data of high enough quality to allow the differences in the Friedel pairs to be statistically relevant compared to the

uncertainties associated with the intensities due to Monte Carlo integration, *i.e.* the R_{symm} values will need to be sufficiently small, but how small is currently unknown. SAD and MAD only became commonplace in protein crystallography once the beam quality became high enough to allow for very accurate and precise measurements of the intensities associated with the Friedel pairs; the uncertainties in the wavelength and pulse durations between pulses may make a seeded XFEL necessary. However, the photoelectron ejection from the K-shell of heavy metals will increase the difference in the Friedel pairs, possibly facilitating the use of MAD or RIP at the LCLS (Chapman, 2011). The use of SIR or MIR should be achievable with the current Monte Carlo integration of the intensities, although very precise measurements of the intensities will be needed, which may necessitate a larger number of patterns than necessary for samples requiring molecular replacement. The use of heavy metal additives may prove problematic with nanocrystals, and therefore the incubation of protein nanocrystals with heavy-metal derivatives should be closely studied.

Many experiments remain to be done regarding the viability of femtosecond nanocrystallography. Determining whether nanocrystals are more perfect than large protein crystals could very well define the role of femtosecond nanocrystallograghy moving forward. For instance, if the nanocrystals are shown to be intrinsically better ordered than the large protein crystals, femtosecond nanocrystallography could become the preferred choice of structure determination. Femtosecond nanocrystallography may be used especially for difficult-to-crystallize proteins and proteins that are prone to X-ray-induced radiation damage only. However, there is also a possibility that femtosecond nanocrystallography can be shown to be the superior technique for structural biology moving forward.

The ability to determine the structure of difficult-to-crystallize proteins, especially of membrane proteins and the protein complexes, could offer new avenues of drug design based on rational design, with the drugs specifically designed for the ac-

tive site of the protein of interest. This could significantly decrease the cost of drug development and limit the side effects of a drug. Additionally, the ability to outrun damage processes could allow the determination of the undamaged structures of many protein that are prone to X-ray-induced radiation damage, such as metalloproteins. To cite one specific case, femtosecond nanocrystallography may present the only opportunity to determine the structure of the undisturbed manganese cluster of photosystem II; the manganese cluster is responsible for the water-splitting reaction that has created the oxygenic environment on earth. The oxygenic atmosphere supports life as it is known on earth, and so the knowledge of the fundamental process that allows mankind to exist in its current form, could be now attainable through the use of femtosecond nanocrystallography. However, knowledge of structure of photosystem II could allow for the rational design of solar cells based upon the known structure, possibly contributing a method of producing new solar cells designed in a manner similar to what nature produced.

Chapter 5

OUTLOOK AND CONCLUSIONS

Membrane proteins are involved in many vital cellular functions and pathways, such as photosynthesis, cellular respiration, and signal transduction, amongst many others. However, the number of solved (unique) membrane protein structures is less than 300. Although X-ray protein crystallography is the workhorse of structural biology, there are many difficulties associated with the technique. The two main difficulties of conventional X-ray protein crystallography that the femtosecond nanocrystallography technique attempted to address were the necessity for large, well-ordered protein crystals - which could take years to grow and optimize, but the conditions may never be found in the first place - and the X-ray-induced radiation damage deteriorating the quality of the data and reconstructed electron-density map.

The femtosecond X-ray protein nanocrystallography project was initiated in order to determine whether membrane protein crystallites could be utilized to obtain high-resolution X-ray diffraction patterns, with the ultimate goal of producing molecular- and atomic-resolution electron-density maps to solve the structures of these complex proteins. Many experiments were necessary during the course of the project. Experiments were related to sample production, such as the development of methods to produce and characterize membrane protein nanocrystals, testing of the reproducibly of membrane protein crystallite production. Further, some experiments were related to sample quality, such as determining the diffraction potential and minimum size of membrane protein crystallites. Other experiments and theoretical developments were necessary for the development of sample handling and delivery systems, as well as the development of novel data acquisition and analysis schemes.

Large amounts of data and existing results have been acquired during the course of the femtosecond X-ray protein nanocrystallography project.

The project has shown the first proof-of-principle that femtosecond nanocrystallography can be established as a novel technique for structure determination of membrane proteins. The success of the project shows that the diffract-before-destroy principle extends to sub-nanometer resolutions for biological imaging and crystallography experiments. The femtosecond nanocrystallography data show that the technique overcomes the X-ray-induced radiation damage problem in X-ray protein crystallography. The data collection for the femtosecond X-ray protein nanocrystallography experiments was achieved from PSI crystallites that were at room temperature in the harvesting buffer G_0. Thereby, the femtosecond X-ray protein nanocrystallography technique overcomes several challenging aspects of conventional X-ray protein crystallography, namely cryogenically cooling of the protein crystal and the use of cryo-protectants in order to prevent ice formation.

One of the major initial results of the femtosecond nanocrystallography project was that very small crystals of membrane proteins exist, containing as few as one hundred unit cells. The fact that small membrane protein crystals exist practically necessitates that soluble protein crystals of this size exist and that these crystallites can be used for structural studies to molecular and atomic resolutions. However, the novel technique of femtosecond nanocrystallography must be utilized to allow the use of the small crystallites for structure determination experiments.

One very interesting result that may have far-reaching consequences is that the femtosecond nanocrystallography experiments show that the nanocrystals and microcrystals of PSI are superior in quality to the large, single crystals of PSI used in con-

ventional crystallography. The diffraction patterns collected contained very sharp spots to the highest resolution, which is in stark contrast to the conventional case for PSI, where spot overlap generally reduced the effective resolution of the crystallographic dataset. The superior quality of the protein nanocrystals and microcrystals may be the most important discovery of all, with regard to the results obtained so far.

The femtosecond X-ray protein nanocrystallography method is already a high-throughput technique. As the higher repetition rates of the LCLS and other XFELs are achieved, the number of patterns needed for a complete data set could be collected in less than an hour. This feature of the method should allow it to become a general technique that is amenable with the protein structure initiatives as well as industrial interests.

An exciting feature of the femtosecond X-ray protein nanocrystallography technique is that it will allow for time-resolved X-ray crystallography experiments, making use of the superior temporal resolution of the LCLS and other XFELs. The high temporal resolution offered by XFELs will allow molecular movies to be determined of molecules at work. In some regards, this would truly be the pinnacle of structural biology.

Outlook

Although the first results of femtosecond X-ray protein nanocrystallography have been published just four months ago (Chapman et al., 2011), the technique is rapidly improving and maturing. The 3-Å resolution diffraction patterns collected at the CXI beamline of the LCLS were recorded at the very first attempt to obtain high-resolution data. The excitement is rapidly building in the structural biology community to harness the potential of the femtosecond nanocrystallography technique that could revolutionize the capabilities of the structural biology community.

Femtosecond X-ray protein nanocrystallography will offer a new path to the

structure of difficult-to-crystallize proteins, one in which a crystallite of any size may be sufficient for structure determination. The most exciting aspect of the technique is that femtosecond X-ray protein nanocrystallography will allow the structures of membrane proteins to be determined without the ill-effects of X-ray-induced radiation damage on the sample.

Ultimately, the true power of femtosecond nanocrystallography will be in its ability to harness the full spatio-temporal resolution offered by X-ray free electron lasers. The technique could allow for the temporal resolution of ultrafast spectroscopy with the spatial resolution of X-ray crystallography, which is an unprecedented combination in science. The immediate implications are in determining the photocatalytic cycles of the photosynthetic proteins, as well as watching light-induced undocking of the proteins involved in the electron transfer of photosynthesis. This could lead to a very fundamental understanding of the basic processes of life and could have implications in artificial protein design and engineering.

The potential applications for femtosecond nanocrystallography are far reaching and could have a major impact on many areas of science and engineering. Results obtained from the method could be used in every area from basic science to the rational design of drugs. However, as is true with most new technologies, some major uses of femtosecond nanocrystallograhy will not be realized or appreciated for many years or decades to come. Although much is left to be done in regard to bringing the femtosecond nanocrystallography method to its full potential, the impact of the technique may be felt in many disciplines. The future of structural biology will be bright, and femtosecond X-ray protein nanocrystallography may present the technique necessary for a paradigm shift from the (mostly) static X-ray protein crystallography of today to the (mostly) time-resolved X-ray protein crystallography of the future.

Ago, H., Kanaoka, Y., Irikura, D., Lam, B.K., Shimamura, T., Austen, K.F., and Miyano, M. Crystal structure of a human membrane protein involved in cysteinyl leukotriene biosynthesis. *Nature* **448**(7153) 609–U12 (2007).

Altarelli, M. From 3rd-to 4th-generation light sources: Free-electron lasers in the X-ray range. *Crystallography Reports* **55**(7) 1145–1151 (2010).

Aro, E.M., Virgin, I., and Andersson, B. Photoinhibition of Photosystem-2 - inactivation, protein damage and turnover. *Biochimica Et Biophysica Acta* **1143**(2) 113–134 (1993).

Arthur, J., Materlik, G., Tatchyn, R., and Winick, H. The LCLS - a 4th-generation light-source using the SLAC linac. *Review of Scientific Instruments* **66**(2) 1987–1989 (1995).

Ayvazyan, V., Baboi, N., Bohnet, I., Brinkmann, R., Castellano, M., Castro, P., Catani, L., Choroba, S., Cianchi, A., Dohlus, M., Edwards, H.T., Faatz, B., Fateev, A.A., Feldhaus, J., Flottmann, K., Gamp, A., Garvey, T., Genz, H., Gerth, C., Gretchko, V., Grigoryan, B., Hahn, U., Hessler, C., Honkavaara, K., Huning, M., Ischebeck, R., Jablonka, M., Kamps, T., Korfer, M., Krassilnikov, M., Krzywinski, J., Liepe, M., Liero, A., Limberg, T., Loos, H., Luong, M., Magne, C., Menzel, J., Michelato, P., Minty, M., Muller, U.C., Nolle, D., Novokhatski, A., Pagani, C., Peters, F., Pfluger, J., Piot, P., Plucinski, L., Rehlich, K., Reyzl, I., Richter, A., Rossbach, J., Saldin, E.L., Sandner, W., Schlarb, H., Schmidt, G., Schmuser, P., Schneider, J.R., Schneidmiller, E.A., Schreiber, H.J., Schreiber, S., Sertore, D., Setzer, S., Simrock, S., Sobierajski, R., Sonntag, B., Steeg, B., Stephan, F., Sytchev, K.P., Tiedtke, K., Tonutti, M., Treusch, R., Trines, D., Turke, D., Verzilov, V., Wanzenberg, R., Weiland, T., Weise, H., Wendt, M., Will, I., Wolff, S., Wittenburg, K., Yurkov, M.V., and Zapfe, K. Generation of GW radiation pulses from a vuv free-electron laser operating in the femtosecond regime. *Physical Review Letters* **88**(10) (2002).

Barty, A., C., C., A., A., N., T., L., L., White, T., Andreasson, J., Arnlund, D., Bajt, S., Barends, T., Barthelmess, M., Bogan, M., Bosted, C., Boutet, S., Bozek, J., Coffee, R., Coppola, N., Davidsson, J., DePonte, D., Doak, R., Ekeberg, T., Elser, V., Epp, S., Erk, B., Fleckenstein, H., Foucar, L., Fromme, P., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C., Hartmann, R., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Holton, J., Hunter, M., Johansson, L., Kassemeyer, S., Kimmel, N., Kirian, R., Liang, M., Maia, F., Malmerberg, E., Marchesini, S. Martin, A., Nass, K., Neutze, R., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Scott, H., Schlichting, I., Schulz, J., Seibert, M., Shoeman, R., Sierra, R., Soltau, H., Spence, J., Stellato, F., Stern, S., Strder, L., Ullrich, J., Wang, X., Weidenspointner, G., Weierstall, U., Wunderer, C., and Chapman, H. Self-terminating diffraction in femtosecond X-ray nanocrystallography reduces measured radiation damage. *in preparation* (2011).

Bergh, M., Huldt, G., Timneanu, N., Maia, F., and Hajdu, J. Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction. *Quarterly*

Reviews of Biophysics **41**(3-4) 181–204 (2008).

Bergh, M., Timneanu, N., and van der Spoel, D. Model for the dynamics of a water cluster in an X-ray free electron laser beam. *Physical Review E* **70**(5) (2004).

Bilderback, D.H., Brock, J.D., Dale, D.S., Finkelstein, K.D., Pfeifer, M.A., and Gruner, S.M. Energy recovery linac (ERL) coherent hard X-ray sources. *New Journal of Physics* **12** (2010).

Bogan, M.J., Benner, W.H., Boutet, S., Rohner, U., Frank, M., Barty, A., Seibert, M.M., Maia, F., Marchesini, S., Bajt, S., Woods, B., Riot, V., Hau-Riege, S.P., Svenda, M., Marklund, E., Spiller, E., Hajdu, J., and Chapman, H.N. Single particle X-ray diffractive imaging. *Nano Letters* **8**(1) 310–316 (2008).

Bonifacio, R., Pellegrini, C., and Narducci, L.M. Collective instabilities and high-gain regime in a free-electron laser. *Optics Communications* **50**(6) 373–378 (1984).

Boutet, S. and Williams, G.J. The coherent X-ray imaging (CXI) instrument at the linac coherent light source (LCLS). *New Journal of Physics* **12** (2010).

Bozek, J.D. AMO instrumentation for the LCLS X-ray fel. *European Physical Journal-Special Topics* **169** 129–132 (2009).

Burmeister, W.P. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallographica Section D-Biological Crystallography* **56** 328–341 (2000).

Caleman, C., Huldt, G., Maia, F., Ortiz, C., Parak, F.G., Hajdu, J., van der Spoel, D., Chapman, H.N., and Timneanu, N. On the feasibility of nanocrystal imaging using intense and ultrashort X-ray pulses. *Acs Nano* **5**(1) 139–146 (2011).

Chapman, H., Ullrich, J., and Rost, J.M. Intense X-ray science: the first 5 years of FLASH. *Journal of Physics B-Atomic Molecular and Optical Physics* **43**(19) (2010).

Chapman, H.N. X-ray imaging beyond the limits. *Nature Materials* **8**(4) 299–301 (2009).

Chapman, H.N., Barty, A., Bogan, M.J., Boutet, S., Frank, M., Hau-Riege, S.P., Marchesini, S., Woods, B.W., Bajt, S., Benner, H., London, R.A., Plonjes, E., Kuhlmann, M., Treusch, R., Dusterer, S., Tschentscher, T., Schneider, J.R., Spiller, E., Moller, T., Bostedt, C., Hoener, M., Shapiro, D.A., Hodgson, K.O., Van der Spoel, D., Burmeister, F., Bergh, M., Caleman, C., Huldt, G., Seibert, M.M., Maia, F., Lee, R.W., Szoke, A., Timneanu, N., and Hajdu, J. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Physics* **2**(12) 839–843 (2006).

Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., DePonte, D.P., Weierstall, U., Doak, R.B., Maia, F., Martin, A.V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R.L., Epp, S.W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M.N., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M.J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnel, K.U., Messerschmidt, M., Bozek, J.D., Hau-Riege, S.P., Frank, M., Hampton, C.Y., Sierra, R.G., Starodub, D., Williams, G.J., Hajdu, J., Timneanu, N., Seibert, M.M., Andreasson, J., Rocker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schroter, C.D., Krasniqi, F., Bott, M., Schmidt, K.E., Wang, X.Y., Grotjohann, I., Holton, J.M., Barends, T.R.M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., and Spence, J.C.H. Femtosecond X-ray protein nanocrystallography. *Nature* **470**(7332) 73–U81 (2011).

Chapman, H.N., Hau-Riege, S.P., Bogan, M.J., Bajt, S., Barty, A., Boutet, S., Marchesini, S., Frank, M., Woods, B.W., Benner, W.H., London, R.A., Rohner, U., Szoke, A., Spiller, E., Moller, T., Bostedt, C., Shapiro, D.A., Kuhlmann, M., Treusch, R., Plonjes, E., Burmeister, F., Bergh, M., Caleman, C., Huldt, G., Seibert, M.M., and Hajdu, J. Femtosecond time-delay X-ray holography. *Nature* **448**(7154) 676–679 (2007).

Chapman, H.N. Personal communication (2011).

Cherezov, V. and Caffrey, M. Membrane protein crystallization in lipidic mesophases. a mechanism study using X-ray microdiffraction. *Faraday Discuss* **136** 195–212; discussion 213–29 (2007).

Cherezov, V., Clogston, J., Papiz, M.Z., and Caffrey, M. Room to move: Crystallizing membrane proteins in swollen lipidic mesophases. *Journal of Molecular Biology* **357**(5) 1605–1618 (2006).

Cherezov, V. Microcrystals in membrane protein crystallization screens (2011).

Chien, E.Y.T., Liu, W., Zhao, Q.A., Katritch, V., Han, G.W., Hanson, M.A., Shi, L., Newman, A.H., Javitch, J.A., Cherezov, V., and Stevens, R.C. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330**(6007) 1091–1095 (2010).

Cong, W.T., Hwang, S.Y., Jin, L.T., He, H.Z., and Choi, J.K. High-throughput negative detection of SDS-PAGE separated proteins and its application for proteomics. *Electrophoresis* **31**(2) 411–420 (2010).

Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A., and Riekel, C.

Small is beautiful: protein micro-crystallography. *Nature Structural Biology* **5** 634 (1998).

DePonte, D.P., Weierstall, U., Schmidt, K., Warner, J., Starodub, D., Spence, J.C.H., and Doak, R.B. Gas dynamic virtual nozzle for generation of microscopic droplet streams. *Journal of Physics D-Applied Physics* **41**(19) (2008).

Doniach, S. Studies of the structure of matter with photons from an X-ray free-electron laser. *Journal of Synchrotron Radiation* **3** 260–267 (1996).

Drenth, J. *Principles of Protein X-ray Crystallography*. Springer Science+Business Media, LLC, New York, 3rd edition (2007).

Duisenberg, A.J.M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *Journal of Applied Crystallography* **25** 92–96 (1992).

Edward, J.T. Molecular volumes and Stokes-Einstein equation. *Journal of Chemical Education* **47**(4) 261–270 (1970).

Emma, P., Akre, R., Arthur, J., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P., Coffee, R., Decker, F.J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gilevich, S., Hastings, J., Hays, G., Hering, P., Huang, Z., Iverson, R., Loos, H., Messerschmidt, M., Miahnahri, A., Moeller, S., Nuhn, H.D., Pile, G., Ratner, D., Rzepiela, J., Schultz, D., Smith, T., Stefan, P., Tompkins, H., Turner, J., Welch, J., White, W., Wu, J., Yocky, G., and Galayda, J. First lasing and operation of an angstrom-wavelength free-electron laser. *Nature Photonics* **4**(9) 641–647 (2010).

Fischer, N., Hippler, M., Setif, P., Jacquot, J.P., and Rochaix, J.D. The Psac subunit of Photosystem I provides an essential lysine residue for fast electron transfer to ferredoxin. *Embo Journal* **17**(4) 849–858 (1998).

Fromme, P., Bottin, H., Krauss, N., and Setif, P. Crystallization and electron paramagnetic resonance characterization of the complex of Photosystem I with its natural electron acceptor ferredoxin. *Biophysical Journal* **83**(4) 1760–1773 (2002).

Fromme, P. and Mathis, P. Unraveling the Photosystem I reaction center: a history, or the sum of many efforts. *Photosynthesis Research* **80**(1-3) 109–124 (2004).

Fromme, P., Melkozernov, A., Jordan, P., and Krauss, N. Structure and function of Photosystem I: interaction with its soluble electron carriers and external antenna systems. *Febs Letters* **555**(1) 40–44 (2003).

Fromme, P., Schubert, W.D., and Krauss, N. Structure of Photosystem-I - suggestions on the docking sites for plastocyanin, ferredoxin and the coordination of P700. *Biochimica Et Biophysica Acta-Bioenergetics* **1187**(2) 99–105 (1994).

Fromme, P. and Witt, H.T. Improved isolation and crystallization of Photosystem I for structural analysis. *Biochimica Et Biophysica Acta-Bioenergetics* **1365**(1-2) 175–184 (1998).

Fromme, P., editor. *Photosynthetic Protein Complexes*. Wiley-Blackwell (2008).

Fromme, P. and Fromme., R. High resolution structure of Photosystem I complexed with ferredoxin (2011).

Gabrielsen, M., Gardiner, A.T., Fromme, P., and Cogdell, R.J. *Membrane Protein Crystallization: Approaching the Problem and Understanding the Solutions*, volume 63 of *Current Topics in Membranes*, pages 127–149. Elsevier Academic Press Inc, San Diego (2009).

Gabrielsen, M., Nagy, L.A., DeLucas, L.J., and Cogdell, R.J. Self-interaction chromatography as a tool for optimizing conditions for membrane protein crystallization. *Acta Crystallographica Section D-Biological Crystallography* **66** 44–50 (2010).

Geloni, G., Saldin, E., SAMOylova, L., Schneidmiller, E., Sinn, H., Tschentscher, T., and Yurkov, M. Coherence properties of the European XFEL. *New Journal of Physics* **12** (2010).

Gnodtke, C., Saalmann, U., and Rost, J.M. Ionization and charge migration through strong internal fields in clusters exposed to intense X-ray pulses. *Physical Review A* **79**(4) (2009).

Gonzalez, A., Thompson, A.W., and Nave, C. Cry-protection of protein crystals in intense X-ray-beams. *Review of Scientific Instruments* **63**(1) 1177–1180 (1992).

Haas, D.J. and Rossmann, M.G. Crystallographic studies on lactate dehydrogenase at -75 degrees C. *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry* **B 26** 998–& (1970).

Hau-Riege, S.P., Boutet, S., Barty, A., Bajt, S., Bogan, M.J., Frank, M., Andreasson, J., Iwan, B., Seibert, M.M., Hajdu, J., Sakdinawat, A., Schulz, J., Treusch, R., and Chapman, H.N. Sacrificial tamper slows down sample explosion in FLASH diffraction experiments. *Physical Review Letters* **104**(6) (2010).

Hau-Riege, S.P., London, R.A., Chapman, H.N., and Bergh, M. Soft-X-ray free-electron-laser interaction with materials. *Physical Review E* **76**(4) (2007).

Hau-Riege, S.P., London, R.A., and Szoke, A. Dynamics of biological molecules irradiated by short X-ray pulses. *Physical Review E* **69**(5) (2004).

Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and X-ray-diffraction. *Proceedings of the Royal Society of London Series*

B-Biological Sciences **241**(1300) 6–8 (1990).

Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quart. Rev. Biophys.* **28** 171–193 (1995).

Hoganson, C.W. and Babcock, G.T. A metalloradical mechanism for the generation of oxygen from water in photosynthesis. *Science* **277**(5334) 1953–1956 (1997).

Holroyd, R. *Fundamental processes in radiation chemistry*. Interscience, New York, 1st edition (1968).

Holton, J.M. A beginner's guide to radiation damage. *Journal of Synchrotron Radiation* **16** 133–142 (2009).

Holton, J.M. and Frankel, K.A. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D-Biological Crystallography* **66** 393–408 (2010).

Hope, H. Cryocrystallography of biological macromolecules - a generally applicable method. *Acta Crystallographica Section B-Structural Science* **44** 22–26 (1988).

Horsefield, R., Norden, K., Fellert, M., Backmark, A., Tornroth-Horsefield, S., van Scheltinga, A.C.T., Kvassman, J., Kjellbom, P., Johanson, U., and Neutze, R. High-resolution X-ray structure of human aquaporin 5. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* **105**(36) 13327–13332 (2008).

Howells, M.R., Beetz, T., Chapman, H.N., Cui, C., Holton, J.M., Jacobsen, C.J., Kirz, J., Lima, E., Marchesini, S., Miao, H., Sayre, D., Shapiro, D.A., Spence, J.C.H., and Starodub, D. An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy. *Journal of Electron Spectroscopy and Related Phenomena* **170**(1-3) 4–12 (2009).

Hunter, M.S., DePonte, D.P., Shapiro, D.A., Kirian, R.A., Wang, X., Starodub, D., Marchesini, S., Weierstall, U., Doak, R.B., Spence, J.C.H., and Fromme, P. X-ray diffraction from membrane protein nanocrystals. *Biophysical Journal* **100**(1) 198–206 (2011).

Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauss, N. Three-dimensional structure of cyanobacterial Photosystem I at 2.5 angstrom resolution. *Nature* **411**(6840) 909–917 (2001).

Kam, Z., Shore, H.B., and Feher, G. Crystallization of proteins. *Journal of Molecular Biology* **123**(4) 539–555 (1978).

Kirian, R.A., Wang, X.Y., Weierstall, U., Schmidt, K.E., Spence, J.C.H., Hunter, M.,

Fromme, P., White, T., Chapman, H.N., and Holton, J. Femtosecond protein nanocrystallography-data analysis methods. *Optics Express* **18**(6) 5713–5723 (2010).

Kirian, R.A., White, T.A., Holton, J.M., Chapman, H.N., Fromme, P., Barty, A., Lomb, L., Aquila, A., Maia, F., Martin, A.V., Fromme, R., Wang, X.Y., Hunter, M.S., Schmidt, K.E., and Spence, J.C.H. Structure-factor analysis of femtosecond micro-diffraction patterns from protein nanocrystals. *Acta Crystallographica Section A* **67** 131–140 (2011).

Klukas, O., Schubert, W.D., Jordan, P., Krauss, N., Fromme, P., Witt, H.T., and Saenger, W. Localization of two phylloquinones, Q(K) and Q(K)', in an improved electron density map of Photosystem I at 4-angstrom resolution. *Journal of Biological Chemistry* **274**(11) 7361–7367 (1999a).

Klukas, O., Schubert, W.D., Jordan, P., Krauss, N., Fromme, P., Witt, H.T., and Saenger, W. Photosystem I, an improved model of the stromal subunits Psac, Psad, and Psae. *Journal of Biological Chemistry* **274**(11) 7351–7360 (1999b).

Koerner, L.J., Philipp, H.T., Hromalik, M.S., Tate, M.W., and Gruner, S.M. X-ray tests of a Pixel Array Detector for coherent X-ray imaging at the Linac Coherent Light Source. *Journal of Instrumentation* **4** (2009).

Krauss, N., Hinrichs, W., Witt, I., Fromme, P., Pritzkow, W., Dauter, Z., Betzel, C., Wilson, K.S., Witt, H.T., and Saenger, W. 3-dimensional structure of system-I of photosynthesis at 6 angstrom resolution. *Nature* **361**(6410) 326–331 (1993).

Krauss, N., Schubert, W.D., Klukas, O., Fromme, P., Witt, H.T., and Saenger, W. Photosystem I at 4 angstrom resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna, system. *Nature Structural Biology* **3**(11) 965–973 (1996).

Kurisu, G., Zhang, H., Smith, J.L., and Cramer, W.A. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**(5647) 1009–14 (2003).

Leslie, A.G.W. The integration of macromolecular diffraction data. *Acta Crystallographica Section D-Biological Crystallography* **62** 48–57 (2006).

Liu, Z., Yan, H., Wang, K., Kuang, T., Zhang, J., Gui, L., An, X., and Chang, W. Crystal structure of spinach major light-harvesting complex at 2.72-Å resolution. *Nature* **428**(6980) 287–292 (2004).

Loll, B., Kern, J., Saenger, W., Zouni, A., and Biesiadka, J. Towards complete cofactor arrangement in the 3.0 angstrom resolution structure of Photosystem II. *Nature* **438**(7070) 1040–1044 (2005).

Malkin, A.J., Kuznetsov, Y.G., Land, T.A., Deyoreo, J.J., and McPherson, A. Mechanisms of growth for protein and virus crystals. *Nature Structural Biology* **2**(11) 956–959 (1995).

Malkin, A.J. and Thorne, R.E. Growth and disorder of macromolecular crystals: insights from atomic force microscopy and X-ray diffraction studies. *Methods* **34**(3) 273–299 (2004).

Marchesini, S., He, H., Chapman, H.N., Hau-Riege, S.P., Noy, A., Howells, M.R., Weierstall, U., and Spence, J.C.H. X-ray image reconstruction from a diffraction pattern alone. *Physical Review B* **68**(14) (2003).

Margiolaki, I. and Wright, J. Powder crystallography on macromolecules. *Aact Crystallographica Section A* **64** 169–180 (2008).

Matthews, B.W. Solvent content of protein crystals. *Journal of Molecular Biology* **33**(2) 491–& (1968).

McPherson, A. Introduction to protein crystallization. *Methods* **34**(3) 254–265 (2004).

McPherson, A., Malkin, A.J., and Kuznetsov, Y.G. The science of macromolecular crystallization. *Structure* **3**(8) 759–768 (1995).

Meents, A., Gutmann, S., Wagner, A., and Schulze-Briese, C. Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proceedings of the National Academy of Sciences of the United States of America* **107**(3) 1094–1099 (2010).

Messinger, J. Evaluation of different mechanistic proposals for water oxidation in photosynthesis on the basis of Mn4OxCa structures for the catalytic site and spectroscopic data. *Physical Chemistry Chemical Physics* **6**(20) 4764–4771 (2004).

Mueller, M., Jenni, S., and Ban, N. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Current Opinion in Structural Biology* **17**(5) 572–579 (2007).

Muller, M.G., Slavov, C., Luthra, R., Redding, K.E., and Holzwarth, A.R. Independent initiation of primary electron transfer in the two branches of the Photosystem I reaction center. *Proceedings of the National Academy of Sciences of the United States of America* **107**(9) 4123–4128 (2010).

Murray, J.W., Garman, E.F., and Ravelli, R.B.G. X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *Journal of Applied Crystallography* **37** 513–522 (2004).

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section*

D-Biological Crystallography **53** 240–255 (1997).

Nave, C. A description of imperfections in protein crystals. *Acta Crystallographica Section D-Biological Crystallography* **54** 848–853 (1998).

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., and Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**(6797) 752–757 (2000).

Ostermeier, C. and Michel, H. Crystallization of membrane proteins. *Current Opinion in Structural Biology* **7**(5) 697–701 (1997).

Owen, R.L., Rudino-Pinera, E., and Garman, E.F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proceedings of the National Academy of Sciences of the United States of America* **103**(13) 4912–4917 (2006).

Paithankar, K.S., Owen, R.L., and Garman, E.F. Absorbed dose calculations for macromolecular crystals: improvements to RADDOSE. *Journal of Synchrotron Radiation* **16** 152–162 (2009).

Parkin, S. and Hope, H. Macromolecular cryocrystallography: Cooling, mounting, storage and transportation of crystals. *Journal of Applied Crystallography* **31** 945–953 (1998).

Perrakis, A., Cipriani, F., Castagna, J.C., Claustre, L., Burghammer, M., Riekel, C., and Cusack, S. Protein microcrystals and the design of a microdiffractometer: current experience and plans at EMBL and ESRF/ID13. *Acta Crystallographica Section D-Biological Crystallography* **55** 1765–1770 (1999).

Perrin, F. Brownian movement of an ellipsoid (II). - free rotation and depolarisation of fluourescences. - translation and diffusion of ellipsoidal molecules. *Journal De Physique Et Le Radium* **7** 1–11 (1936).

Petsko, G.A. Protein crystallography at sub-zero temperatures - cryo-protective mother liquors for protein crystals. *Journal of Molecular Biology* **96**(3) 381–& (1975).

Popot, J.L., Berry, E.A., Charvolin, D., Creuzenet, C., Ebel, C., Engelman, D.M., Flotenmeyer, M., Giusti, F., Gohon, Y., Herve, P., Hong, Q., Lakey, J.H., Leonard, K., Shuman, H.A., Timmins, P., Warschawski, D.E., Zito, F., Zoonens, M., Pucci, B., and Tribet, C. Amphipols: polymeric surfactants for membrane biology research. *Cellular and Molecular Life Sciences* **60**(8) 1559–1574 (2003).

Porra, R.J., Thompson, W.A., and Kriedemann, P.E. Determination of accurate extinction coefficients and simultaneous-equations for assaying chlorophyll-a and chlorophyll-b extracted with 4 different solvents - verification of the concentration of chlorophyll standards by atomic-absorption spectroscopy. *Biochimica Et*

Biophysica Acta **975**(3) 384–394 (1989).

Prive, G.G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**(4) 388–397 (2007).

Pushkar, Y.L., Yano, J., Sauer, K., Boussac, A., and Yachandra, V.K. Structural changes in the Mn4Ca cluster and the mechanism of photosynthetic water splitting. *Proceedings of the National Academy of Sciences of the United States of America* **105**(6) 1879–1884 (2008).

Quiney, H.M. and Nugent, K.A. Biomolecular imaging and electronic damage using X-ray free-electron lasers. *Nature Physics* **7**(2) 142–146 (2011).

Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., Schertler, G.F., Weis, W.I., and Kobilka, B.K. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**(7168) 383–7 (2007).

Riekel, C., Burghammer, M., and Schertler, G. Protein crystallography microdiffraction. *Current Opinion in Structural Biology* **15**(5) 556–562 (2005).

Robinson, I. and Harder, R. Coherent X-ray diffraction imaging of strain at the nanoscale. *Nature Materials* **8**(4) 291–298 (2009).

Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Yao, X.J., Weis, W.I., Stevens, R.C., and Kobilka, B.K. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* **318**(5854) 1266–73 (2007).

Rousseau, F., Setif, P., and Lagoutte, B. Evidence for the involvement of PSI-E subunit in the reduction of ferredoxin by Photosystem-I. *Embo Journal* **12**(5) 1755–1765 (1993).

Rupp, B. *Biomolecular Crystallography*. Garland Science, New York (2010).

Sayre, D. Some implications of a theorem due to shannon. *Acta Crystallographica* **5**(6) 843–843 (1952).

Schagger, H. Tricine-SDS-PAGE. *Nature Protocols* **1**(1) 16–22 (2006).

Scherrer, P. Determination of the size and inner structure of colloid particles by X-rays. *Nach. Ges. Wiss. Gttingen* **26** 2 (1918).

Scott, H.A. Cretin - a radiative transfer capability for laboratory plasmas. *Journal of Quantitative Spectroscopy & Radiative Transfer* **71**(2-6) 689–701 (2001).

Seibert, M.M., Boutet, S., Svenda, M., Ekeberg, T., Maia, F., Bogan, M.J., Timneanu,

N., Barty, A., Hau-Riege, S., Caleman, C., Frank, M., Benner, H., Lee, J.Y., Marchesini, S., Shaevitz, J.W., Fletcher, D.A., Bajt, S., Andersson, I., Chapman, H.N., and Hajdu, J. Femtosecond diffractive imaging of biological cells. *Journal of Physics B-Atomic Molecular and Optical Physics* **43**(19) (2010).

Setif, P.Q.Y. and Bottin, H. Laser flash absorption-spectroscopy study of ferredoxin reduction by Photosystem-i in synechocystis SP PCC-6803 - evidence for submicrosecond and microsecond kinetics. *Biochemistry* **33**(28) 8495–8504 (1994).

Setif, P.Q.Y. and Bottin, H. Laser FLASH absorption-spectroscopy study of ferredoxin reduction by Photosystem-I - spectral and kinetic evidence for the existence of several Photosystem-I ferredoxin complexes. *Biochemistry* **34**(28) 9059–9070 (1995).

Shapiro, D.A., Chapman, H.N., DePonte, D., Doak, R.B., Fromme, P., Hembree, G., Hunter, M., Marchesini, S., Schmidt, K., Spence, J., Starodub, D., and Weierstall, U. Powder diffraction from a continuous microjet of submicrometer protein crystals. *Journal of Synchrotron Radiation* **15** 593–599 (2008).

Shintake, T., Tanaka, T., Hara, T., Togawa, K., Inagaki, T., Kim, Y.J., Ishikawa, T., Kitamura, H., Baba, H., Matsumoto, H., Takeda, S., Yoshida, M., and Takasu, Y. Status of SPring-8 compact SASV source FEL project. *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment* **507**(1-2) 382–387 (2003).

Snell, E. Personal communication (2011).

Spence, J.C.H. and Doak, R.B. Single molecule diffraction. *Physical Review Letters* **92**(19) (2004).

Spence, J., Kirian, R., Wang, X., Weierstall, U., Schmidt, K., White, T., Barty, A., Chapman, H., Marchesini, S., and Holton, J. Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Optics Express* **19** 2866–2873 (2011).

Steller, I., Bolotovsky, R., and Rossmann, M.G. An algorithm for automatic indexing of oscillation images using fourier analysis. *Journal of Applied Crystallography* **30** 1036–1040 (1997).

Struder, L., Eppa, S., Rolles, D., Hartmann, R., Holl, P., Lutz, G., Soltau, H., Eckart, R., Reich, C., Heinzinger, K., Thamm, C., Rudenko, A., Krasniqi, F., Kuhnel, K.U., Bauer, C., Schroter, C.D., Moshammer, R., Techert, S., Miessner, D., Porro, M., Halker, O., Meidinger, N., Kimmel, N., Andritschke, R., Schopper, F., Weidenspointner, G., Ziegler, A., Pietschner, D., Herrmann, S., Pietsch, U., Walenta, A., Leitenberger, W., Bostedt, C., Moller, T., Rupp, D., Adolph, M., Graafsma, H., Hirsemann, H., Gartner, K., Richter, R., Foucar, L., Shoeman, R.L., Schlichting, I., and Ullrich, J. Large-format, high-speed, X-ray pnCCDs combined

with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources. *Nuclear Instruments & Methods in Physics Research Section a-Accelerators Spectrometers Detectors and Associated Equipment* **614**(3) 483–496 (2010).

Teng, T.Y., Srajer, V., and Moffat, K. Initial trajectory of carbon monoxide after photodissociation from myoglobin at cryogenic temperatures. *Biochemistry* **36**(40) 12087–12100 (1997).

Tommos, C. and Babcock, G.T. Oxygen production in nature: A light-driven metalloradical enzyme process. *Accounts of Chemical Research* **31**(1) 18–25 (1998).

Umena, Y., Kawakami, K., Shen, J.R., and Kamiya, N. Crystal structure of oxygen-evolving Photosystem II at a resolution of 1.9 Å. *Nature* **advance online publication** (2011).

Wagner, W. and Pruss, A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *Journal of Physical and Chemical Reference Data* **31**(2) 387–535 (2002).

Wampler, R.D., Moad, A.J., Moad, C.W., Heiland, R., and Simpson, G.J. Visual methods for interpreting optical nonlinearity at the molecular level. *Accounts of Chemical Research* **40**(10) 953–960 (2007).

Warkentin, M. and Thorne, R.E. Slow cooling of protein crystals. *Journal of Applied Crystallography* **42** 944–952 (2009).

Warren, B. *X-ray Diffraction*. Cambridge University Press, Dover, 2nd edition (1990).

White, S.H. The progress of membrane protein structure determination. *Protein Science* **13**(7) 1948–1949 (2004).

Witt, H.T., Witt, I., Krauss, N., Hinrichs, W., Fromme, P., and Saenger, W. Crystals and structure of Photosystem-1. *Biophysical Journal* **66**(2) A2–A2 (1994).

Witt, I., Witt, H.T., Difiore, D., Rogner, M., Hinrichs, W., Saenger, W., Granzin, J., Betzel, C., and Dauter, Z. X-ray characterization of single-crystals of the reaction center-I of water splitting photosynthesis. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics* **92**(12) 1503–1506 (1988).

Woolfson, M. *An introduction to X-ray crystallography*. Cambridge University Press, Cambridge, 2nd edition (1997).

Yano, J., Kern, J., Irrgang, K.D., Latimer, M.J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A., and Yachandra, V.K.

X-ray damage to the Mn4Ca complex in single crystals of Photosystem II: A case study for metalloprotein crystallography. *Proceedings of the National Academy of Sciences of the United States of America* **102**(34) 12047–12052 (2005).

Yano, J. and Yachandra, V.K. Where water is oxidized to dioxygen: Structure of the photosynthetic mn4ca cluster from X-ray spectroscopy. *Inorganic Chemistry* **47**(6) 1711–1726 (2008).

Young, L., Kanter, E.P., Krassig, B., Li, Y., March, A.M., Pratt, S.T., Santra, R., Southworth, S.H., Rohringer, N., DiMauro, L.F., Doumy, G., Roedig, C.A., Berrah, N., Fang, L., Hoener, M., Bucksbaum, P.H., Cryan, J.P., Ghimire, S., Glownia, J.M., Reis, D.A., Bozek, J.D., Bostedt, C., and Messerschmidt, M. Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466**(7302) 56–U66 (2010).

Zouni, A., Witt, H.T., Kern, J., Fromme, P., Krauss, N., Saenger, W., and Orth, P. Crystal structure of photosystem II from synechococcus elongatus at 3.8 angstrom resolution. *Nature* **409**(6821) 739–743 (2001).

Appendix A

THERMOSYNECHOCOCCUS ELONGATUS GROWTH MEDIA

Micronutrient solution:	0.5 mL of concentrated H_2SO_4 solution, 2.28 g of $MnSO_4 \cdot H_2O$, 0.5 g of $ZnSO_4 \cdot 7H_2O$, 0.5 g H_3BO_3, 0.025 g $CuSO_4 \cdot 5H_2O$, 0.025 g $Na_2MoO_4 \cdot 2H_2O$, 0.045 g $CoCl_2 \cdot 6H_2O$, filled to 1 L with millipure water
$FeCl_3$ solution:	0.2905 g $FeCl_3$ filled to 1 L with millipure water
Medium D10 (10X):	1 g nitriloacetic acid, 0.8 g $CaCl_2 \cdot 6H_2O$, 1.0 g $MgCl_2 \cdot 7H_2O$, 0.08 g NaCl, 5 mL micronutrient solution, 10 mL $FeCl_3$, 1.03 gKNO_3, 6.89 g $NaNO_3$, fill to 1 L with millipure water
Phosphate stock:	(1.11 g Na_2HPO_4, filled to 1 L with millipure water

The *T. elongatus* cells are grown in a media prepared by adding the 10X stocks, adjusting the pH to 8.2, and diluting to volume using DI water.