EXPLORATION OF METHODS FOR SERIAL MICROCRYSTALLOGRAPHY AT STORAGE RING X-RAY SOURCES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jennifer Lynn Wierman

January 2017

EXPLORATION OF METHODS FOR SERIAL MICROCRYSTALLOGRAPHY AT STORAGE RING X-RAY SOURCES

Jennifer Lynn Wierman, Ph.D.

Cornell University 2017

Protein crystallography has made the largest contribution to our knowledge of protein structure. However, it is well known that many biologically important proteins do not readily form large enough crystals for traditional crystallography. Successful serial microcrystallography (SMX) studies have been performed at X-ray Free-Electron Lasers, but they are limited in experimental availability. We look to more accessible light sources, such as storage ring sources, for the development of SMX. However, improvements to the conventional experiment are required to make SMX viable at storage rings. Here, we explore several devices and techniques designed given this consideration.

To isolate crystal diffraction, the sample environment should contribute zero background scatter outside the crystal, since excess scatter obscures the weak microcrystal signal. We will show the feasibility of using atomically-thin, gas-tight graphene to reduce background scatter as a crystal mount and suggest using it as a window material for SMX. We will also explore two microcrystal delivery devices, a microfluidic chip and a viscous jet injector, for use in SMX. While both of these devices show promise for optimizing various aspects of the crystal delivery system, both contribute more background scatter than is acceptable for an ideal SMX experiment at a storage ring source.

Merging diffraction from multiple microcrystals is necessary when a complete data set cannot be determined from a single microcrystal. When micro-

crystal diffraction is weak enough that Bragg peaks are no longer visible, merging through conventional techniques fails since crystal orientation cannot be obtained through Bragg peak indexing. We will explore proof-of-principle experiments which show that indexing data frames on a per-frame basis is unnecessary for a structure solution, when reciprocal space intensities can be reconstructed using the EMC algorithm.

In principle, serial microcrystallography is feasible at storage ring sources if improvements in beamline setups, sample chamber construction and microcrystal diffraction analysis evolve to optimize the diffraction of microcrystals.

BIOGRAPHICAL SKETCH

Jennifer Wierman was born in Pendleton, Oregon in 1985. She pursued her B.A. in biology from Seattle Pacific University, received her B.S. in physics from the University of Washington in May 2010, and started graduate studies in biophysics at Cornell University the following fall.

To my best friend and partner in life, Adam, who chooses to adventure with me through this all.

ACKNOWLEDGEMENTS

My graduate adviser, Sol Gruner, has guided me with wisdom, patience and grace. He encourages me to reconsider problems with a physicist mind, and analyze every aspect of my work down to fundamentals. More, he expertly navigates just about every subject, including how to enjoy life outside the laboratory. I am fortunate and grateful to have the opportunity to work with him for these past six years. Many members of the Gruner group, past and present, deserve thanks for their support and company. Hugh Philipp, Mark Tate, and Marty Novak have been instrumental in organizing, machining and executing experiments, including helping with data collection and analysis. Undergraduate student Gabrielle Illava continued my graphene project with enthusiasm, and I couldn't have put it into better hands. I'm thankful for the support of Joel Weiss and Julian Becker, as they are always willing to lend a hand, an eye, or an ear when l need it. As the Gruner Group Gals, Kate Shanks, Veronica Pillar and Adrian Baur have been a part of my support system within the larger Gruner group (though we never got our GGG t-shirts!). I would not have been as successful as a graduate student if not for the mentorship of Kate Shanks, and I'm honored to be considered her friend. From APS beamtime to short story book clubs, she's always provided a wonderful example of how to be a successful scientist and a friend.

I'd like to thank my special committee members, Lois Pollack and Brian Crane. They have guided me since my first year in my graduate career. From every meeting with them, I am encouraged to pursue my goals, with insights to new solutions and directions to follow. Veit Elser, Ti-Yen Lan, and Kartik Ayyer form the pivotal collaboration which has guided much of the work contained in this dissertation. Ti-Yen, in particular, has worked very closely with me to

produce the protein crystal proof-of-principle experiments in this dissertation. Without him, this work would've floundered long ago. They all deserve many thanks and I look forward to more work with them!

My graduate career could be defined by nearly constant collaborations. From my first collaboration in graphene with Jonathan Alder and Paul McEuen, to microfluidic chips with Michael Heymann, Achini Opthalage and Seth Fraden, I'm grateful for the excitement, experience and willingness to share projects with me that came with each new experiment. Beamtime runs have been more entertaining and educational by their sides. I am also thankful for various collaborations with John Spence, Uwe Weierstall, Nadia Zatsepin, Garrett Nelson, Dan James, Chelsie Conrad and Bob Fischetti. Though short in duration, they provided me with more experience and skill that I would've earned on my own.

Macromolecular Diffraction at CHESS (MacCHESS) staff have been significant in teaching me not only the basics of crystallography, but in allowing me to tinker away in their hutches. I can't thank Marian Szebenyi enough for the tutoring she's given me over the years. At early stages of my graduate career, Chae Un Kim provided pivotal mentorship and life guidance while he was still at MacCHESS. In addition to laughs and fun outside of the hutches, Irina Krisunokov, Richard Gillilan, Jesse Hopkins, Bill Miller, Scott Smith, and Tiit Luuk formed the crucial support team for any project I was working on. My right-hand man, Mike Cook, is the most supportive and wonderful partner in experiment and in crime. His hard work and dedication have been a foundation for much of my experimental work. Between the two of us, we can conquer the world (or so we think).

The fantastic research environment at Cornell High Energy Synchrotron

Source (CHESS) is due to the CHESS staff. Detector master Zak Brown, x-ray capillary wizard Rong Huang, and beam line rocket scientists Aurther Woll, Darren Dale, and Jacob Ruff always welcomed me and worked diligently to provide me with the best CHESS could offer. My first beamline experience was guided by Alexander Kazimirov, who has since passed. The ease and gentleness with which he handled a bumbling graduate student in her first x-ray experiment is a goal of mine to honor.

I've spent some time in the company of Elspeth Garman and James Holton. I am humbled by their breadth of knowledge and insight in our field, and their willingness to engage in deep discussion on radiation damage, emittance, or complex thinking. In each conversation, I've come away not only knowing much more about my field, but understanding just how little I already know.

This dissertation is based partly on research conducted at the CHESS, which is supported by the National Science Foundation and the National Institutes of Heath/National Institute of General Medical Sciences under NSF award DMR-1332208, and using the MacCHESS facility, which is supported by award GM-103485 from the National Institute of General Medical Sciences, National Institutes of Health. In my early graduate years, I was funded by the Biophysics Training Grant under NIH award GM-008267.

Before the Gruner Lab, my first rotation at Cornell was in the lab of Michelle Wang and I'm grateful for all the skills she and others in her lab have taught me. Michael Hall and Ben Smith patiently tutored me in computational biophysics with optical traps, in addition to tricks in making the best coffee ever. My undergraduate years at the University of Washington ultimately shaped my path to Cornell. I worked in the nanopore DNA-sequencing group with Jens Gundlach and Ian Derrington. Their enthusiastic balance of work, biophysics, and play

encouraged me to find it in my own career. I'm thankful for Marcus Collins introducing me to crystallography and motivating me to email Sol. Michael Schick proved to be an excellent mentor, and his door remains open to this day for his sage advice and belly-laughs. Vladimir Chaloupka taught me the beauty of science in society and art, and my love for Bach would never have come about if he hadn't introduced me to fugues. Marjorie Olmstead and Steven Ellis were both influential instructors and incredible people. I never feared stopping by either of their offices for advice or company.

My family has always been the rock of my life and their unwavering love and support is an example for me everyday. Since I was young, my parents supported everything I wanted to do, including the multiple sports, music lessons, hobbies and interests. When I wanted to be and ER surgeon, they knew I could do it. When I wanted to be a biological physicist, they knew I could do it. They have supported every decision I've made. My particle-physicist brother, Kevin, has enjoyed the ups and downs of university careers with me. Single texts from him always brightened my day, and I'm so excited and proud to have walked along our respective physics-paths together.

Lastly, my best friend and partner in life, Adam, deserves more thanks than I can put into words. He met me after I decided to leave the West Coast for Ithaca, and joined me on the East Coast anyway. He watched me spiral into beamtimes, and caught me out the other end. He's inspired me to make every minute of the day count, to run on trails in the woods and on mountains, and to be the best version of myself. He's a daily reminder that the path in life you take is the one you choose. Thank you, Adam.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADE	Acoustic droplet ejection
ADSC	Area Detector Systems Corporation
AGIPD	Adaptive Gain Integrating Pixel Detector
ALS	Advanced Light Source
APS	Advanced Photon Source
AS	Australian Synchrotron
ASU	Arizona State University
BW	Bandwidth
CCD	Charge-coupled device
CHESS	Cornell High Energy Synchrotron Source
CLS	Canadian Light Source
COC	Cyclic olefin copolymer
CS-PAD	Cornell SLAC Pixel Array Detector
CVD	Chemical-vapor deposition
DESY	Deutsches Elektronen-Synchrotron
DLS	Diamond Light Source
DNA	Deoxyribonucleic acid
EMC	Expand-Maximize-Compress
ERL	Energy Recovery Linac
ESRF	European Synchrotron Radiation Facility
FLASH	Free-electron LASer in Hamburg
GDVN	Gas Dynamic Virtual Nozzle
GPCR	G-protein coupled receptor
HEWL	Hen egg white lysozyme
KB	Kirkpatrick-Baez

LCLS	Linac Coherent Light Source
LCP	Lipidic cubic phase
MacCHESS	Macromolecular Facility at CHESS
MESH	Microfluidic Electrokinetic Sample Holder
MM-PAD	Mixed-mode PAD
MR	Molecular replacement
MX	Macromolecular crystallography
NGS	Next generation light sources
PAD	Pixel array detector
PDB	Protein Data Bank
PDMS	Poly(dimethylsiloxane)
PEG	Polyethylene glycol
PETRA III	Positron-Electron Tandem Ring Accelerator III
PF	Photon Factory
PMMA	Poly(methyl methacrylate)
relp	Reciprocal lattice point
s.c.c.m	Standard cubic centimeters per minute
SACLA	SPring-8 Angstrom Compact Free Electron Laser
SASE	Self amplified spontaneous emission
SFX	Serial femtosecond crystallography
SHG	Second harmonic generation
SLAC	SLAC National Accelerator Laboratory
SLS	Swiss Light Source
SMX	Serial microcrystallography
SNR	Signal-to-noise ratio
SPring-8	Super Photon Ring

SR	Synchrotron radiation
SSRL	Stanford Synchrotron Radiation Light Source
STXL	Scanning transmission x-ray microscopy
TPEF	Two photon-excited fluorescence
UESR	Ultra-low-Emittance Storage Ring
UFXC32k	Ultra Fast X-ray Camera 32k
VD	Vapor diffusion
XFEL	X-ray free-electron laser

LIST OF SYMBOLS

CHAPTER 1

INTRODUCTION

Macromolecules such as proteins, nucleic acids and lipids perform the vast majority of functions within all living cells. They can form large multi-unit complexes, working in concert to control reactions, respond to stimuli, direct transport, and replicate deoxyribonucleic acid (DNA). A fundamental assumption of molecular biology is that the structure of macromolecules defines their function and is central to understanding the dynamics of all living matter. Disease and cancer research, in particular, require understanding these dynamics and controlling them for structure-based drug design and the advancement of human health.

Since molecular bonds and interactions occur at atomic scale, a high resolution measurement of molecular structure is vital to interpret macromolecular shape. This is a true challenge due to lack of electron scattering power from a single molecule, like a protein. Macromolecular x-ray crystallography (MX) solves the intensity problem by enhancing x-ray scattering with a protein crystal lattice, sending diffraction into distinct intensity trajectories (schematic shown in Figure 1.1) which satisfy Bragg's Law:

$$2d \sin\theta = n\lambda \tag{1.1}$$

where d is the interplanar lattice spacing, λ is the x-ray wavelength, and 2θ is the scattering angle of the x-rays from the incident beam axis. The recorded Bragg peak intensities can be analyzed to obtain an accurate three-dimensional (3D) model of the molecule. Since x-ray wavelength is on the order of an angstrom, MX can provide high, atomic-level resolution.

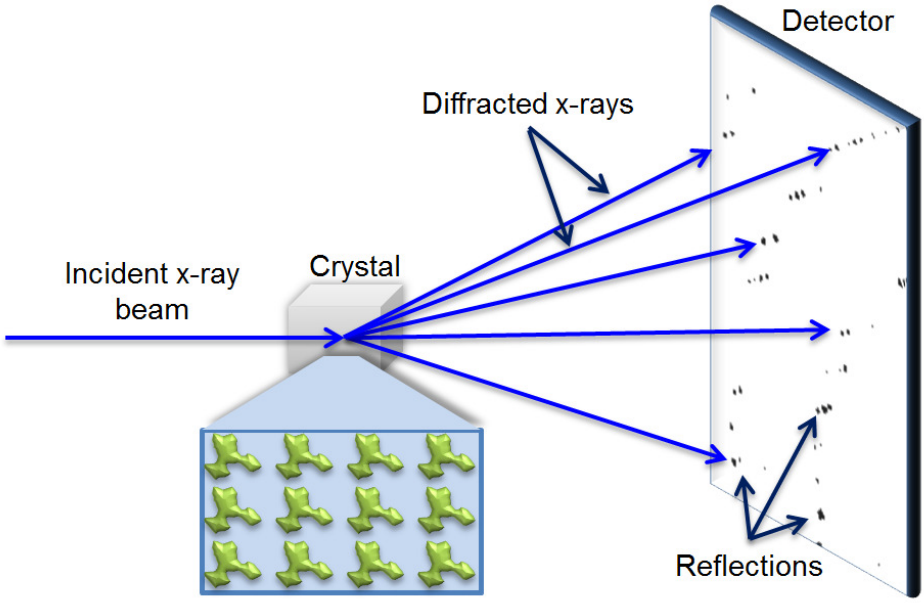


Figure 1.1: Schematic of typical crystal scattering onto a detector. The crystal lattice is formed by a regular lattice array of (green) proteins. Diffracted x-rays (blue) satisfy Bragg's law and form reflection spots which are subsequently recorded by a detector in data frames.

MX is the most widely used technique for 3D structure determination. As the catalog of known protein structures expands, it becomes apparent that many of the most biologically relevant proteins do not readily form large crystals necessary for strong diffraction. Instead they form small crystals on the order of a micrometer or smaller [1, 2, 3, 4, 5, 6, 7, 8]. This is most apparent in crystallization growth screening and in unique crystallization environments, such as when grown *in vivo* and in a special membrane-mimic, lipidic cubic phase [9, 10, 11]. As more difficult-to-crystallize proteins are studied, this is catalyzing a shift in the future of MX towards the study of small crystals - microcrystallography - and will require an adjustment to traditional crystallography approaches. The goal of this dissertation is to explore and develop microcrystallography methods.

In the ideal experiment, a complete data set is obtained by recording enough diffraction from a single protein crystal to solve the structure of the protein. Since diffraction only occurs where the Ewald sphere intersects with the reciprocal lattice point (relp), many crystal orientations relative to the beam must be recorded in order to cover enough of reciprocal space for a structure solution. Usually, this requires collecting a series of data images from a rotating protein crystal until radiation damage ultimately limits the amount of useful information diffracted from the crystal. Techniques for merging data from multiple crystals at many orientations in the beam must be used for a structure solution if complete data sets cannot be collected from a single crystal. This technique of obtaining complete data sets from multiple small protein crystals is referred to as serial microcrystallography (SMX), now at the forefront of modern microcrystallography.

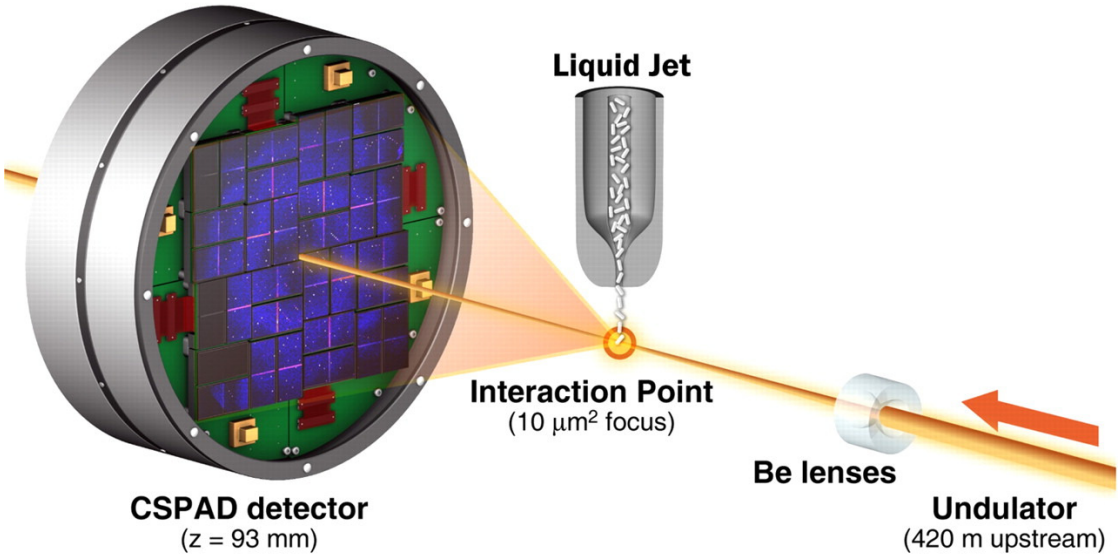


Figure 1.2: Experimental geometry for serial femtosecond crystallography at an XFEL, *e.g.*, the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory (SLAC). Single crystals flow in a liquid jet to a focused interaction point where the single 120 Hz repetition rate XFEL pulses diffract from microcrystals and are recorded on a CS-PAD detector (Cornell SLAC pixel array detector) [2].

As radiation damage limits useful protein crystal diffraction lifetimes, scientists initially looked toward next generation light sources (NGS) as a solution for SMX [2, 12]. Successful experiments at a x-ray free-electron laser (XFEL) introduced the concept of "diffract-before-destroy", where an extremely bright x-ray pulse diffracts from the sample and is recorded before the damage from the x-ray pulse tears apart the sample. A single data image is recorded from each sample before the next sample is delivered to the beam, and the process repeats until enough images with unique crystal orientations are collected to solve a structure. Shown schematically in Figure 1.2, this process is referred to as serial femtosecond crystallography (SFX), due to the tens of femtosecond duration of the XFEL pulse. While SFX encourages the microcrystallography field to grow steadily at XFELs, beamtime availability is constrained for the near future, compelling the development of necessary tools and techniques for SMX to be tested at storage ring sources.

In the meantime, the field is looking to push limitations at current bright sources, including storage-ring-based synchrotron radiation (SR) facilities. Microcrystallography has already seen some success at bright storage ring sources (which will be reviewed in Chapter 2), but it is clear that a complete redesign of the entire crystallography experiment is necessary to advance. In theory, diffraction from micron-sized crystals should satisfy requirements for a complete structure solution [13, 14], and yet the limits of what is possible in microcrystallography at storage ring sources have yet to be experimentally explored. This is because current beamlines and crystal sample chambers add x-ray scatter which obscures the weaker signal from small crystals. The goal is to optimize storage-ring-based experimental setups in favor of enhancing diffracted x-rays from the smallest crystals and minimizing any unwanted background scatter.

Microcrystal size presents unavoidable limitations, because the number of diffracted x-rays decreases proportionally with crystal size. Longer exposure times and brighter sources can increase the x-ray dose delivered to the crystal. However, the dose per protein molecule before it is radiation damaged is, at best, constant with time. Thus it is necessary to obtain merged data from many microcrystals to equal the diffraction information from a single large crystal. Traditionally, to merge data sets, it is necessary to know the orientation of each crystal for every collected image. With any given source, there exists a lower limit to the size of a microcrystal where too few x-rays are collected per image to determine the crystal orientation for merging data sets before radiation damage sets in. Such an image is considered to have "sparse data" and is traditionally discarded as useless. In collaboration with the Elser group at Cornell University, we have developed the *Expand-Maximize-Compress* (*EMC*) algorithm which, when applied to sparse crystallographic data, allows reconstruction of a complete data set in reciprocal space [15, 16, 17]. Through these proof-of-principle exercises, we conclude that requiring enough diffraction from individual crystals to determine orientation on a per-crystal basis is unnecessary to build a 3D reciprocal space Bragg pattern, given enough random distributions of crystal orientations. In principle, this eliminates the restriction on source brilliance and makes SMX an extremely attractive option at SR sources.

The goal of this dissertation is to explore current methods for serial microcrystallography at synchrotron-based storage ring sources and to suggest a reconstruction in favor of an ideally zero background microcrystal experiment.

1.1 Limitations in protein crystallography

The main function of crystallography is to distinguish Bragg-reflected x-rays, or structural "information", from background scattered x-rays, and to determine the 3D reciprocal space Bragg intensities from two dimensional (2D) data images. Given a fixed source of x-rays, the amount of structural information accessible in a data image is ultimately constrained by three limitations: radiation damage, background scatter, and crystal quality and size. These limitations must be considered when designing the experimental setup necessary for serial microcrystallography.

1.1.1 Radiation damage

X-ray damage limits the amount of recordable structural information from each biological sample. The types of damage are classified into two categories: primary and secondary. Primary damage occurs from the direct interaction of x-rays and atoms, resulting in photoelectron ejection. This process is linearly dependent on dose (absorbed energy per sample mass) and is unavoidable. Secondary damage arises from free radical formation, is diffusion limited (and is thus time and temperature dependent) and can continue even after the beam is turned off [18]. Both types of damage cause bond scissions, chemical alterations and charge movement, which eventually deteriorate the crystal lattice and destroy the constructive interference needed for defined Bragg peaks [19]. This can be measured through changes in the unit cell dimensions, increased Wilson B values, decreased Bragg spot diffraction intensity and loss of high resolution data.

Fortunately, secondary damage can be mitigated if the sample is sufficiently cryocooled, turning the intermolecular fluid water into a glass. Slowing the diffusion of ions through the sample effectively limits the damage to local sites of photoelectric ejection and the total damage rate becomes linear with total dose [20]. However, cooling samples has several drawbacks: possible crystalline ice formation, cryoprotectant effects on crystal, and thermal effects on the protein. When cooled, crystalline ice formation within the mother liquor solution in and around the crystal can affect the lattice structure and produce distinct ice diffraction rings that interfere with accurate Bragg spot measurement. To avoid these effects, crystals are treated before cooling with cryoprotectants, such as oil, polyethylene glycol (PEG), or glycerol. This cryoprotection and cryocooling step has been shown to negatively affect crystal lattice structure, either through the additional handling or cryoprotectant interaction with the protein, and increases the mosaic spread of the diffraction spots [21, 22, 23]. Cryocooling has also been shown to remodel the conformational distribution of more than 35% of protein side chains and eliminate packing defects necessary for functional motions [24]. Lastly, motions critical in catalysis, such as ligand binding and allosteric reactions, do not occur in cryocooled samples [25]. Studies which aim to examine dynamic protein structures are then limited to biologically inactive states, if samples need to be cooled. In the context of smaller crystals, it is possible that cryocooling across microcrystals may produce smaller mosaicity increases mentioned above, but this remains to be studied in depth.

In the absence of glassy ice, radiation damage is time-dependent. Studies suggested that the effects of secondary radiation damage may be outrun with higher dose rates [26]. Given sufficiently high dose rates at storage ring sources and fast detectors, it may be possible to dose microcrystals quickly enough at

room temperature to increase the information gathered before damage is limiting. With a single exposure per room temperature microcrystal, this still results in hundreds or thousands of crystal samples needed [13]. The benefit of bright x-ray beams from NGS, compared to current storage ring sources, is that fewer non-cryocooled crystals may be needed in order to collect enough diffraction for a complete data set. However, given enough samples, radiation damage doesn't limit microcrystallography to NGS.

1.1.2 Background scatter

Every diffraction image contains two sources of signals: diffraction from the crystal and scatter from everything else in the path of the beam to the detector. The latter, referred to as background scatter, is a source of "noise" and complicates attempts to solve crystal structure from Bragg peaks. One measure of background scatter is the signal-to-noise ratio (SNR) , which is the ratio of average intensity, $\langle I \rangle$, of the Bragg peaks within the data set to their average standard error, $\langle \sigma(I) \rangle$, expressed as $\langle I \rangle / \langle \sigma(I) \rangle$. The SNR ideally should be higher than 2 at any peak in order to distinguish a Bragg peak photon hit from background scattered photon hits [13], thereby setting a lower limit to the background scatter that can be tolerated. Sources of background scatter include all matter in the x-ray beam including air in the beam path, window and mounting materials, cryoprotectants, and water within the crystal. In order to distinguish the signal from a very small crystal, it is necessary to reduce the number of background scattered photons in data images. This requires a complete overhaul of anything the x-ray beam encounters other than the crystal and the detector and will be discussed in detail in Chapter 3.

1.1.3 Crystal size and quality

Numerous attempts to determine protein structure have failed due to difficulties in obtaining an adequately large crystal for good quality diffraction. Until recently, crystals needed to be larger than tens of microns across to have a chance at producing obvious Bragg peak spots. This is in part due to larger contributions in noise superimposed on diffraction patterns, but also due to less-bright sources in early beamline setups. Analyzing Bragg peak intensities is an essential step in determining crystal structure. From Darwin's law, we know the intensity I of any Bragg diffraction spot (photons/spot) from a crystal is directly proportional to size of the crystal [13, 27]:

$$I \propto I_0 \frac{V_{xtal}}{V_{cell}^2} |\mathbf{F}(\mathbf{hkl})|^2 \tag{1.2}$$

where I_0 is the intensity of the incident beam (photons/s/m^2), V_{xtal} is the volume of the illuminated crystal, V_{cell} is the volume of the unit cell, and $\mathbf{F}(\mathbf{hkl})$ is the structure factor of the unit cell at the relp (hkl). The structure factor describes the amplitude of the diffracted x-rays as a function of the Thomson scattering cross section of the electrons within the unit cell [28].

The number of Bragg diffracted photons needed to solve a structure is dependent on the total number of electrons in the molecule being imaged. Eventually, with decreasing crystal size, the number of x-rays contributing to a Bragg peak in a single diffraction image becomes indistinguishable from the background. Such x-ray diffraction images also contain information since x-rays from Bragg diffraction are still being recorded, but the images are considered sparse, and are usually discarded as worthless (this will be addressed later in

Section 1.2.5).

Diffraction quality depends not only on the total diffraction intensity, but also crystal quality and mosaicity. In fact, if the sample is a well-ordered crystal hundreds of micrometers in diameter, then a rotating anode will be sufficient to solve the structure given enough time (the experiments described in Chapters 5 and 6 use this kind of x-ray source). The quality and the mosaicity of the crystal can be measured from the resolution strength and angular spread of the Bragg diffraction spots. Many factors can influence the measurement. In addition to physical over-handling while delivering the crystal to the x-ray beam, crystal mosaicity can be inherently compromised by a combination of disordered crystal-surface proteins and complex interactions within the crystal itself. Almost all large protein crystals are formed from small mosaic blocks of more perfectly-ordered domains [4, 29] and it is presumed that decreasing the size of the crystal to that of the more perfect mosaic block will improve the diffraction quality of the crystal. This was seen in studies on the G protein-coupled receptor (GPCR) complex reported by Weierstall *et al.* [8] at an XFEL and a storage ring source in 2014. While imaging crystals at the storage ring source, the larger GPCR crystals diffracted poorly, hindering structure solution. Ergo, they brought smaller receptor complex crystals to the LCLS, and crystals diffracted well enough to solve their structure. Other studies have also suggested that smaller crystals may yield higher-quality diffraction [3, 7]; however, conclusive investigations into the quality of crystal diffraction as a function of crystal size have not been reported as of this writing.

As a note on small crystal size, techniques used for handling large crystals don't function efficiently for handling thousands of smaller crystal sizes needed

in SMX. But this is simply an engineering problem and not, in principle, a limiting factor for the field. Novel techniques for microcrystal handling are progressing in the community and will be discussed in Chapter 2.

1.2 Features of a microcrystallography experiment

As the predominant technique to experimentally determine atomic 3D structures, MX has contributed over 90% of the protein structures deposited in the Protein Data Bank (PDB). The essence of a MX experiment involves everything from the x-ray source to the structure solution, including beamline optics, crystal growth, crystal transfer to the beam, detectors and post-imaging analysis. Optimizing each of these elements for microcrystallography is key to expanding the applicability of SMX. The next section will walk us through the characteristic elements, beginning with what is considered to be a microcrystal, and highlight elements significant to this dissertation.

1.2.1 In numbers: microcrystals

Although there is no formal agreement within the crystallography community for what is technically considered a microcrystal, it is agreed that crystals smaller than a few tens of microns across begin to be a challenge for traditional crystallographic methods. For the purpose of this dissertation, microcrystallography will be limited to the study of protein crystals smaller than a few microns across (therefore this also includes nanocrystals).

First, it is useful to estimate the number of microcrystals needed for a struc-

ture solution in order to provide a context for a microcrystallography experiment. Consider a microcrystal with a volume of $(2\ \mu m)^3$ and a unit cell edge of 100 Å. What is required for a data set with a 2-3 Å resolution structure solution? From first principles, Holton and Frankel [13] predict that if every diffracted photon is recorded and there is zero contributing background, a complete 2 Å resolution data set could be obtained from a single cryocooled lysozyme crystal sphere 1.2 μm in diameter. How does this scale to the $(2\ \mu m)^3$ microcrystal? Taking into account increased photoelectric escape in microcrystals [30, 31], Gruner and Lattman [14] predict in a "best case" scenario that a single microcrystal 2 μm across with a 100 Å unit cell should yield a 2 Å resolution data set [14]. They compare to a "worst case" scenario of Sliz *et al.* [32]: extrapolating from experimental results using the technology available in 2003, a 3.5 Å resolution data set could be obtained from a cryocooled $(20\ \mu m)^3$ crystal. Recall that at cryocooled temperatures, radiation damage is linear with dose (see Section 1.1.1 and [20]), and that diffracted information intensity is defined by the total volume sampled. Therefore in this "worst case" scenario, the $(20\ \mu m)^3$ crystal volume needed for a 2 Å resolution structure solution can be scaled to 10^3 cryocooled crystals $(2\ \mu m)^3$ in size. In short, only a few thousand microcrystals are needed for an adequate structure solution. Recent studies have already shown this number to be accurate [11, 33]. These will be detailed further in Chapter 2.

Let us not limit ourselves to cryocooled conditions. Assume that uncooled crystals suffer approximately 100 times the damage per dose compared to cryocooled crystals [30]. Even at room temperature, the effects of radiation damage only increase our numbers by a factor of ~100. For a 2 Å resolution data set using a $(2\ \mu m)^3$ crystal, the "best case" scenario requires 10^3 crystals and the "worst case" requires 10^5 crystals. At worst, around a hundred thousand microcrystals

will be needed obtain a complete data set for a $(2\ \mu m)^3$ microcrystal. Between these extremes of "best" and "worst" is reality, where beamline improvements would include advanced detectors and background reduction techniques not available to Sliz *et al.* [32] in 2003. And while 10^5 crystals seems cumbersome, the total volume of 10^5 $(2\ \mu m)^3$ crystals is nearly comparable to that of a single $(100\ \mu m)^3$ crystal.

The challenge is in practicing microcrystallography, *i.e.*, handling such large numbers of tiny crystals with short exposure lifetimes (milliseconds for microcrystals), and processing the resulting large data sets.

1.2.2 X-ray light sources

Returning to the characteristic microcrystallography experiment, SMX requires a source of low-divergence x-rays with a small focus to match the size of the crystal cross-sectional area. Matching beam and crystal size reduces background scatter from the material surrounding the crystal and low divergence limits the contribution of the beam to mosaic spread. To compare strength and quality of x-ray sources, the characterization measure of x-rays beams is spectral brightness:

$$brightness = \frac{photons}{second \cdot mrad^2 \cdot mm^2 \cdot 0.1\%BW} \tag{1.3}$$

where divergence of the beam is measured in $mrad$, size of the beam in mm^2, and 0.1% BW denotes a bandwidth (BW) centered around the central x-ray frequency ω. Integrating over beam area, divergence and bandwidth yields the flux (number of photons per second per unit area), which is a simplified measure of comparison for microbeamlines. Attributes worth noting include x-ray

energy or wavelength, specific bandwidth of the beam, angular distribution, size of the source, time variations, and polarization. Each of these attributes need consideration and can be individually optimized, depending on the type of experiment being performed; however, it is the simple measure of flux on the crystal which is of importance in this dissertation.

Storage ring light sources

The evolution of MX parallels much of x-rays light source development, as increasingly brighter sources have enabled dramatic leaps in MX capabilities. Historically, x-ray diffraction was first developed on conventional x-ray machines. The basics of an x-ray machine are an electrode pair, a cathode and an anode inside a vacuum tube, accidentally discovered in 1895 by Wilhelm C. Röntgen [34]. When a high voltage is applied, electrons accelerate from the cathode and strike the anode, emitting sharp peaks of x-ray energies characteristic of the electronic structure of the anode material superimposed on broadband Bremsstrahlung. It wasn't until the early 1950's that significant improvements to this system were made with a rotating anode, which increased the brightness by an order of magnitude, as seen in Figure 1.3. This achievement was quickly surpassed by the first generation of SR storage ring sources in the mid-to-late 1960's. SR is produced by accelerating ultrarelativistic charged particles transverse to their primary direction of motion, such as along a curved path. Originally intended as high energy physics particle colliders, the first generation of synchrotrons created strong magnetic fields using simple dipole bending magnets to deflect the particle beam around the curvature of the ring. This deflection produced a parasitic fan of incoherent and broad spectrum SR. The "unwanted" radiation

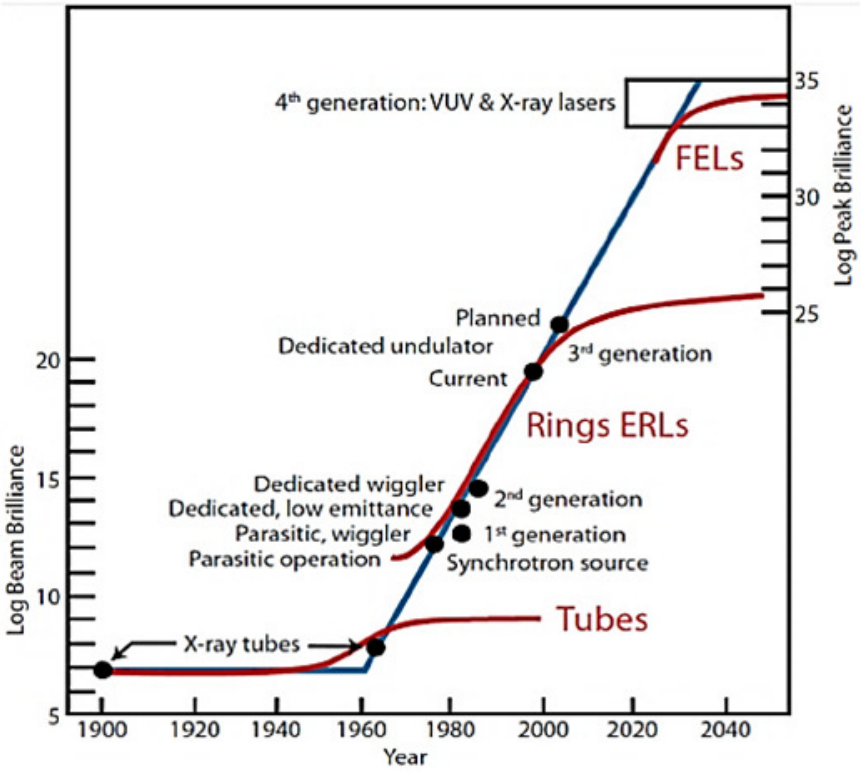


Figure 1.3: Historical log graph shows enormous increase in brightness of x-ray sources in the past half century. Tubes represent x-ray tubes including rotating anode x-ray tubes introduced in the 1960's and used in Chapters 5 and 6. Rings represent the synchrotron facilities of storage rings. ERLs represent the predicted brightness of Energy Recovery LINACS, and "FELs" represent current and future peak brightness capabilities of x-ray free-electron lasers. Courtesy of M. Rivers [35].

proved to be a breakthrough in x-ray science as it provided several orders of magnitude increase in spectral brightness over the x-ray tube and introduced synchrotrons as a very promising source for intense x-rays. In the 1960's and 1970's, insertion devices were designed and installed specifically to enhance the intrinsic brightness of the SR source. Composed of periodic arrays of dipole magnets, termed "wigglers" and "undulators", these insertion devices oscillate the charged particles thereby producing radiation cones at each bend in the oscillation. Wigglers produce radiation cones over a continuous spectrum similar to that of bending magnets, but give higher brightness as seen in Figure 1.4.

Owing to a relatively weak magnetic field, an undulator emits radiation cones at each bend in the beam trajectory which overlap and give rise to a constructive interference effect that results in one or a few spectrally narrow peaks in a beam that is highly collimated in both the horizontal and vertical directions [36].

At the same time as these developments, the second generation of syn-

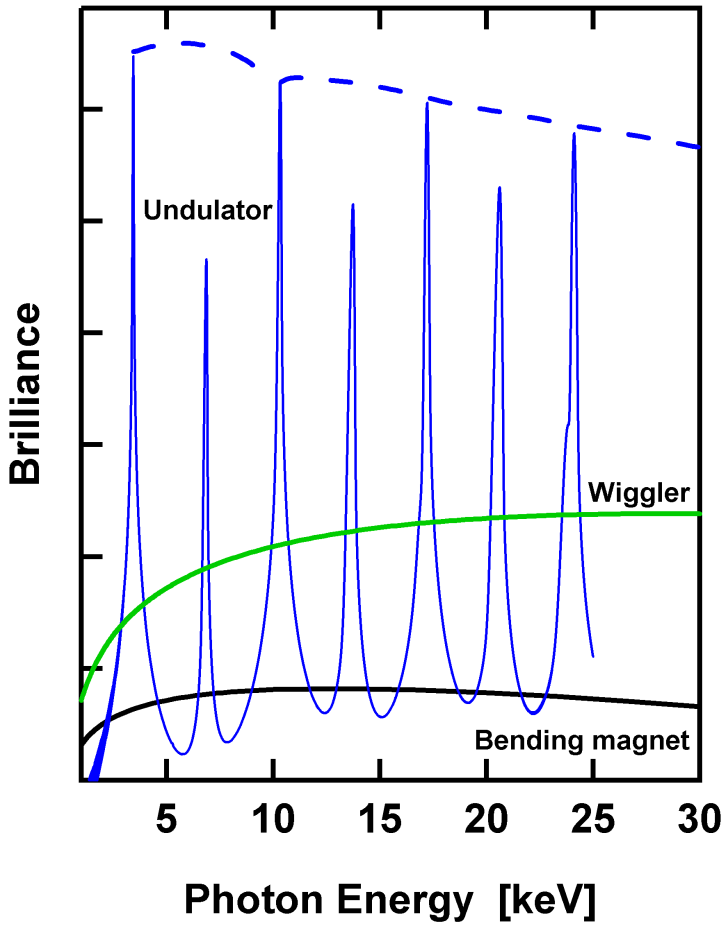


Figure 1.4: Graphical representation of the spectra from various sources of synchrotron radiation. Bending magnets produce a smooth spectrum of incoherent radiation. In comparison, wigglers also produce an incoherent x-ray beam resulting in a broad spectrum, but is brighter than a bending magnet source. The spectrum from undulators shows the effect of constructive interference, producing discrete "peaked" energies which are harmonically-related. Figure courtesy of DESY [37].

chrotrons dedicated solely for x-ray production started coming online and showed, again, orders of magnitude increase in spectral brightness from the first generation. Over the next few decades, third generation SR sources saw increasingly transversely coherent x-ray beams with the installation of permanent insertion devices. This also resulted in lower emittance and further increases in brightness. Since the late 1990's, modern third generation synchrotron sources feature undulators and optimized beam optics to improve the efficiency of light generation.

As monochromatic x-rays are typically needed in MX experiments, monochromators are installed after the insertion devices to reduce the spectrum to more narrow bandwidth x-rays. Downstream optics such as mirrors, compound refractive lenses (seen in Figure 1.2), Fresnel zone plates, polycapillaries, multi-bounce and single-bounce monocapillaries, and other x-ray refractive optics are installed to focus and collimate the beam to the desired sample beam size [38]. An example of a typical focusing optic is a set of Kirkpatrick-Baez (KB) mirrors (seen in Figures 2.3 and 2.4), which focus the beam both horizontally and vertically to the focal spot where the sample is located.

Next generation x-ray light sources

Next generation sources, including energy recovery linacs (ERLs), ultra-low-emittance "diffraction-limited" storage rings (UESRs) and XFELs, are starting to come online and feature transversely coherent production of x-ray radiation. Current examples of the XFEL type of facility are SLAC's LCLS, Deutsches Elektronen-Synchrotron's (DESY) Free-electron LASer in Hamburg (FLASH), and RIKEN's Spring-8 Angstrom Compact free electron LAser (SACLA). XFELs

can produce highly brilliant femtosecond pulses of coherent x-rays on the order of 10^{33} photons/s/mrad^2/mm^2/0.1% BW. The typical pulse duration can be on the order of 10-100 femtoseconds, and have a peak of about 10^{12} photons per pulse. In order to produce such radiation, electrons are accelerated by a UV laser from a photo-cathode into a linear accelerator with radio-frequency cavities. The electrons reach relativistic velocities before passing into a long array of undulators where the electromagnetic field of all electrons are modulated until the fields superimpose in phase. This creates an increase in the amplitude of the total electromagnetic field, and produces coherent and bright x-rays. Unfortunately, the so called "SASE" (self amplified spontaneous emission [39]) conditions to produce such a beam also make the XFEL beam unstable, adding timing jitter and energy uncertainty which produce images that differ in spectrum and intensity from shot to shot. Still, experimental time at SASE XFELs is highly competitive and the x-rays they produce are a unique source for experimentalists. Future seeded XFELs have promise for more stable beams [40].

In practice, the ultrashort duration pulses from XFELs enable data collection from microcrystals before radiation damage occurs. These pulses also are highly destructive and typically vaporize samples within a single pulse. Many microcrystals are required to obtain a complete data set. However, the routine availability of XFEL sources is still on the horizon, as at present only three users in the world are on an x-ray free electron laser station at any time.

Storage rings are still the most widely used source for protein structure determination and steady improvements have expanded the feasibility of advanced diffraction experiments. The tuning flexibility, availability of beamtime and beam stability will continue to make storage rings a very relevant and active

source for crystallographers for many years to come. As such, several storage ring beamlines designed for microcrystallography will be discussed in Section 2.1.

1.2.3 Microcrystal growth and delivery

A typical crystallography experiment starts with crystallizing a protein of interest (see Figure 1.5 for the traditional MX pipeline). As thousands of microcrystals are required for merging in SMX (determined in Section 1.2.1), methods are needed for producing large numbers of homogeneously sized, high quality microcrystals; this will be discussed in detail in Section 2.2 and Chapter 4.

Figure 1.5: Simplified pipeline of traditional crystallography. First, protein crystals are grown from purified proteins in solution. They are delivered to the x-ray beam where they are imaged. The images are indexed to establish orientation, and datasets are integrated, scaled and merged together. The resulting intensity distribution is phased and a structure is built. Lastly, refinement cycles are performed to optimize the structure to the diffraction data. Each step is critical to be performed before moving onto the next in the traditional pipeline.

The next step is collecting diffraction data from thousands of microcrystals at a beamline. Since each microcrystal may only survive a single x-ray exposure, acquiring data from thousands of microcrystals necessitates delivering them unharmed in quick succession to the x-ray beam to fit into a feasible experimental time window. Additionally, enough orientations must be captured to explore sufficient reciprocal space for a structure determination (which is highly dependent on crystal symmetry). This means the delivery of microcrystals must allow enough flexibility to explore those orientations. Merging partial data sets from a

random distribution of orientations should provide a multiplicity of reflections which follows a binomial distribution [41]. Any departure from this shows preferential orientation of the microcrystals (which occurs in some fixed-target devices) and will limit the data collected for certain reflections. Various delivery devices and methods will be discussed in Chapter 2 and Chapter 4. Knowing the background scatter contributions from each of the devices is key to limiting additional scatter. Crystal visibility, orientation and timing of delivery into the beam is critical to designing experiments.

1.2.4 Detectors

Recording the diffraction from protein crystals is next in the crystallography pipeline. Electronic detectors such as charge-coupled devices (CCD) and pixel array detectors (PADs) currently dominate the crystallography field. Detectors can be classified into two groups, photon-counting (most PADs) and photon-integrating detectors (special PADs and CCDs) which differ in the technique of recording photons. Several characteristics of detectors need to be considered for a SMX experiment: pixel size, count rate, energy range, dynamic range, quantum detective efficiency, and energy resolution.

Fast-framing, highly efficient x-ray detectors are required to collect the quick succession of microcrystal exposures from SMX. As radiation damage limits the lifetimes of small crystals, especially at room temperature, fast-framing detectors are also needed to record images before radiation damage effects propagate, even in stationary samples. Considering the extremely low number of photons diffracted from microcrystals, every Bragg diffracted x-ray needs to be

recorded with high SNR. Measured Bragg x-ray signals can be compromised by variations in absorption of the detector window, partial transparency of the detector sensor to x-rays, and calibration imperfections, but these limitations can be made small [14]. SMX also requires high efficiency detectors to distinguish accurate photon hits.

Today's fast-framing detectors, such as the mixed-mode PAD (MM-PAD) [42] developed in the Gruner Group at Cornell University, and hybrid photon counting detectors, such as the EIGER developed by the Swiss Light Source (SLS) and commercialized by Dectris, have demonstrated that detectors do not need to be the limiting factor for SMX.

1.2.5 Data analysis and structure solution

Once thousands of crystals are delivered to the beam in random orientations, consider a complete diffraction data set with thousands of individual crystal images. If the images can be indexed to determine crystal orientation relative to the beam, then the images can be merged and analyzed to solve a complete 3D structure (see the pipeline in Figure 1.5). However, if the images are severely Poisson-noise limited, *i.e.*, the images contain so few x-rays that no single image contains enough information to determine the orientation of the crystal, then conventional techniques to merge images fail. This presents a limit of "indexability" for the data processing, and such images unable to be indexed are considered "sparse". Recently, this limitation has been relaxed with the introduction of the *EMC* algorithm [43, 44]. Although the *EMC* algorithm was originally developed for single particle imaging experiments at XFELs, Gruner and Elser

suggested that *EMC* can be used to reconstruct reciprocal space from the unindexable frames of randomly-oriented microcrystals, if enough measurements or frames are available. The protein crystal structure can then be solved from the 3D reconstructed reciprocal space using more traditional techniques. This dissertation will introduce the application of the *EMC* algorithm for serial microcrystallography in Chapters 5 and 6.

1.3 Dissertation Organization

Chapter 2 of this dissertation will cover a more detailed review of the current state and techniques of microcrystallography. Chapter 3 will address the SMX limitation of background scatter more carefully, with special attention to microbeam production, thin window materials and mounting materials for SMX. Chapter 4 will explore two specific examples of delivery devices intended for SMX, a microfluidic chip and a viscous jet injector, and lessons from their use. To address the improvement of software to handle sparse data from microcrystals, Chapter 5 will introduce the *EMC* algorithm and three initial proof of principle experiments. These experiments include *EMC* applied to 2D and 3D shadow radiography, which were performed by others but serve as an introduction to Section 5.5, describing a *EMC* reconstruction of diffraction from a large inorganic crystal using a rotating anode. To simulate SMX using *EMC*, Chapter 6 will detail a crystallography experiment where the 3D reciprocal space diffraction intensity from a large protein crystal was reconstructed using *EMC* algorithm. Chapter 7 will conclude with discussion on requirements for a practical future in SMX with special attention to beamline conditions, sample delivery devices, and limitations which may arise in data handling with the *EMC* algorithm.

CHAPTER 2

CURRENT STATE OF MICROCRYSTALLOGRAPHY AT STORAGE RING SOURCES

Until recently, crystallographers considered microcrystals as a nuisance, indicating a need for continued screening in order to produce sufficiently large crystals for diffraction. Now with fruitful studies of SMX at XFELs, there is interest in developing tools to expand microcrystallography capabilities at all SR sources. Many of the following tools were developed for use at XFELs, but their use is not limited to those facilities. The flexibility built into many of the following experimental tools, whether intended or not for XFEL use, will aid SMX at other SR sources. In the context of SMX at storage ring sources, it is helpful to detail the current status of instrumentation and methods and acknowledge present experimental limitations which will require additional optimization for practical SMX.

Successful demonstrations of serial microcrystallography have already been performed at storage ring sources. Table 2.1 shows several recent studies, both at cryocooled and room temperatures, completed at storage ring sources with crystals a few tens of micrometers across or smaller. While several studies have also shown the feasibility of using micro or nano-focused beams on larger crystals of 10-100 μm to demonstrate various microcrystallography tools [45, 46, 47, 48], this dissertation will focus on those specifically designed for microcrystals smaller than 10 microns in diameter.

The number of studies done on microcrystals smaller than ~10 μm is increasing, but it is clear that there is still much difficulty in obtaining structures from crystals smaller than a few microns. In the next few sections, we will

Study	Crystal size	Beam size (FWHM)	Beamline	Flux* (ph/s)
Cryocooled				
Boudes *et al.*	(5 -15 μm)^3	10 × 10 μm^2	MX2 at AS	3.6×10^{11}
Gati *et al.*	(0.9 × 0.9 × 10) μm^3	4 × 5 μm^2	P14 at PETRA III	1.2×10^{12}
Roedig *et al.*	(4 μm)^3	50 × 50 μm^2	I24 at DLS	5.7×10^{10}
Room temp.				
Botha *et al.*	(10 × 10 × 30) μm^3	10 × 30 μm^2	PXII at SLS	2×10^{12}
Lyubimov *et al.*	(10-15 μm)^3	10 × 10 μm^2	12-2 at SSRL	2×10^{12}
Murray *et al.*	(10 μm)^3	10 × 10 μm^2	23-ID-D at APS	
Nogly *et al.*	(5 × 5 × 1-5) μm^3	2 × 3 μm^2	ID13 at ESRF	9.1×10^{12}
Stellato *et al.*	(3 × 3 × 6) μm^3	6 × 9 μm^2	P11 at PETRA III	2×10^{11}
Tsukui *et al.*	(10 × 10 × 50) μm^3	50 × 100 μm^2	BL38B1 at SPring-8	

Table 2.1: Microcrystallography studies performed at storage ring x-ray sources, showing crystal size, resolution of structure solved, at what beamline the study was performed and the approximate flux (in photons per second) used in the study, if provided by the study. Studies which exceed tens of microns are outside of the context of this dissertation, and are not included. Both cryocooled and room temperature studies are shown. Of special note are the studies performed on crystals only a few microns across. Data from: Boudes *et al.*[49], Gati *et al.*[11], Roedig *et al.*[50], Botha *et al.*[51], Lyubimov *et al.*[52], Murray *et al.*[47], Nogly *et al.*[53], Stellato *et al.*[33], and Tsukui *et al.*[54].

walk through the experimental setup from Section 1.2, pointing out the current progress of SMX. Keep in mind that SMX will require a great number of homogeneously sized microcrystals delivered rapidly to microbeams with little excess surrounding material to contribute to background scattering. SMX will also require methods to analyze the large data sets collected, including images with little or no diffraction from microcrystals.

2.1 Microcrystallography beamlines at storage ring sources

One of the first concerns in microcrystallography is finding a sufficiently bright x-ray source which most closely matches the cross-section of the microcrystal

of interest. Section 1.2.2 briefly covered many different light sources, but of interest to this dissertation are storage ring sources, and their microbeam capabilities. Ideally, if the footprint of the beam exactly illuminates the sample, then the amount of background scatter from anything surrounding the sample is reduced (more detail in Chapter 3). Table 2.2 is an incomplete list of currently available storage ring source macromolecular micro-beamlines, as of August 2016.

Upgrades to current second and third generation facilities such as CHESS, APS, Diamond, SPring-8 and ESRF will continue to improve beam size and flux availability. In the very near future, other advanced third generation SR sources such as MAX IV and NSLS II are coming online. These are based on complex magnet lattice designs to provide microbeams of ~1 μm or smaller with flux of $\sim 10^{13}$ photons/second (ph/s) [55]. Microfocusing x-rays is critical to strengthen signal from microcrystals without adding detrimental background scatter from extraneous material around the crystal. Thus, matching the beam size to that of the crystal is essential. Beamlines with tunability in this fashion are especially attractive for microcrystallographers.

There are several methods to create microbeams, including using collimators to reduce overall size of the beam, and using optics to microfocus the beam to a focal point. Collimators can achieve specific beam footprint sizes without adding divergence to the beam, while microfocusing optics add some divergence (see Table 2.3 for details on focal spot size, divergence and gain over collimated beams).

Microbeamlines need to have low-divergence to avoid increasing the angular resolution of the Bragg diffraction peak spots past the point where peaks are

Table 2.2: Estimates of current storage ring x-ray source protein microcrystallography beamlines. Microbeam size, energy range and flux available are listed. Microbeam sizes are achieved through a variety of methods, including capillary focusing and collimating, and only include micro beams smaller than 20 microns. Beam size and flux are either reported from beamline specifications provided by the facility or the most recent publications from that facility. The exceptions are measurements at CHESS, performed by J. Wierman. Beam characteristics are current as of August 2016. ALS: Advanced Light Source, Berkeley, CA, USA (courtesy of James Holton). APS: Advanced Photon Source, Argonne, IL, USA [56]. AS: Australian Synchrotron, Melbourne, Australia [49]. CLS: Canadian Light Source, Saskatoon, Saskatchewan, Canada [57]. CHESS: Cornell High Energy Synchrotron Source, Ithaca, NY, USA. DLS: Diamond Light Source, South Oxfordshire, UK [58]. ESRF: European Synchrotron Radiation Facility, Grenoble, France [53, 59]. PETRA III: Positron-Electron Tandem Ring Accelerator III, Deutsches Elektronen-Synchrotron, Hamburg, Germany [60]. PF: Photon Factory, The High Energy Accelerator Research Organization, Tsukuba, Japan [61]. SPring-8: Super Photon Ring, Hyōgo Prefecture, Japan [62, 63]. SLS: Swiss Light Source, Villigen, Switzerland [64]. SSRL: Stanford Synchrotron Radiation Light Source, Menlo Park, CA, USA [52]. Table on next page.

Beamline	Energy range	Beam size	Flux (ph/s)
ALS			
8.3.1	5-17 keV	15 μm	7×10^{9}
APS			
21ID-D	6.5-20 keV	10 μm	5×10^{12}
23ID-B	3.5-20 keV	5 μm	1.4×10^{11}
23ID-D	5-20 keV	5 μm	1.9×10^{11}
AS			
MX2	5.5-28 keV	10 μm	3.6×10^{11}
CLS			
08ID-1	6-18 keV	5 μm	2×10^{10}
CHESS			
A2	5-70 keV	< 20 μm	3×10^{10}
G3	8-13.6 keV	< 20 μm	1.3×10^{12}
DLS			
I24	6.2-17.7 keV	5 μm	1×10^{11}
ESRF			
ID13	7-30 keV	0.25 μm	3×10^{8}
		2×3 μm	9.1×10^{12}
ID23-1	14.2 keV	10 μm	9.5×10^{10}
ID23-2	14.2 keV	10 μm	1.1×10^{12}
ID29	6-20 keV	10 μm	8.7×10^{11}
PETRA III			
P11	5.5-30 keV	1 μm	2×10^{11}
P14	10 keV	5 μm	5×10^{12}
PF			
BL-1A	2.7-3.0 keV	13 μm	9×10^{10}
SPring-8			
BL32XU	8.5-20 keV	1 μm	6×10^{10}
BL41XU	6.5-35 keV	2 μm	1.7×10^{12}
SLS			
PXI	5.7-17.5 keV	1 μm	2×10^{11}
SSRL			
12-2	6.7-17.0 keV	10 μm	2×10^{12}

individually discernible.

All components which make up the beamline can be optimized to eliminate extraneous scattering, and more will be discussed on beamline optimization in Chapter 3. There is yet to be a microfocus beamline at a storage ring source which is fully optimized for reduced background scatter that can provide a small, low-divergence beam per crystal. If the goal is to image crystals less than one micron across, much has yet to be done to engineer an experimental setup which makes the only source of x-ray scatter the crystal itself.

Optic	**Focal spot size**	**Divergence**	**Gain**
Single-Bounce Capillaries	0.25-100 μm	1-10 mrad	2 to 10^3
Multi-Bounce Capillaries	0.09-50 μm	1-10 mrad	2 to 10^3
Polycapillaries	10-100 μm	<350 mrad	100 to 10^4
Fresnel Zone Plates	0.03-30 μm	0.1-350 mrad	2 to 10^6
Refractive lenses	0.25-30 μm	0.1 to 10 mrad	2 to 4000
KB mirrors	>0.09 μm	0.03 to 0.1 mrad	3 to 10^6
Waveguides	>0.03 μm	1 to 10 mrad	2 to 4000

Table 2.3: Various microfocusing optics, including typical focal spot ranges, divergence, and flux gain in photons/s/μm^2 over collimated beams. See Cornaby *et al.* [38] for more details.

2.2 Microcrystal growth

Controlling the growth and quality of protein microcrystals is essential to ensuring quality data collection, as discussed in Section 1.2.3. Given a pure protein sample, the principle is to increase supersaturation of the protein in solution until crystals nucleate and then allow crystal growth until the soluble protein is exhausted. To grow microcrystals ideally, the environmental conditions can be optimized during crystal growth to induce a high number of nucleation sites

and produce crystals in large numbers and of identical size, as this will ease merging data of multiple crystals into single data sets. Predicting what conditions will yield adequate crystals requires knowledge of the phase space of the specific protein and, typically, successful crystallographers screen thousands of precipitation conditions until quality crystals appear. Until recently, methods for producing microcrystals have remained mostly unexplored, as small crystals have generally been thought to be an unfortunate result in the search for large crystals.

The most widespread method for traditional crystallography is vapor diffusion (VD). A protein solution droplet is sealed into a chamber with a higher precipitant concentration reservoir and left to equilibrate until the concentrations match, shown in Figure 2.1(a and b). At some point, supersaturation is achieved within the protein droplet, nucleation occurs and is followed by the gradual growth of crystals. This method can take minutes to weeks, depending on the protein, and works very well for growing large, well-ordered crystals. Although it can produce microcrystals, VD yields small volumes of crystal showers which require tedious harvesting, and homogeneous growth is uncontrolled. This makes VD unsuitable for serial microcrystallography, though it has been shown to be successful in a brute force approach [65].

Batch crystallization combines a protein and a precipitant solution into a single, supersaturated slurry directly in the nucleation phase of the protein, shown in Figure 2.1(c). Though this method requires substantially more effort in pre-screening in order to accurately measure the nucleation phase [66], batch crystallization produces large numbers of homogeneously-sized crystals (very handy for serial microcrystallography) and is the prevailing method for produc-

ing microcrystals [2, 6, 7, 10, 67, 68, 69, 70, 71, 72].

A technique called lipidic cubic phase (LCP) crystallization has been very successful in crystallizing membrane proteins [73]. Within either crystallization methods above, LCP forms a 3D matrix of lipids, creating a scaffold for membrane proteins to assemble. This method typically forms very small crystals with lower solvent content and better ordering for membrane proteins than traditional methods [10]. Growth within other viscous media, such as silica hy-

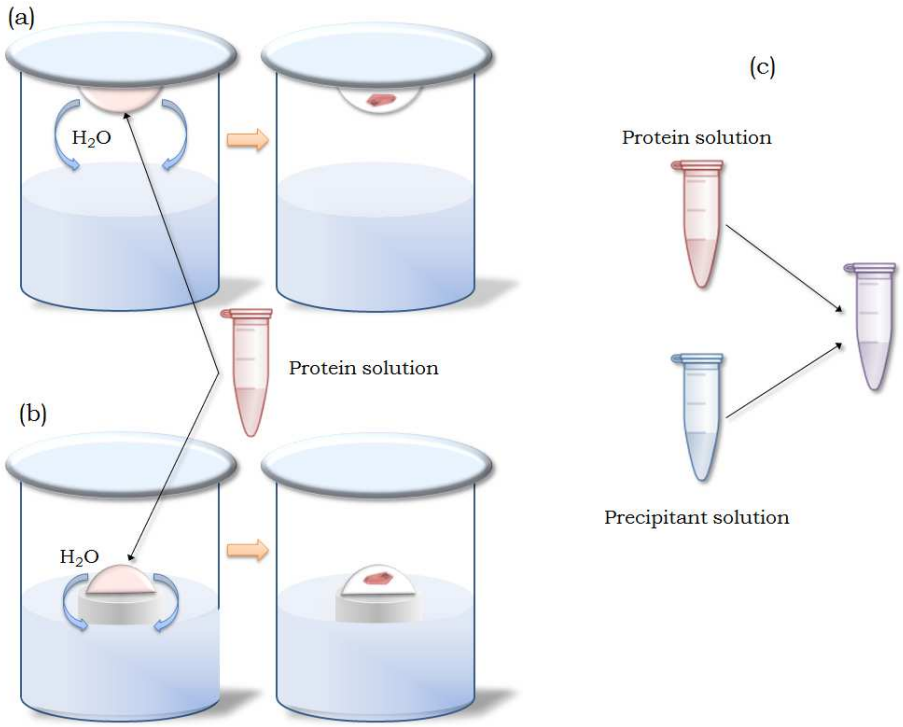


Figure 2.1: Schematic representation of vapor diffusion and batch crystallization methods (not to scale). The vapor diffusion method separates the crystallization droplet from the reservoir solution. Within the crystallization droplet in (a) hanging drop or (b) sitting drop configuration is purified protein solution and buffer solution. The reservoir below consists of buffer with precipitant and other chemical additives which can promote crystal growth. Both types of environments are sealed, and equilibrate until the concentrations of the droplet and reservoir match and, given the right conditions, produce a crystal (red). (c) Shows the batch crystallization method where purified protein solution is mixed directly with the precipitant solution.

drogels and agarose, has been shown to produce higher quality crystals when compared to VD [74, 75]. Microcrystals can also form inside the membranes of bicelles of a well-defined lipid−amphiphile mixture [76]. The bicelles create a more native environment for membrane proteins than micelles, which likely stabilizes the conditions needed to form crystals. It is also easier to handle than LCP mixtures, as it acts more like a fluid at room temperature, and crystals are easier to manipulate. It is speculative whether microcrystals can be grown quickly in large numbers within gel-like medias, but nonetheless studies have used LCP crystallization in serial microcrystallography at XFELs [8, 53, 70, 77].

An attractive approach to quality-control microcrystal growth involves a careful and dynamic tailoring of the protein crystal nucleation and growth environment. Multi-well microfluidic chips can excel at micro-managing the solution content of individual crystal growth chambers, and tuning crystallization conditions per well to produce single crystals of high quality in controlled sizes [78, 79, 80]. Some of these devices can also be used as a delivery mechanism to the beam, seen in Section 2.3. The Fraden Group at Brandeis University has developed a microfluidic chip which uses nanoliter-sized droplets and negative feedback to control concentration and to generate single microcrystals [79]. This will be discussed in detail in Chapter 4. In a slightly different approach with microfluidics, The Ros Group at Arizona State University has developed a nanowell microfluidic device which can simultaneously screen 170 protein and precipitant conditions in nanoliter quantities through gradient generation and elastomeric valving (shown in Figure 2.2 [80]). To enable diffraction *in situ* within a low-background microfluidic chip, the Perry Group at University of Massachusetts-Amherst developed a graphene-based thin-film chips using micro batch and counter-diffusion growth methods to grow crystals [78]. Growing

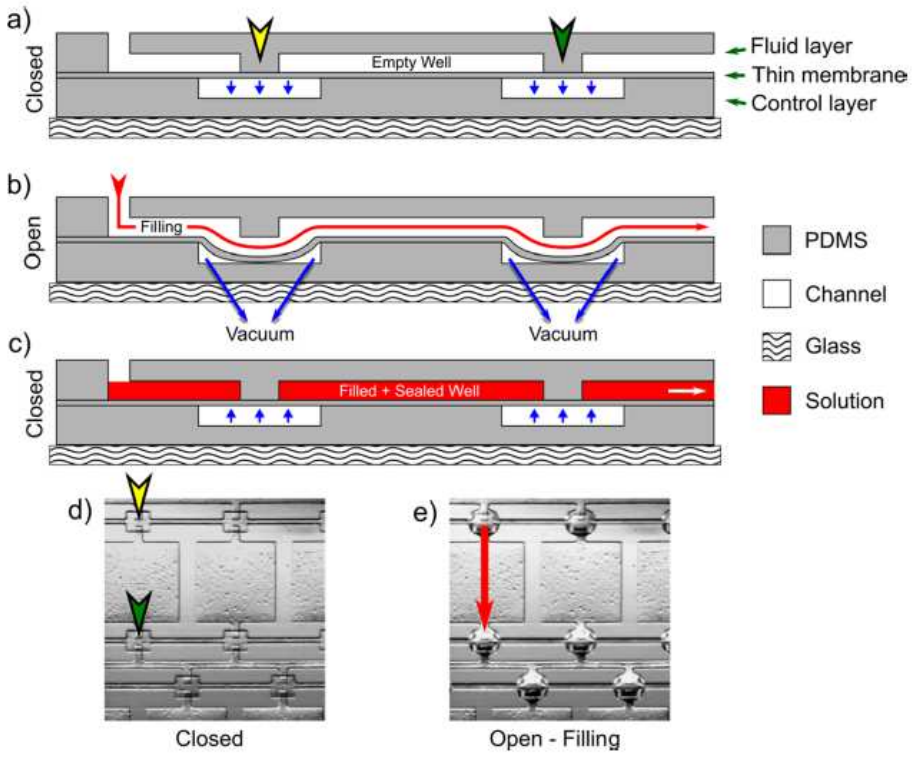


Figure 2.2: A schematic of the "doormat" valve and filling process from Abdallah *et al.* [80]. (a) The channels are blocked where a thin membrane is sealed against poly(dimethylsiloxane) (PDMS) (arrows) when there is no pressure applied. (b) When a vacuum is applied, the membrane is pulled away from the channel, allowing the channels fill with solution. (c) When filling is complete, the vacuum is relaxed and the valve returns to the original state, sealing the wells with solution inside. (d) Bright field shows a perpendicular view with the valves (corresponding arrows) closed and (e) open with a red arrow showing the channel pathway. Image courtesy of Abdallah *et al.* [80].

thousands of crystals within a x-ray-transparent device which can then deliver crystals *in situ* to the beam in random orientations easily is a very appealing approach for serial microcrystallography.

In parallel to the *in vitro* crystallization methods described above, *in vivo* crystallization within insect and mammalian cells was introduced to produce quality microcrystals from proteins which don't form *in vitro* crystals in their native form [11, 49, 71, 81]. It is possible that this method commandeers the

host cell's mechanism for protein storage and toxin protection to build small, well ordered protein crystals within the cell. Over-expression of the protein *in vivo* leads to crystal production which can either be imaged within the cells (*in-cellulo*), or harvested separate for data collection. Crystals on the order of 10-15 μm in length and 0.5 - 1.0 μm in width were grown within a cell by Redecke *et al.* [71] and subsequently isolated for diffraction. Large scale production of these microcrystals has already been shown by Gati *et al.* [11], though the overhead for cell growth, harvesting and transfer may be cumbersome for a pipeline SMX experiment.

There are several techniques for handling microcrystals once they've been grown, such as microchips which can isolate microcrystals into a narrow distribution of sizes (down to tenths of micrometers) [82], but are beyond the scope of this work.

2.3 Delivery devices

Serial microcrystallography depends highly on being able to consecutively shuttle many microcrystals into the x-ray beam. Gas-focused liquid jets used to perform SFX of microcrystals at XFELs (seen in Figure 1.2) have inspired the development of a bevy of small-crystal delivery devices at other current SR sources. The methods can be separated into two categories: flowing suspensions of crystals across the beam and fixed-target sample holders. Recent advances in source brightness, beamline optics and fast-framing detectors are enabling more routine room temperature data collection, and consequently many of the examples outlined below take advantage of these conditions.

2.3.1 Flowing suspensions of crystals

Ideally, microcrystals would be rapidly delivered into the beam with minimal surrounding material contributing to background scatter, save for what is required to prevent dehydration. For rapid exposure times, microcrystals need not remain in the beam for long, so a thin layer of mother liquor covering the crystal should suffice for hydration. In pursuit of this ideal, several groups have developed droplet or stream systems to transit crystals across the beam.

Low-viscosity injectors

The gas dynamic virtual nozzle (GDVN), or liquid jet injector, was the first device to deliver crystals in a flowing suspension of liquid across an XFEL beam at the LCLS [83]. Crystals suspended in a buffer solution are supplied in a liquid stream which is then focused to a 4 μm jet with an external gas sheath (Figure 1.2) [2, 7, 12, 67, 71, 83, 84, 85]. Being liquid-based, the devices have a fast flow rate of 10-25 μL per minute, and in combination with a pulsed source this results in fairly low hit rates (single crystal diffraction events) of 2-10%. Low hit rates reflect missing a majority of the sample, which in turn generates a high sample consumption of protein (10-100 mg) and long data collection times to collect a sufficient number of diffraction hits. Sample consumption is less problematic at storage ring sources, where the source is pulsed at very high rates, but fast flow rate, sample consumption and data acquisition time will still be too high for many experiments. With high crystal density or fragile crystal samples, the device also has issues with clogging of the injector and possible shearing of fragile crystals, which make it unappealing for comprehensive application.

A method to reduce the consumption rate of liquid-based delivery devices uses electric fields in place of gas to focus the liquid jet. Sierra *et al.* [87] initially developed the microfluidic electrokinetic sample holder (MESH). This device incorporates a high electric field to focus the liquid free surface into a tight jet. At the comparatively low flow rate of 0.14-3.1 µL per minute, the jet eventually breaks apart into highly charged droplets which are then shuttled across the beam [87]. The vacuum of the sample chamber causes congealing and freezing problems with liquid jets. Though it is incompatible with the GDVN, adding glycerol or PEG to the MESH liquid can increase viscosity and avoid these problems, and also adds heavier background contribution than simple mother liquor. To avoid adding heavy scatterers, Sierra *et al.* [86] then developed a co-centric flow electrospinning injector (CoMESH), seen in Figure 2.3, which adds

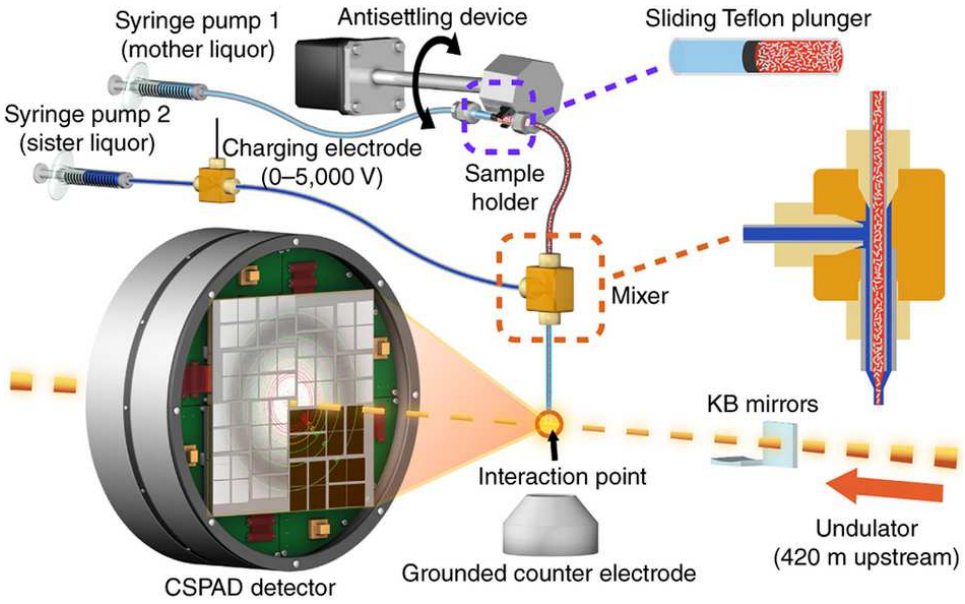


Figure 2.3: A schematic of the CoMESH injector where two pumps drive mother liquor containing microcrystals (red) and a hydrating, electrically charged "sister" liquor (blue) into a mixer (dashed orange square). The sister liquor creates a cocentric sheath around the mother liquor which then passes through the x-ray beam at the "interaction point" (orange circle). The diffraction is then recorded on a detector, such as the CS-PAD. Image courtesy of Sierra *et al.* [86].

a sheath of low-background hydration to surround the jet with comparable flow rates to the MESH injector. The slow-flowing MESH and CoMESH injectors may impart electrostatic charge on crystals (but the effect has yet to be explored). These devices can easily be adapted for storage-ring-based experiments.

High-viscosity injectors

Recently, several groups have developed higher viscosity injectors which reduce the flow rate to 0.05-0.48 μL per minute. This increases the crystal hit rate, decreases the sample consumption and therefore reduces the total number of crystals needed for an experimental run [8, 70, 87, 88]. In most viscous jet injector designs, hydraulic pressure is amplified from a pump to a nozzle to extrude viscous media through a column-defining capillary. The column is then stabilized

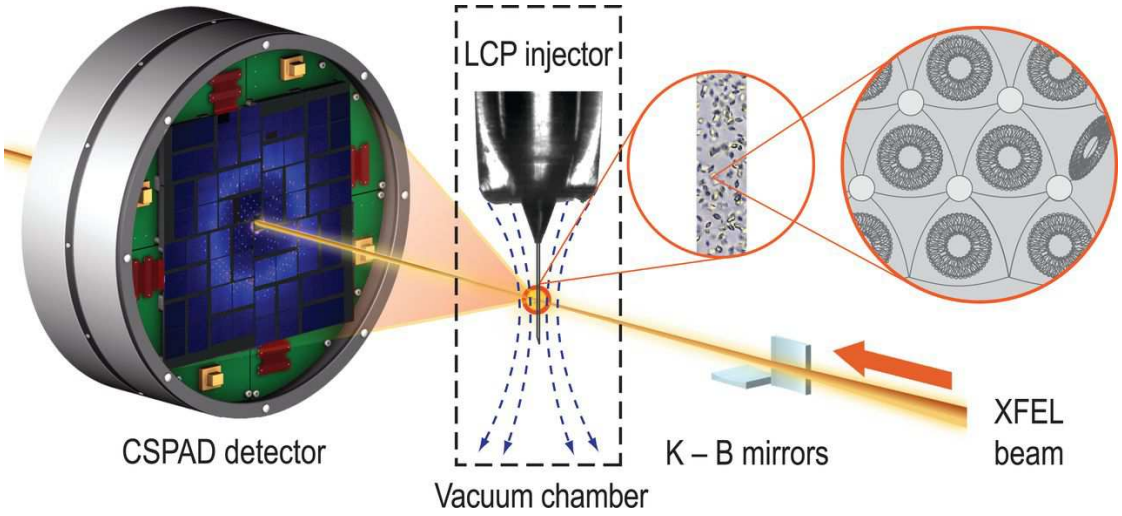


Figure 2.4: Schematic of the LCP injector setup at an XFEL. The LCP injector experimental setup shows an incoming XFEL pulse focused with KB mirrors interacting with microcrystals dispersed in LCP. The column of LCP is usually 20-50 μm in diameter and crystal diffraction data from the embedded microcrystals are recorded on a fast framing detector, such as the CS-PAD. Although a vacuum chamber is necessary at XFELs, this does not have to be the case at storage ring sources. Image courtesy of Liu *et al.* [70].

with a gas stream for diffraction and works in both vacuum and atmospheric conditions [51]. Originally designed for LCP-based membrane protein crystals, these viscous jet injectors are not limited by the medium used, and studies have shown success with agarose [65, 89], mineral oil-based grease [88], petroleum jelly [51] and PEG [90]. A specific viscous jet injector designed at Arizona State University (ASU) will be discussed in Chapter 4.

One disadvantage of the viscous jet injector extrusion is the size restriction of the capillary-defining column. Thus far, a ratio of ~5:1 capillary diameter to crystal dimension has been used to avoid clogging of the column (but see Section 4.3.3). Although the gel column itself adds material which contributes to scatter, the background can be reduced when using agarose in comparison to LCP or grease [89]. Unfortunately, this still adds material within the beam footprint surrounding the crystal. In searching for the microcrystal solution for delivery devices, viscous jet injectors may not be able to broach the limit of what is possible in background reduction at storage ring sources as there must be a minimum amount of gel around the crystal to maintain stability of the column which will add to the background scatter.

Recently, time resolved crystallography has benefited from mixing injectors developed in the Pollack Lab at Cornell University [91], which can rapidly mix reactant and sample milliseconds before exposure in the beam.

Pulsed droplets

Another delivery option is to create droplets from a sample reservoir containing microcrystals which then intersect with the beam. These droplets can be created

by a piezo-driven nozzle [92, 93] or directly from reservoirs [94, 95]. The nozzle, which defines droplet size, uses timed pulses from a generator to produce 80 μm diameter droplets to specific heights to coincide with XFEL pulses. To produce more closely matched droplet to microcrystal size, Soares *et al.* [94] designed an acoustic droplet ejection (ADE) system which can gently transfer microcrystals from crystallization trays or slurries onto mounts, such as mesh grids or a moving Kapton belt, and then deliver the crystals directly into the beam. Droplets can be scaled to approximately 12 μm, and additional solvent can be wicked away after mounting [96]. Roessler *et al.* [95] improved upon this for injection on-demand from crystal growth microplates with nanoliter to picoliter volumes.

All of the acoustic devices add background scatter, either by the amount of solution in the pulsed liquid droplets around the microcrystals, or in additional mounting material such as Kapton or mesh in acoustic ejection. Ideally, the amount of additional materials needs to be minimized for SMX.

The method of flowing microcrystal suspensions remains the workhorse of SMX, but background scatter contributions from the fluids which transmit the microcrystals across the beam, viscous or not, will remain a hard limitation for these devices. Perhaps more compromising is that the maximum hit rate of crystals typically doesn't exceed 10-30%, which sets a fairly high requirement for total numbers of crystals needed. When considering the numbers of microcrystals required in SMX, this does not compare well to the higher hit rate methods such as raster-scanning fixed-target devices.

2.3.2 Fixed-target imaging

Positioning the crystals within a fixed target sample holder increases crystal hit rate, as the microcrystals can be positioned into predetermined, or more predictable locations. In general, microcrystals are held stationary within the sample holder and moved through the beam either in a predetermined pattern or through raster scanning. Examples of fixed-target holders model the technological range available to crystallographers, from traditional cryocooling nylon loops to complex microfluidic chips and grids. As discussed in Section 1.2.3, there are concerns of potential orientation bias from fixed-target samples holders [97] and added background from support material with fixed-target devices, though careful design can avoid most of the problems.

Goniometer-based loops, capillaries and meshes

In an important milestone for microcrystallography at a storage ring source, Gati *et al.* [11] performed a helical raster-scan of over four hundred patterns from needle-like cryocooled cathepsin B crystals. The microcrystals were suspended across a 0.7 mm diameter loop in a thin film of mother liquor, as shown in Figure 2.5 [11]. The microcrystal hits were scaled to the overall diffraction strength in each pattern and merged into a complete 3 Å resolution data set. Though traditional in preparation, the slurry in the loop doesn't localize microcrystals to any specific region, making microcrystal identification a fairly tedious scanning process [77]. However, the method uses very little sample in comparison to flowed samples and doesn't show a preferential orientation bias of the microcrystals due to the flexibility of the helical scans.

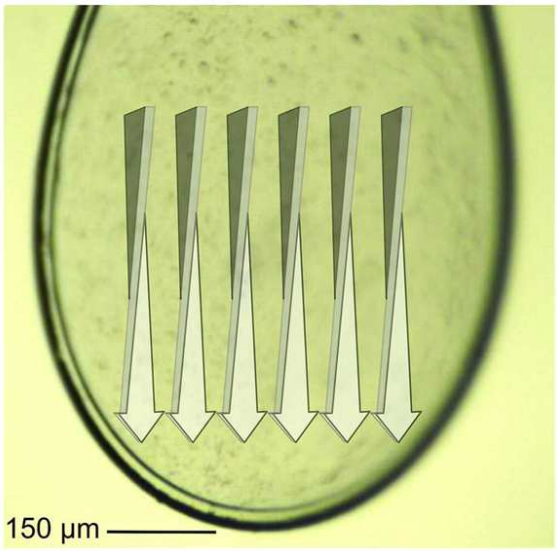


Figure 2.5: Serial microcrystallography in a fixed-target cryoloop. Schematic macroscopic illustration of the serial helical line-scan approach using a standard cryogenic loop. The helical scans allow greater flexibility in recording complete rotational coverage. Image courtesy of Gati *et al.* [11].

The first room temperature application of SMX at a storage ring was an elegantly simple model system experiment performed by Stellato *et al.* [33]. A series of short exposures from 3 µm × 3 µm × 6 µm lysozyme crystals in a 10 µm thin glass capillary produced a 2.1 Å structure from 40,000 single-crystal diffraction patterns [33]. Notably, this experiment proved that merging data obtained at storage ring sources can be used to solve the structure from tens of thousands of randomly oriented single-microcrystal diffraction patterns. The authors point out that the of merging data sets is not dependent on the delivery method; simply that serial microcrystallography is possible at storage rings sources, regardless of crystal mounting challenges. Of note is that both of these studies, Gati *et al.* [11] and Stellato *et al.* [33], required every image to be indexed in order to be merged into the complete data set.

In another goniometer-based delivery method, crystals grown *in vivo* within the protective environment of the cell can be left intact and mounted on mesh

grid support for diffraction. This technique of imaging *in cellulo* reduces the number of possibly damaging handling steps from harvesting *in vivo* grown crystals, thus increasing quality of data. The hit rate also increases, as cells containing crystals are generally easier to identify [98]. Boudes *et al.* [49] recently showed that diffraction from CPV1 polyhedrin within cells mounted on mesh frames were of higher quality and required fewer time consuming steps than purified CPV1 polyhedrin crystals. Although very attractive for proteins which are difficult to grow *in vitro*, this methods requires more nuanced preparation steps (such as cellular growth) and time to ready crystals than other methods for large scale data production.

Grids, sandwiches and chips

Serial microcrystallography benefits greatly from several highly engineered microcrystal-housing chips and grids. These have the ability to hold large numbers of crystals at room temperature, keeping them hydrated, and tend to keep sample consumption very low. In one solution for crystal hydration at room temperature, Coquelle *et al.* [45] designed a silicon nitride sandwich with 500 nm wafers for windows surrounding a microcrystal slurry. The thin windows enabled fairly low-background data collection and randomized orientation from microcrystals, though at random positions within the window.

Other chip designs restrict the locations of crystals to specific regions within the device by patterning hydrophobic and hydrophilic surfaces, or trapping crystals into grid-like patterns with pores or trap arrays, or even adding physical obstructions such as pillars to separate microcrystals [6, 47, 52, 99, 100]. The crystals are then localized for targeted diffraction (as seen in Figure 2.6), which

can increase hit rates to nearly 100% [100]. For example, thin silicon membrane chips with 1 μm holes patterned in a grids were developed to immobilize microcrystals from microliter drops of crystal suspension [50, 99, 101]. To minimize background scatter contributions, the chips can be processed to remove excess mother liquor by wicking.

Room temperature experiments require either enclosing the system in a humidified air chamber during data collection [101], or sealing off the chip with thin polymer or even graphene [78] to avoid dehydration during data collection (discussed in Chapter 3). To expand to higher crystal sample densities, Baxter *et al.* [102] designed a laser-cut 100-200 μm thick polycarbonate grid, which can accommodate up to 7200 samples within wells of 400, 200 or 125 μm in diameter [103]. Again, to function at room temperature, a 5 μm polycarbonate film is added on each side of the grid to prevent dehydration. Much consideration must be placed upon materials within the path of the beam, as any addition of polymer films or windows adds background scatter, though this depends heavily on the material used (and will be discussed in Chapter 3). Lastly, additional steps such as transferring crystals to chips run the risk of sample damage or loss before data collection.

The benefit of incorporating a delivery system into the device intended for crystal growth is that it avoids challenges with harvesting the crystals and eliminates any physical handling before diffraction. One clever method is to use thin crystallography plates to grow crystals and transfer the plate to the beam as a delivery system [104]. This circumvents the risk of crystal handling, but relies on the "uncontrolled" VD methods for crystal growth on plate.

Ideally, microcontrolled crystal growth (as mentioned in Section 2.2) is

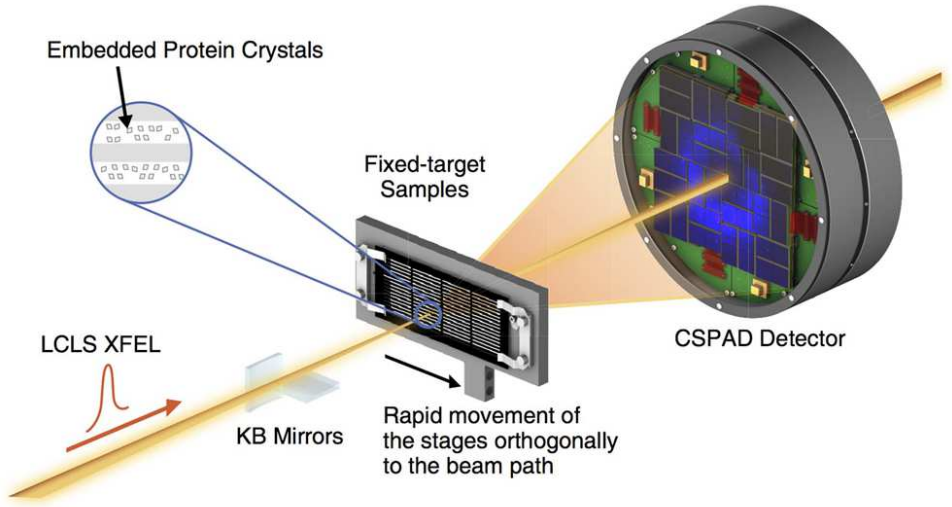


Figure 2.6: Schematic of a fixed-target sample setup at an XFEL. The fixed target experimental setup shows an incoming XFEL pulse focused with KB mirrors interacting with microcrystals dispersed within a sample holder. The holder can be moved rapidly orthogonally to the beam path and crystal diffraction from the embedded microcrystals are recorded on a fast framing detector, such as the CS-PAD. Although a vacuum chamber is necessary at XFELs, this does not have to be the case at storage ring sources. Image courtesy of Hunter *et al.* [6].

achievable directly at the beamline with microfluidics. Recall the ideal SMX with little to no excess background scatter. It is noted that all the examples above have material within the beam path at least an order of magnitude greater than the size of microcrystals. This is a challenge for eliminating sources of background scatter. However, the thin crystallization microfluidic chips introduced in Section 2.2 show promise for fixed target experiments with great rotational flexibility and ultra-thin support material [78, 79, 80, 104, 105]. More on microfluidics will be discussed in Chapter 4 with particular emphasis on a chip developed in collaboration with the Fraden Group at Brandeis [79].

2.4 Detectors

As discussed in Section 1.2.4, small crystal size and radiation damage limits diffraction from microcrystals. This puts high priority on detectors with small pixel size, fast frame rates, and single photon detection with no electronic background noise for experiments in SMX. Currently, most of the experiments performed at storage rings sources (in Table 2.1) use the PILATUS3 6M [106] (Dectris Ltd, Baden, Switzerland), a photon counting detector with 172 μm x 172 μm pixels.

There are several terms which are useful to know when discussing detectors. The rate at which the signal from a pixel is recorded is considered the readout time. Often, this sets a limit to the frame rate of the entire device, which is how quickly an entire image can be recorded and stored. A quick frame rate is essential to capture many diffraction images rapidly when a sample is a flowing quickly across the beam, or to optimize scanning, before radiation damage limits the diffraction signal (see Section 1.2.4). A PILATUS 6M detector typically frames at 25 Hz with a readout time of 2 ms, though newer versions can frame up to 500 Hz with a readout time of 0.95 ms. Recently, Dectris released the EIGER 1M detector [107], which offers a pixel size of 75 x 75 μm, frame rates up to 3 kHz and a readout time of 3.8 μs [107]. Currently, rates of image acquisition range from 133 Hz for the large EIGER X 16M with more than 16 million pixels to 750 Hz for the EIGER X 4M. Detectors promise to reduce the time for data collection significantly.

For all detectors, accurate measurement is not simply a matter of counting x-ray photons as they strike the detector. The signal recorded is only accurate

to the shot noise of whatever is being measured, plus various dependencies on window absorption, sensor stopping power, *etc*. One concern for photon-counting detectors is if the counter doesn't record a photon when it hits between neighboring pixels [108, 109], although some proposed detectors claim to eliminate this concern [110]. Another concern is the saturation of pixels from high numbers of incoming photons and slow counting time.

In most of the studies in this dissertation, we use the fast-framing photon-integrating Mixed-Mode Pixel Array Detector (MM-PAD) [42] or the chip that makes up the CS-PAD detector installed at the LCLS Coherent X-ray Imaging (CXI) beamline [111, 112]. Since these detectors integrate photon hits, they avoid missing photon hits between pixels. Additionally, the MM-PAD has a high dynamic range for detecting anywhere from single photons and to 10^7 photons per pixel and, most importantly, has a frame rate up to 1kHz [42].

For completeness, we note some proposed detectors not yet in commercial use show much promise for SMX, such as the JUNGFRAU 1.0 hybrid pixel detector [113], the AGIPD 1M (Adaptive Gain Integrating Pixel Detector), the Ultra Fast X-ray Camera 32k (UFXC32k) photon-counting dual-counter PAD [114], and the KECK PAD [115] from the Gruner Group.

2.5 Software: Data handling and processing

Due to the large numbers of microcrystals needed, their small size and weaker diffraction, it is often challenging to identify whether or not an image has single, if any, microcrystal Bragg diffraction. Depending on the delivery device and if the position of the microcrystal is known, a microcrystal can be accurately

placed into the beam for diffraction and timed such that images are known to be likely to have microcrystal diffraction. If the crystal size is smaller than visual resolution limits or is obscured by the buffer medium, inline low-dose raster-scanning programs such as *DOZER* [59], *DISTL* [116], *NanoPeakCell* [117], and *STXL* (scanning transmission x-ray microscopy) [118] could recognize and rate crystal hits, recording crystal position within mounting material. With raster-scanning, resolution is limited by the grid scan in beamsize and grid-step size, and provides only a 2D characterization of the crystal. Using online x-ray micro-radiography/microtomography for crystal visualization and positioning will allow for 3D characterization, but is limited by the resolution of the setup, at least in one study, to 2 μm [119]. These methods will help identify microcrystals if present within the beam, and reduce exposure of microcrystals to unnecessary radiation. Offline imaging using second order non-linear imaging of chiral crystals (SONICC) [120] and two photon-excited fluorescence (TPEF) microscopy [121] have also been used to identify crystal position without damaging x-ray radiation (SONICC is especially good at discriminating between protein and salt crystals).

As radiation damage is still a limiting factor with microcrystals, it is sensible to use as little pre-data-collection radiation exposure as possible to maximize the diffraction data quality. The most popular methods either raster-scan or flow crystals across the beam continuously while collecting data, and allow software to analyze images and eliminate those without microcrystal diffraction. *DOZER*, *XDS*, *Cheetah*, and *CrystFEL* all are packages which are developed to handle the large influx of data images and low Bragg peak counts inherent to microcrystallography. Newer programs such as *Cheetah* [122] which was originally designed for XFEL experiments, and more traditional crystallog-

raphy programs such as *HKL-2000* [123], *XDS* [124] and *MOSFLM* [125] utilize Bragg peak spot finding algorithms to identify images with crystal diffraction. For this identification, Bragg peaks necessarily must be discernible from background scatter. Subsequently, individual images are indexed for crystal orientation through fast Fourier transform algorithms [126, 127]. Images are integrated, scaled and merged to form complete 3D Bragg reciprocal space data sets using most of the same programs mentioned above, in addition to others such as *CrystFEL* [127, 128, 129], *PHENIX* [130], and *ADXV* [131]. Some merging programs use Monte Carlo integration, where each Bragg peak is indexed so that the like Miller indicies are summed over the average of crystal sizes and shapes. Lastly, phase information can be approximated to create a model phasing for the experimental data through techniques such as Molecular Replacement (MR) from the *PHENIX* or *MOLREP* [132] programs. Many of the above programs, such as *CrystFEL*, help to build and refine the model, including refinement tools *BUSTER* [133] and *REFMAC* [134], and model building with *Coot* [135].

All of these packages use Bragg peak identification to index and orient data frames on a per-frame basis. If Bragg peaks are not distinguishable, then the images are discarded. This is where the application of *EMC* mentioned in Section 1.2.5 would benefit microcrystallography (discussed in detail in Chapters 5 and 6).

2.6 Summary

The most compelling motivation for continued development of SMX is that the limits of what is possible at storage ring sources are not close to being met, yet.

As crystal size decreases and the techniques above honed, one expects much improvement to be seen with SMX at storage ring sources. Many of the aspects of the crystallography experiment should be optimized for SMX performance including improving beamlines, reducing background scatter, reducing crystal handling, reducing sample consumption, and dealing with the large data set sizes.

This dissertation will look at one aspect which is not addressed anywhere in this chapter review. Every reported structure in the studies above require the data images to be indexed in order to merge into a complete data set. This dissertation explains how applying the *EMC* algorithm can eliminate that boundary of indexability, and obtain information from sparse data frames which are currently considered useless in Chapters 5 and 6.

CHAPTER 3

BACKGROUND REDUCTION FOR SMX

As mentioned in Section 1.1.2, every diffraction image contains two types of photon counts: Bragg diffracted photons from the structured crystal lattice and scattered photons from everything else the beam encounters from the source to the detector. In traditional crystallography, identifying Bragg diffracted photons from other scattered photons is essential to the process of structure solution. Microcrystals deliver very few Bragg photons before they are irreversibly damaged by the radiation. Measurement of these Bragg photons is impeded by superimposed background x-ray scatter. A convenient standard for the quality of this measurement is the signal-to-noise ratio (SNR), $\frac{I}{\sigma(I)} = \frac{N}{\sqrt{N}}$, from Poisson statistics. The "signal" of the Bragg diffracted photons, N, becomes difficult to distinguish from the superimposed "noise" of other photons when their ratio drops below roughly 2. Holton and Frankel explain the implications of this ratio in the "best case" scenario calculation introduced in Section 1.2.1, where a complete data set can be obtained from a 1.2 μm crystal with zero background contributions. It implies that only N = 4 photons per spot are needed to measure a Bragg diffraction spot, if no other noise obscures them. While current SMX experiments are far from ideal, it nonetheless presents an encouraging prospect for microcrystallography if all other noise is eliminated.

Modern "noiseless" x-ray detectors, (discussed in Section 2.4), are capable of detecting every x-ray with nearly 100% efficiency [136]. Since any background scatter contributes to "noise" [137, 138], eliminating those sources will allow detection of every x-ray scattered from a microcrystal to a given resolution. This suggests a need to eliminate background as much as possible. Proof of the bene-

fits of background reduction have already been seen with SMX at XFELs, where the entire beam path is in vacuum, as x-ray windows do not survive the intense x-ray pulse, and the sample is delivered *via* comparatively low-background jet injectors.

This chapter identifies various aspects of the storage ring beamline environment and sample setup which require re-engineering to reduce background scatter. It also describes in detail a path towards a practical ultimate in background reduction by use of atomically thin graphene sheets as a crystal mounting platform for protein crystals [139]. The results show how the overall signal-to-noise ratio per unit dose for x-ray diffraction data from protein crystals can be improved by reducing the mass and density of all material surrounding the crystals.

3.1 Common sources of background scatter

Regardless of makeup, any material within the beampath will scatter x-rays as the electrons within any matter may interact with x-rays. Since x-ray scattering is dependent on the number of the electrons with which the x-rays interact, the electron density and volume of any material within the beam path define the intensity of the scattered x-rays. This scattered intensity scales almost linearly with the width of material along the dimension of the beam.

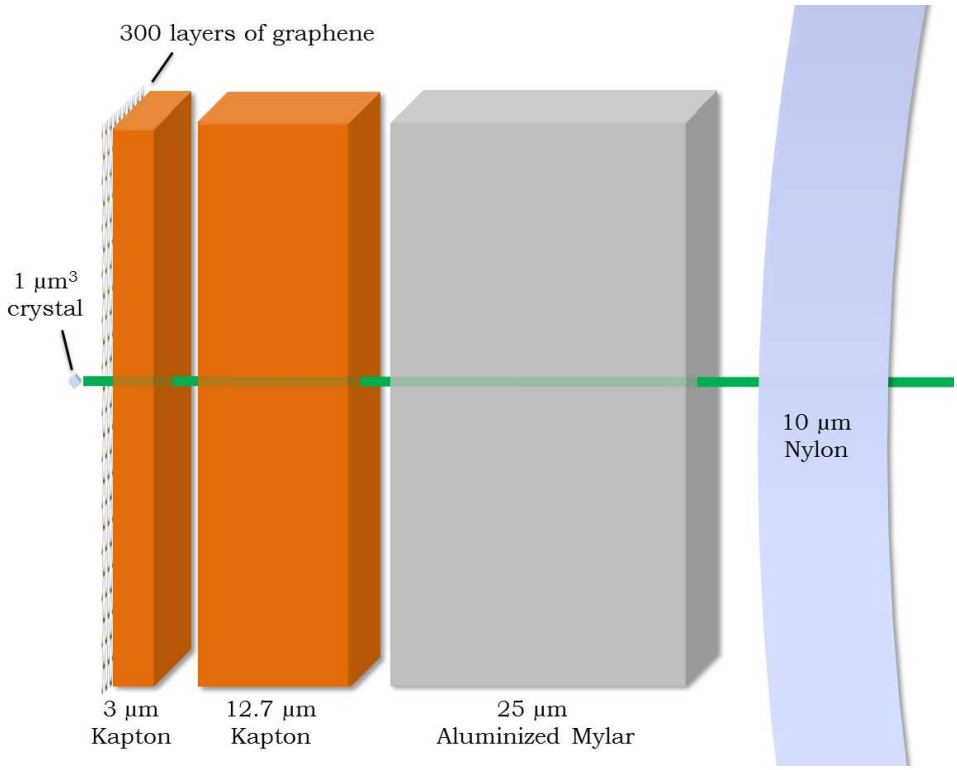


Figure 3.1: Example sources of background scatter within a microbeam for dimensional scale. Given a 1 μm^2 microbeam (green), the amplitude of scatter from material within the footprint is dependent on the electron density and thickness of the material within the beam. The minimum background achievable would be what comes from just the mother liquor within a hydrated crystal, which makes up approximately 50% of the crystal volume. Consider a 1 μm^3 protein crystal shown on the left. Window material, such as the Kapton films drawn to scale (orange), will increase that background proportionally to the volume illuminated. Kapton of 3 μm thickness would scatter ~6 times more background than the signal from the crystal, or a 10 μm nylon cryoloop would scatter ~20 times more. As more material is added, represented by blocks in this figure, background increases in proportion. Using thinner windows, such as multilayer graphene shown next to the crystal, is ideal for background reduction. Even 300 layers of graphene, measuring less than 50 nm thick, would contribute ~12 times less background scatter than the water in the crystal. In practice, only a couple layers of graphene are needed for graphene windows or mounting material, and would insignificantly contribute to the background. Air has roughly 10^{-3} the density of water; hence 1 mm of air contributes more background than the internal water of a 1 μm^3 crystal! This figure is slightly modified from "Biostructural Science Inspired by Next-Generation X-ray Sources" by Sol Gruner and Ed Lattman in Annual Review of Biophysics [14].

3.1.1 Beamline external to the crystal local environment

It is crucial to step back and evaluate the entire beamline environment to identify sources of background scatter. The majority of beamline-specific background scatter contributions come from air within the beam path, or parasitic scattering from optics, pinholes or beam-defining slits [140]. Windows to separate upstream beampipe sections from experimental hutches will also contribute scatter. Any windows to or air within ion chambers will contribute scatter. Finally, even the divergence of the beam will add unnecessary background scatter contributions, often further compounded by parasitic scattering from the other beamline components mentioned above.

3.1.2 Local crystal environment

Perhaps the largest contribution to background noise comes from the local crystal environment. Beginning with the windows which separate the sample environment from the beampath, to the coatings on the crystal to maintain hydration, each layer of material adds background scatter. These materials include external mother liquor [140], coating oil, suspension medias, enclosing capillaries [141], and the polymers that make up most grids and microfluidic chips [78]. Most of these materials are used to hold the crystal in the beam while staving off dehydration and/or protecting the crystal from the damaging effects of ice formation during cryocooling [21, 22, 142, 143, 144, 145, 146, 147, 148, 149]. However, these solutions for mounting, dehydration reduction and cryoprotection scatter x-rays which, for weakly diffracting crystals such as microcrystals, can hide the structure signal. Eliminating or reducing these sources of scatter is

crucial to measuring the small diffraction signals in microcrystallography.

If the beam cross section is larger than the diameter of the crystal, scattering materials contribute in proportion to their volume in the beam. Simply matching the beam size to the crystal cross section limits the interaction of extra material with the beam to that of the crystal footprint (see Figure 3.1). Consider a 1 μm^3 protein crystal held within a (2 μm)^3 thin film of water within the 1 μm^2 beam footprint. A protein crystal is considered to be ~50% water, so the thin film would scatter 4 times more compared to the liquid within the crystal. Add any other materials or air within the beam path, and the background scatter increases proportionately to whatever volume is added; *e.g.*, a 20 μm diameter liquid suspension containing the crystal would contribute 40 times more scatter to the background than the liquid from within the crystal.

The last source of background scatter is the solvent in channels within protein crystals. Typically, protein crystal volume contains 40-60% water which fills the void volume, or channels, between ordered protein molecules. Without the internal solution channels, the architecture of the crystal crumbles, and ordered diffraction is no longer possible. Therefore, the irreducible limit of background contributions in the local crystal environment is the water internal to the crystal.

3.2 Ideal background reduction, in principle

The objective is to experimentally approach this irreducible background limit at storage ring sources and make the inherent ordered protein crystal diffraction the most prominent source of x-rays seen by the detector by eliminating extraneous scatterers. Start with a collimated beam, free of parasitic scattering

from optics, pinholes or beam defining slits. Either eliminate window material or replace components with less electron dense alternatives. Ultra-thin vacuum windows of silicon nitride can reduce the thickness of windows from hundreds of nanometers to tens of nanometers. Graphene can reduce it to atomically-thin layers and effectively eliminate background scatter.

Replacing open air flight paths with helium or, preferably, vacuum flight paths will reduce scatter from air. For room temperature data collection, this might require additional material to protect protein crystals from dehydration, either by hydrating the helium, immediately surrounding the crystal *via* a sample chamber with ultra-thin windows, or using ultra-thin crystal mounting material. But these materials should also be individually evaluated for their background scatter contributions. One can envision graphene-encased microcrystals held in an otherwise complete vacuum path environment where there is no main beam air scatter. In other cases, where a microcrystal must be kept cryocooled, one can imagine a vacuum path that has a graphene vacuum window just short of a crystal held in a helium gas cryostream, followed by another graphene window into a vacuum path that encloses the beam stop and has a conventional vacuum window at the detector. Electron microscope studies [150] and helium gas studies [151] have already shown high-quality graphene sheets to be vacuum, water and helium gas tight on micron length scales. Single layer graphene windows tens of microns across have already been produced, therefore single or multi-layer graphene vacuum windows are feasible and would be a big step for microcrystallography. In the cryocooled case, the total beam path in a cryocooled He gas stream can be reduced to very short lengths, *i.e.*, smaller than a millimeter. The background scatter of such a system can be reduced to negligible levels relative to the diffraction from micron-sized crystals.

There are many experimental ways to approach the irreducible limit, discussed next.

3.3 Review of current background reduction techniques

Many of the current studies in serial microcrystallography at storage ring sources echo the growing awareness for eliminating background scatterers at storage ring sources [11, 33, 49, 47, 50, 53, 101] and various techniques of reducing background scatter are slowly being introduced into practice. Effectively matching beamsize to the crystal, Evans *et al.* [152] reported a dramatic decrease in background signal when shrinking a 8×8 μm^2 beam to 4.5×5 μm^2 beam while imaging 5 μm^3 polyhedra crystals. However, this is not yet common practice at all beamlines. This measure is absolutely critical regardless of additional reduction measures, because unless the microcrystal is the only other component of the system, the x-ray beam will scatter from anything it illuminates. A few storage ring beamlines have incorporated vacuum (*e.g.*, Diamond I04 [153]) or helium chambers (*e.g.*, SPring-8 BL32XU [154]) to reduce scatter from air, but most crystallography beamlines used for microcrystallography have large in-air sample chambers.

To reduce scatter from the local crystal environment, many recent studies have chosen to eliminate material surrounding microcrystals, as Roedig *et al.* [50] and Baxter *et al.* [102] showed by wicking mother liquor solution away from micro-patterned SMX chips (though this required additional cryocooling to avoid dehydration while collecting data). For room temperature samples, the windows of sandwich-style delivery platforms can be replaced with much

thinner silicon nitride wafers, only 500 nm thick [45], or thin polymer films for total sandwich thicknesses less than 40 μm [104]. For liquid jet injectors, a clever redesign of the electrospun injector (mentioned in Section 2.3.1) eliminated the need for denser cryoprotectants, which contribute more intense scatter compared to mother liquor [86]. The use of less dense medias in high-viscosity injectors, such as replacing LCP with agarose, also lowers background scatter [89]. Reducing the size of the extrusion columns further diminishes their background contribution [51]. Even the designs of microfluidic chips discussed in Section 2.3.2 look to reduce the amount of material within the beam path, either by replacing thick polymers with thinner polymer membranes such as Kapton [47], thin silicon membranes [50], and/or graphene [78]. Commercially available silicon nitride windows can be purchased with thickness of 50 nm. Graphene has been shown to be an air-tight window material in electron microscopy and more recently a mounting material in crystallography [139, 153].

For example, Sui *et al.* [78] incorporated graphene into a poly(methyl methacrylate)(PMMA)-based microfluidic chip to grow and deliver microcrystals to the beam while remaining remarkably ultra-thin. Figure 3.2 demonstrates the increase in scattering, seen as a darker ring in (b), with the addition of cyclic olefin copolymer (COC). The SNR within that region is diminished due to the stronger scattering of the polymer, obscuring Bragg peaks. Removal of COC reduces the thickness of the chip by two orders of magnitude, resulting in a dramatic clarity of diffraction signal in (c). A radial integration over the entire diffraction image for various device materials is shown in (a). Sharp peaks in pink and orange identify crystal diffraction and scattering bands around 2 Å and 5 Å represent the more diffuse scatter from COC which obscures peaks within those resolution bands.

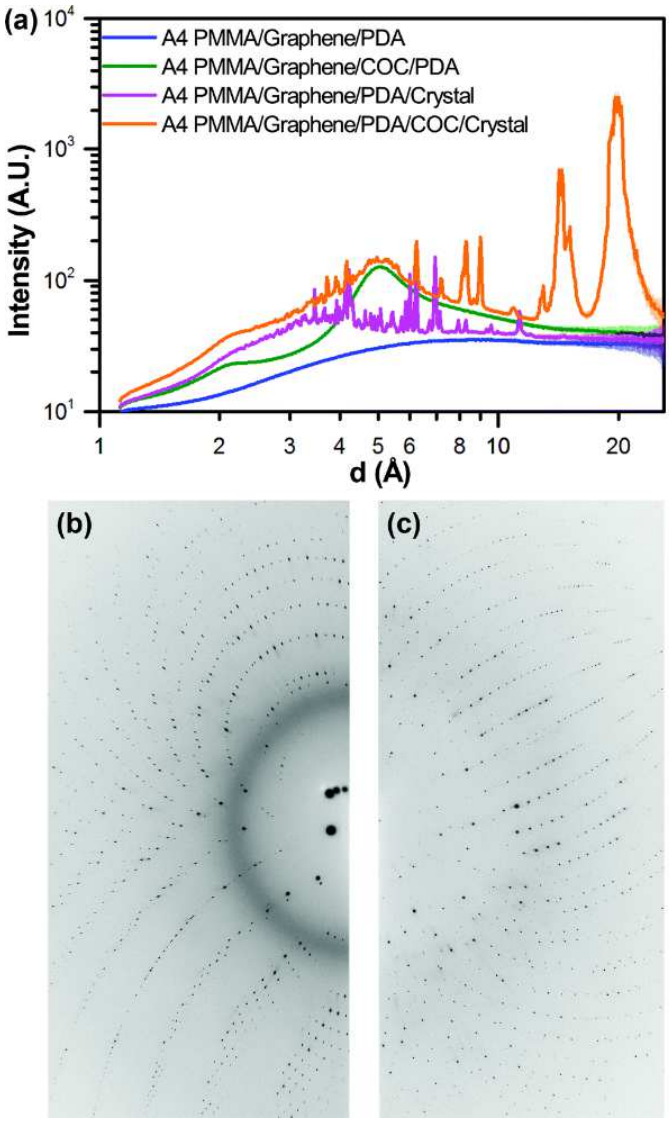


Figure 3.2: Intensity comparisons from various microfluidic chip materials. (a) Radial integration of the x-ray intensity profile showing relative strength of diffraction from a hen egg white lysozyme crystal compared to various microfluidic chip materials as a function of resolution. Sharp peaks identify crystal diffraction, while background contributions show more broad effects. (b) shows the corresponding diffraction images from the A4 PMMA/graphene/PDA/COC/crystal dataset (orange in (a)), which represents the heaviest contributor to scatter, and (c) shows the A4 PMMA/graphene/PDA/COC/crystal dataset (magenta in (a)) representing the thin microfluidic chip that reduces background. Reproduced from Sui *et al.* [78].

It is clear that current SMX experiments (from Chapter 2) are far from the lowest-background ideal proposed in Section 3.2.

3.4 Proof-of-principle: Graphene as background reduction material

Graphene is an attractive barrier material because it is composed of low atomic weight carbon and can be obtained in macroscopic sheets that are only a single atomic layer thick, yet remarkably strong [155, 156]. Graphene is a planar layer of sp^2-bonded carbon atoms and has an attenuation length of 4988 μm for 13.5 keV x-rays. The thickness of each of these graphene layers is about 0.34 nm, so a few layers are essentially transparent to the x-rays used for crystallography.

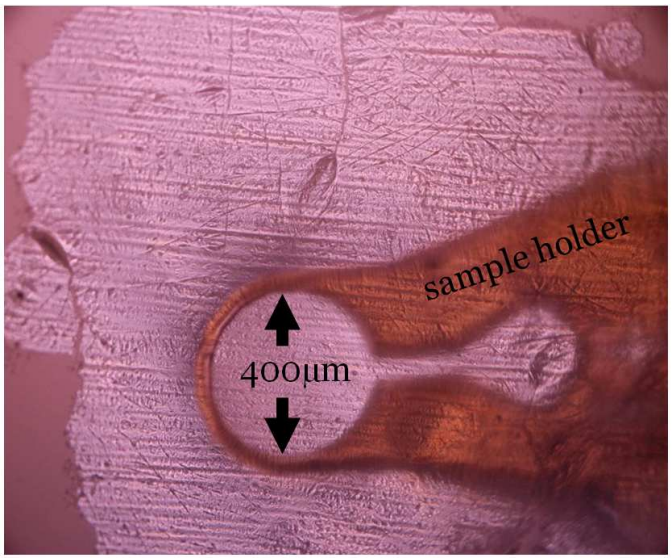


Figure 3.3: An example of a graphene-incorporated sample holder. A sheet of graphene is seen suspended across a 400 μm MiTeGen loop as an example of incorporating graphene into mounting materials. The large sheet is a multilayer of graphene, described later in Section 3.5.1. The graphene can act as a support and a hydration barrier for crystals.

Recent advances in the development of graphene membranes have allowed for relatively easy fabrication of sheets of carbon having the thickness of a single atomic layer using chemical vapor deposition [157]. A single layer of graphene has been shown to be impermeable to standard gases, including helium at several atmospheres of pressure, in addition to being optically transparent [151, 158]. In recent studies, pockets of liquids containing crystals between layers of graphene have been used successfully even in the high-vacuum environment of transmission electron microscopy, resulting in improved image resolution and signal-to-noise ratios for acquired electron microscopy images [150, 159].

If it can be experimentally shown that few-layer graphene can be used as ultra-low background, gas-tight, conformal windows around protein crystals, then the ideal of Section 3.2 is plausible.

3.5 Graphene-wrapping protein crystals for background scatter reduction

The contents of this section have been published in the *Journal of Applied Crystallography* by Jennifer Wierman, Jonathan Alden, Chae Un Kim, Paul McEuen, and Sol Gruner [139].

Incorporating graphene into materials can accomplish two objectives: reducing volume of mounting material and creating a gas-tight barrier. As such, graphene crystal mounts can clearly be of importance in microcrystal diffraction. In the following section, we describe a study which demonstrates the fea-

sibility of graphene as a crystal mounting substrate. We show that graphene may be used to wrap and support protein crystals in both cryocooled and room-temperature crystallographic experiments. This demonstrates the plausibility of using graphene as a solution for eliminating background scatter.

3.5.1 Experimental setup

To compare diffraction in mounting materials, we applied five different mounting methods. These methods include having both samples mounted using more traditional methods and a novel technique involving graphene-wrapped samples. The graphene-wrapped sample preparations include one method in which the sample is left at room temperature for over 10 minutes before flash-cooling, to show the robustness of the graphene covering as a barrier for prevention of dehydration.

Protein crystallization

Tetragonal crystals were prepared using the hanging drop method [160] with lyophilized thaumatin powder from *Thaumatococcus daniellii* (Sigma-Aldrich, St Louis, MO, USA). The powder was resuspended in deionized water to 25 mg/ml and 50 mM Hepes buffer at pH 7. The crystals were grown at 293 K in hanging drops comprising 2 μL of protein solution combined with 2 μL of a reservoir solution containing 0.9 M sodium potassium tartrate. The handing drops were suspended, on a siliconized glass coverslip, over 800 μL wells of reservoir solution in a 24-well plate and the assembly was sealed with vacuum grease. Small clear crystals appeared overnight and were incubated at room

temperature for a few days until truncated bipyramidal crystals approximately 100 μm across were obtained. The crystals used in this study were almost identical in size for each method, with size chosen to match the footprint of the beam. After unsealing, the hanging drops were mixed with glycerol to a final concentration of 0.9 M NaK tartrate with 10% glycerol for cryoprotection. Crystals were equilibrated for 1-2 minutes in this solution before being flash-cooled in liquid nitrogen, oil-coated and flash-cooled or wrapped in graphene, as described below.

Oil-Coating

In cases where a crystal was to be coated in oil as cryoprotection for comparison purposes, the hanging drop was covered directly with NVH oil (Hampton Research, Aliso Viejo, CA, USA) to prevent dehydration. The crystals were gently extracted from the mother liquor and external mother liquor on the crystal removed with gentle swishing so as to leave a tail of solvent behind with a cryoloop (Hampton Research). This was repeated until little-to-no visible solvent was left on the exterior of the crystal. The cryoloop was also used to extract the crystal from the oil droplet with as little external excess oil as possible and flash cooled in liquid nitrogen.

Graphene Chemical Vapor Deposition

Large-grain graphene was grown on copper foil by chemical-vapor deposition (CVD), using a method developed by Li *et al.* [161]. In this method, the copper growth foil was enclosed in a second copper foil to limit exposure to growth

gases, and placed in an evacuated furnace (base pressure 10^{-4} torr). In a slight modification from Li *et al.* [161], the foil was annealed while flowing hydrogen at 60 standard cubic centimeters per minute (s.c.c.m.) at 1253 K for 45 minutes and then cooled to 1203 K. A flow of 3 s.c.c.m. of methane was added to the hydrogen, then the temperature was ramped to 1253 K over the course of an hour, and held at 1253 K for 3 hours, and finally the system was cooled to room temperature and the gases turned off. The resulting graphene had 30-100 μm grains, with small ~2 μm patches of bilayer at graphene nucleation sites.

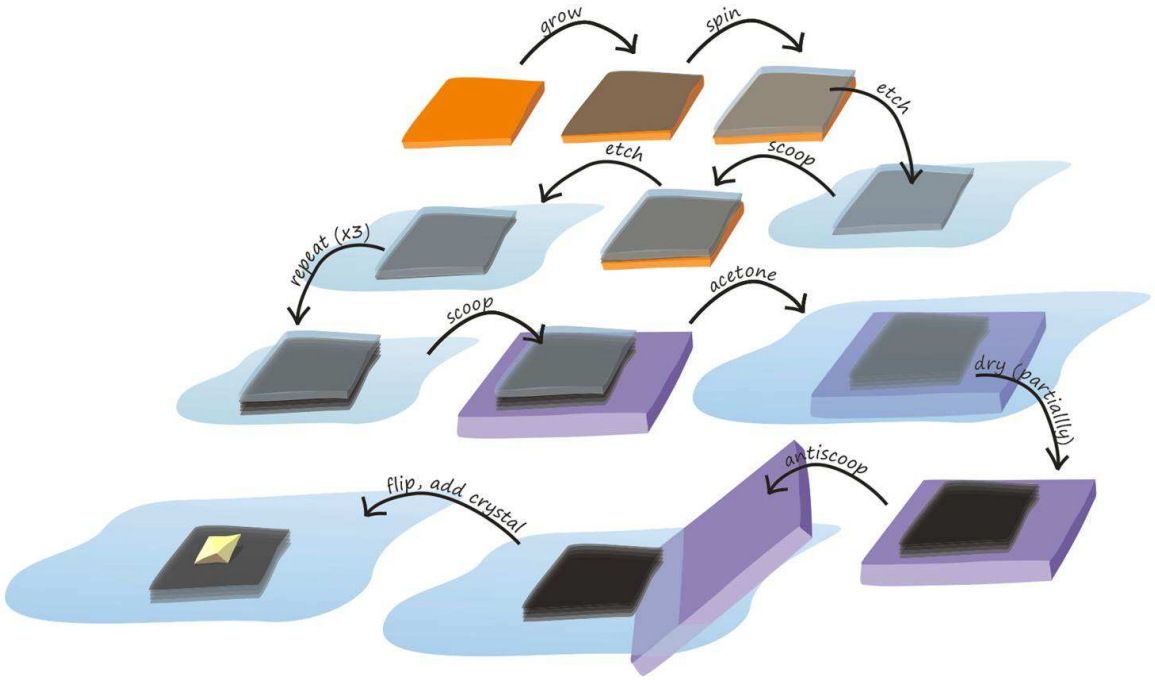


Figure 3.4: A graphic of CVD graphene production and layering. Graphene (dark gray) is grown onto copper (orange) *via* CVD. Next, PMMA is spun onto the graphene to aid the transfer process. The copper is etched from the backside of the graphene and transfered to water several times to rinse the graphene clear of etchant. Another layer of graphene-on-copper is used to scoop the PMMA-graphene layer from the water bath, and allowed to dry. The copper is then etched from the backside of the PMMA-graphene-graphene layers, and the process repeats until the desired number of layers is achieved. The PMMA can be dissolved from the graphene layers with acetone, and the resulting layers of graphene remain floating on the water-air interface. From here, pieces of multilayer graphene can be selected for crystal-wrapping.

Graphene Polymer-Transfer

A schematic of the graphene-polymer transfer is seen in Figure 3.4. We spun 50-200 nm of PMMA onto one piece of the graphene-on-copper. After scratching the graphene off the back of the copper foil with steel wool, the copper was etched by floating it on the surface of an ammonium persulfate-based etchant (Transene APS 100). The PMMA-on-graphene was rinsed in multiple water baths, and transferred to a second PMMA-free graphene-covered copper foil by scooping it out of the water using the foil, and letting it dry. The copper was again etched, and the now-two layers of graphene under PMMA were rinsed. This process was repeated until either three or five layers of graphene are stacked on top of each other underneath the PMMA. Finally, the PMMA-on-graphene was scooped out of the water using a piece of quartz wafer, and the drops of water were gently blown off using a nitrogen gun, while leaving a thin layer of water between the quartz and graphene. The wafer was then dipped in acetone and left for ~3 minutes, during which time the PMMA dissolves and was mostly removed. It was gently lifted out of the acetone and slowly lowered back into the water, at which point the PMMA-free, multilayer graphene lifted off the surface of the quartz (presumably as a result of the thin water layer still existing between the graphene and quartz), to float on the surface of the water. This process yielded intact multilayer graphene on the centimeter-scale, which could then be broken into smaller pieces and transferred to loops and other sample substrates.

Graphene Preparation

The multilayer graphene was broken into smaller pieces of approximately 1 x 1 mm with tweezers while resting on the surface of the water. These graphene pieces were then swept out in a suspension of deionized water with a small copper loop (5 mm in diameter) so that the multilayer graphene floated on the top of the water droplet within the copper loop. The entire copper loop was inverted so that the multilayer graphene floated on the bottom of the suspended water droplet. The drop was then washed with protein crystallization reservoir solution several times, ensuring the conditions within the suspended droplet were near identical to the crystallization hanging droplet for the protein crystals.

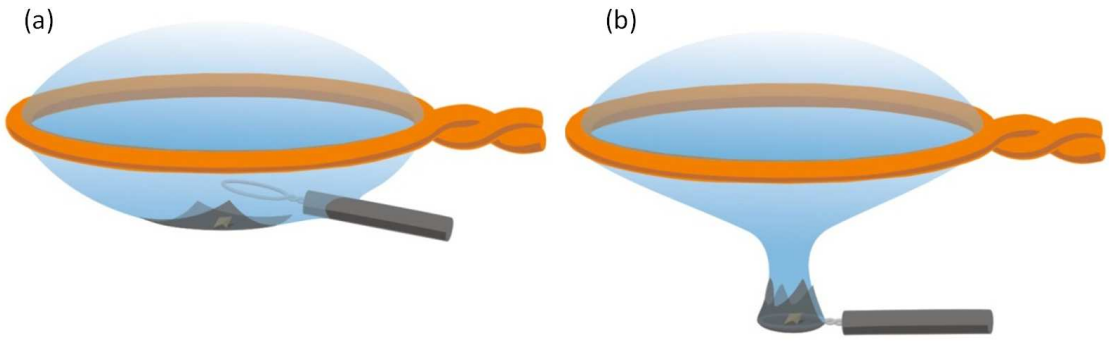


Figure 3.5: Graphic depicting how the graphene and crystal are mounted on a Hampton cryoloop. A copper loop supports a droplet of mother liquor solution. Graphene is floated on the bottom of the droplet, and a crystal is dropped in solution onto the hydrated side of the graphene. (a) The cryoloop is then inserted into the droplet below the copper loop, oriented horizontally with the crystal situated in the center of the loop and (b) dragged out through the bottom of the solvent droplet. This causes the graphene to wrap closely around the crystal and the cryoloop.

Graphene-wrapping

The crystals were transferred *via* a pipette to the suspended drop in the copper loop containing the graphene sheet. The graphene floats at the water-air interface and can be visually located as a dark film with well-defined edges compared to the solution around it. Crystals sank to the bottom of the drop by gravity and settled near the water-air interface. A cryoloop (Hampton Research, Aliso Viejo, CA) was used to carefully sweep the crystal to the center of the multilayer graphene piece using advection, being careful not to contact the crystal, as shown in Figure 3.5(a). Once the crystal was properly seated on the multilayer graphene piece, the cryoloop was used to sandwich the crystal inside the multilayer graphene by sweeping the horizontal loop downward through the droplet, onto the multilayer graphene beside or around the crystal and finally out the bottom of the drop, as shown in Figure 3.5(b). This causes the multilayer graphene to wrap around the crystal as the loop is pulled away from the larger droplet, and the surface tension draws the graphene around it. Figure 3.6 shows two different crystals enclosed within three-layer and five-layer graphene sheets, for a total of either ~ 6 layers or ~10 layers of graphene in the path of the beam. In most samples, the graphene wraps closely around the crystal and suspends it within the cryoloop as in Figure 3.6. Figure 3.6(c) shows how the graphene creases within the loop from the contours of the crystal.

The following three post-treatments were applied to the five-layer graphene-wrapped-crystal samples:

(1) Samples were immediately flash cooled and diffraction data collected (an example is shown in Figure 3.7(d)).

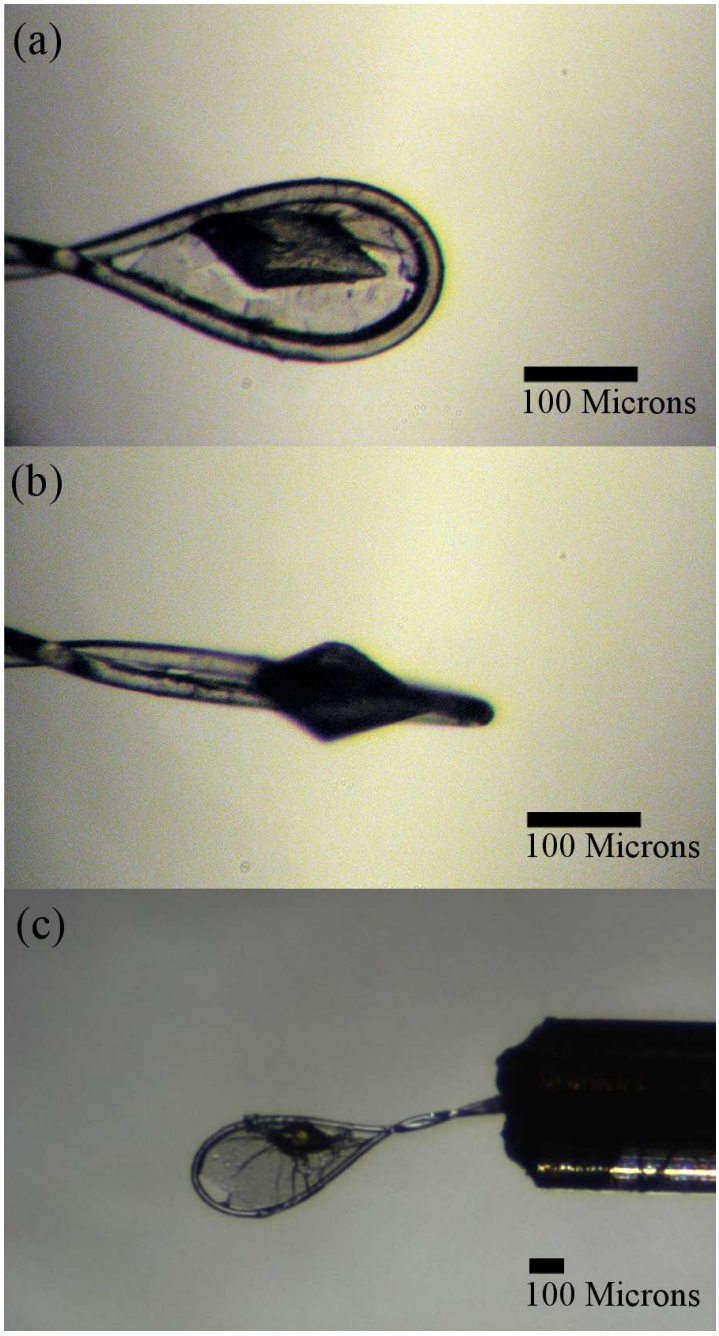


Figure 3.6: Cryocooled thaumatin crystals wrapped in graphene on a Hampton cryoloop. This figure shows two samples of crystals wrapped within graphene, supported in a cryoloop using the graphene as a scaffold. One sample wrapped in three layers of graphene is shown at orthogonal angles in (a) and (b), with features of the crystal prominently above the plane of the cryoloop (a), as well as the crystal placement within the cryoloop (b). (c) shows the crinkling and creasing of five-layer graphene along the contours of another crystal. All three images show the graphene completely covering the thaumatin crystals.

(2) Samples were air-dried for ten minutes at room temperature they were flash cooled and diffraction data were collected (shown in Figure 3.7(e)).

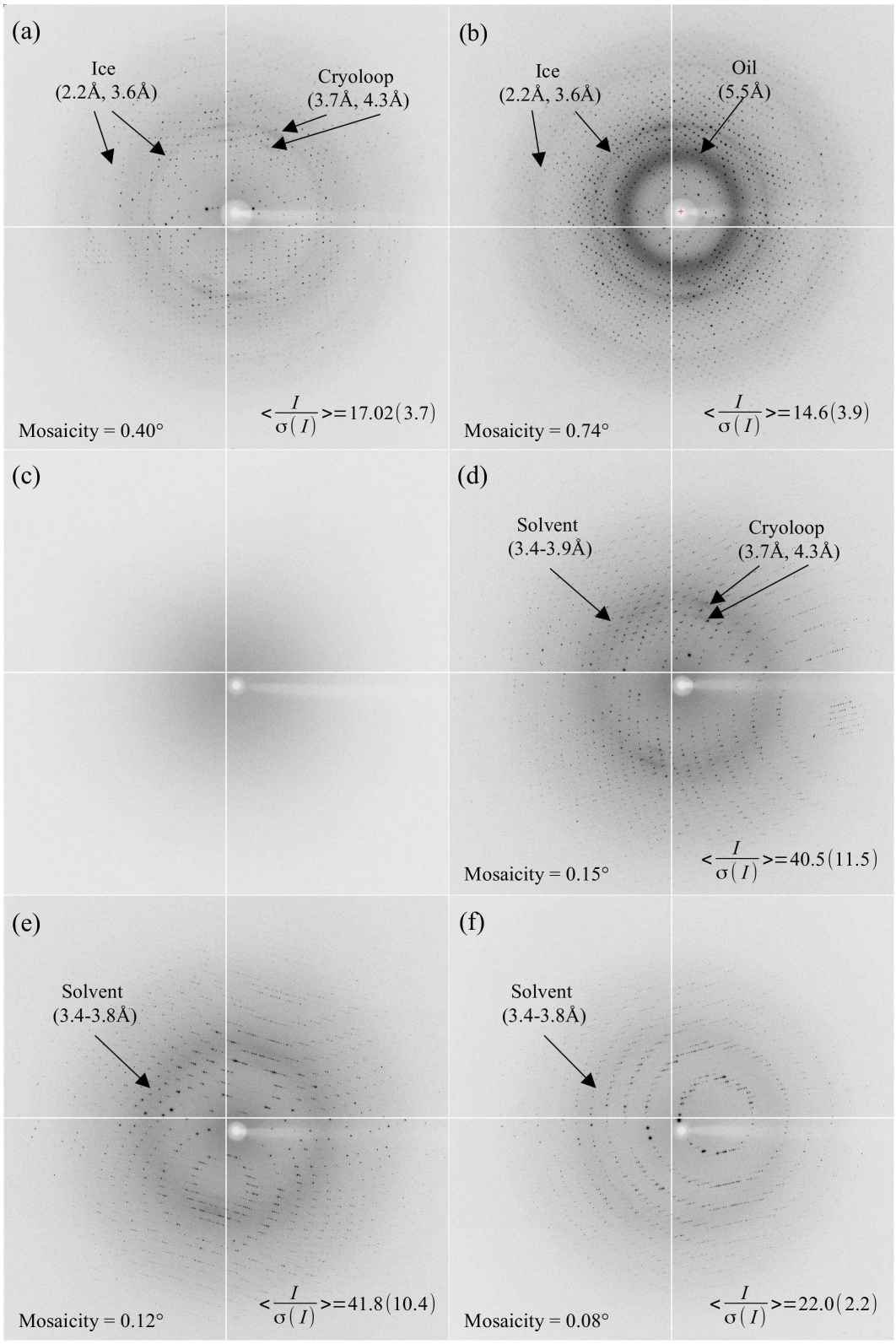
(3) Diffraction data were obtained at room temperature, from samples that were allowed to remain in air at room temperature for approximately 5 minutes before data collection (shown in Figure 3.7(f)); the delay is due to the time needed begin the diffraction experiment.

Crystallographic data collection and processing

Crystallographic x-ray diffraction data were collected at CHESS at the F1 beamline station (λ = 0.9179 Å, E= 13.508 keV) using a 100 μm monochromatic x-ray beam from a 24-pole wiggler. For data collection, an Area Detector Systems Corporation (ADSC) Quantum 270 (Q270) detector was placed such that the face was perpendicular to the beam, at a distance of 200 mm from the sample, corresponding to a largest inscribed circle of resolution of 1.4 Å. For each crystal sample, complete datasets were imaged at 100 K (or at ~300 K for the room temperature graphene-wrapped samples) with a 1° oscillation step-size during each one second exposure, resulting in a 90° rotation overall. The software program *HKL-2000* was used to index, refine, integrate and scale each 90° data set [162]. Parameters including unit cell size, resolution, mosaicity, redundancy, completeness, $\langle I \rangle / \langle \sigma(I) \rangle$ and R_{sym} were evaluated for every data set during the scaling process and compared between data sets.

Samples consisting of an empty cryoloop and ten layers of graphene suspended across the cryoloop were also examined in order to determine the background scattering baseline arising from sources outside the crystal (air paths,

Figure 3.7: Diffraction patterns from five different methods including conventional and graphene-wrapped thaumatin crystals. (a) Diffraction from a crystal flash cooled immediately after equilibration in crystal preparation. The faint ice rings (arrows) are due to the small amount of cryoprotectant used in the solvent. (b) Diffraction from a sample coated in oil to prevent dehydration and flash cooled in liquid nitrogen before imaging. Again, there are ice rings from the small glycerol content of the solvent. (c) Diffraction from a sample of graphene stretched across a cryoloop without crystals or solvent. The resultant scatter, which includes that from graphene, is indistinguishable from background scatter (data not shown). (d) Diffraction from a crystal wrapped in five-layer graphene and then directly flash cooled before diffraction. Note the absence of the ice rings, indicating that the amount of solution surrounding the crystal is small enough to eliminate the growth of ice crystals. (e) Diffraction from a crystal left at room temperature for ten minutes after being wrapped in five-layer graphene, before being flash cooled and exposed to x-rays. (f) Diffraction from a graphene-wrapped crystal diffracting at room temperature without any cooling. Figure on next page.


(a)
Ice
(2.2Å, 3.6Å)
Cryoloop
(3.7Å, 4.3Å)
Mosaicity = 0.40°
$<\frac{I}{\sigma(I)}>=17.02(3.7)$
(b)
Ice
(2.2Å, 3.6Å)
Oil
(5.5Å)
Mosaicity = 0.74°
$<\frac{I}{\sigma(I)}>=14.6(3.9)$
(c)
(d)
Solvent
(3.4-3.9Å)
Cryoloop
(3.7Å, 4.3Å)
Mosaicity = 0.15°
$<\frac{I}{\sigma(I)}>=40.5(11.5)$
(e)
Solvent
(3.4-3.8Å)
Mosaicity = 0.12°
$<\frac{I}{\sigma(I)}>=41.8(10.4)$
(f)
Solvent
(3.4-3.8Å)
Mosaicity = 0.08°
$<\frac{I}{\sigma(I)}>=22.0(2.2)$

collimator scatter, *etc.*). No attempt was made to eliminate air scatter arising from the air path from the collimator to the beamstop.

3.5.2 Results

Tetragonal thaumatin crystals were determined to have a space group of $P4_12_12$ and with solvent content of 55-60% (*v/v*) calculated using the Matthews coefficient and CCP4 programming suite [134]. Figure 3.7 shows example diffraction images, at the same contrast level, taken samples obtained using conventional methods of flash cooling as well as incorporating graphene. The diffraction images of crystals without graphene-wrapping are shown in Figure 3.7(a) and Figure 3.7(b). The average mosaicity (of at least three crystals in each case) of flash cooled crystals (*e.g.*, as in Figure 3.7(a)) was 0.36° and $\langle I \rangle / \langle \sigma(I) \rangle$ was 21.4, whereas oil protected crystals (*e.g.*, in Figure 3.7(b)) had averages of 0.67° and 18.0, respectively. The empty graphene cryoloop image (Figure 3.7(c)) showed diffraction indistinguishable from background images with no loop and graphene mounted in the beam path (not shown). The scatter that is seen is primarily from air in the beam path.

Figures 3.7(d-f) show representative diffraction images from crystals wrapped in graphene in three different conditions. Crystalline ice formation would result in ice rings corresponding to a spacing of around 3.65 Å: no such pattern was observed for any of the graphene-wrapped samples. There is also minimal diffuse scatter in all images from the solution, both from around the crystals or in the inner solvent channels.

Figure 3.7(d) shows the diffraction image of a crystal which was flash cooled

directly after being wrapped in five-layer graphene, *i.e.*, five layers on either side of the crystal. Diffraction arcs from the nylon cryoloop are seen at spacings corresponding to ~3.7 Å and 4.3 Å. The crystal diffracts to 1.5 Å at the edge of the image, and to ~1.4 Å at the corners. If the detector were moved closer to the crystal, we believe the crystal would have diffracted even further. A solvent ring at 3.4 Å to 3.9 Å is believed to arise primarily from solvent within the crystal.

A diffraction pattern from a graphene-wrapped thaumatin crystal which had been allowed to stand in air for ten minutes prior to flash cryocooling is seen in Figure 3.7(e). The diffraction is very similar to the case of immediate cryocooling (*e.g.*, Figure 3.7(d)). There are no sections of cryoloop present in the diffraction images, and the background scattering from solution in/around the crystal is even fainter. The image diffracts to the edge of the detector, which corresponds to 1.5 Å.

Lastly, Figure 3.7(f) shows a diffraction image of a graphene-wrapped crystal that was not cryocooled and for which diffraction data were collected at room temperature. There is reduced background scatter, specifically from the resolution where water rings typically appear. However, the crystal diffracts to approximately 1.6 Å and does not diffract to the edges of the detector. This is believed to be due to the effect of radiation damage at room temperature.

Table 1 shows the diffraction data from representative data sets in each of the five experimental conditions. The second and third columns give results for comparison from thaumatin crystals not wrapped in graphene, but treated using the more conventional methods of flash-cooling with cryoprotectants without and with oil coating, respectively. The fourth column shows results from a graphene-wrapped crystal cooled directly after mounting within the graphene

in the cryoloop. The fifth column shows data from a graphene-wrapped crystal that was left to air-dry after mounting in graphene in a cryoloop for ten minutes before flash cooling. The final column shows data from a graphene-wrapped crystal which was left at room temperature for the entire exposure set.

Method	**Flash cooled**	**Flash cooled in oil**	**Graphene wrapped, flash cooled**	**Graphene wrapped, delay, flash cooled**	**Graphene wrapped, RT**
Space Group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit-cell param. (Å)	a, b = 58.25, c = 150.40	a, b = 58.62, c = 150.85	a, b = 57.80, c = 150.22	a, b = 57.80, c = 150.30	a, b = 58.70, c = 151.60
Resolution range (Å)	50.0-1.60 (1.63-1.60)	50.0-1.60 (1.63-1.60)	50.0-1.60 (1.63-1.60)	50.0-1.60 (1.63-1.60)	50.0-1.60 (1.63-1.60)
No. of reflections	33845(2034)	28021(1431)	34506(1675)	34650(1679)	35967(1769)
Redundancy	6.5(5.8)	6.6(4.7)	6.9(5.3)	7.1(6.7)	7.1(6.6)
Completeness (%)	95.1(100.0)	98.0 (80.6)	99.8(100.0)	99.9(99.9)	99.8(99.8)
Rsym	8.6(33.3)	14.7(15.2)	7.3(30.4)	6.6(20.5)	14.2(87.2)
$\langle I \rangle / \langle \sigma(I) \rangle$	17.0(3.7)	14.6(3.9)	40.5(11.5)	41.8(10.4)	22.0(2.2)
Mosaicity	0.40	0.74	0.15	0.12	0.08

Table 3.1: A comparison of complete thaumatin diffraction data and statistics from three different methods of graphene-wrapped crystal mounting, along with two standard graphene-free techniques. Values in parenthesis are of the highest resolution shell. For room temperature (RT) graphene-wrapped crystals, four data sets were analyzed. In the remaining four methods, three data sets each were analyzed.

3.5.3 Discussion

Five layers of graphene wrapped around a thaumatin crystal produce background scatter indistinguishable from the background scatter in the absence of

graphene. This agrees with expectations from a material only ten carbon atoms thick. The lifetimes of the graphene-wrapped cooled and room temperature crystals were equivalent to those of crystals obtained using existing methods [141, 147, 163]. There was no visual evidence of any effect from the physical contact of the graphene and the crystals.

Graphene-wrapped crystals give high quality diffraction datasets without the need for external hydration

Complete datasets were acquired using the method of wrapping graphene around thaumatin crystals. This eliminates need for external hydration methods involving enclosing oils or capillaries. The results from graphene wrapping are comparable to other achieved using other methods of background reduction, such as the high-pressure cryocooling with capillary shielding method [141, 164]. The graphene-wrapping technique gives reproducible results for thaumatin crystals. Graphene-wrapped crystals that were cryocooled diffracted to high resolution with mosaicities that were comparable to or better than those achieved with standard cryocooling procedures. The ease with which ice rings were eliminated is probably a consequence of the reduction in solvent external to the crystal. The absence of external coating material led to a reduction in background scatter and thus to an increase in the signal-to-noise ratio.

Graphene extends the time-dependent window for flash cooling crystals during sample mounting and shows extension to room temperature studies

Our study showed that a crystal could be handled in air for over ten minutes after being wrapped in graphene but before being cryocooled, yet still remain hydrated well enough to yield superb diffraction results. These crystals showed no difference in diffraction quality compared to the samples immediately cryocooled in our study. This suggests that the graphene provides an environmental seal, thereby slowing evaporation and keeping the crystals hydrated for some period of time. This is in fact confirmed in studies by Sui *et al.* [78].

In unpublished studies, we observed a graphene-wrapped crystal to diffract to over 2 Å resolution after remaining at room temperature for approximately 18 months without any additional hydration protection other than the graphene. During analysis, it was noted that the unit cell had shifted, and the proteins had repacked inside the unit cell, forming additional molecular contacts. Much more study is needed to fully evaluate what process is occurring in this observation. Recent studies on the permeability of water through cracks in layers of graphene [165] suggests there is much need to produce single, perfect sheets of graphene in order to maintain the gas-tight barrier.

However, our results show that crystals can diffract successfully at room temperature within graphene for at least ten minutes after preparation. Incidentally, the presence of graphene in place of surrounding materials with higher thermal mass, such as excess solvent or coating oil, facilitates a more successful cooling of the sample [141]. We believe this could lead to better results when cryocooling crystals.

The effects of quality and size suitable for crystallography

Electron microscope [150, 159] and helium studies [151] prove that graphene can be vacuum tight, at least on nanometer-to-micron length scales. It is presently difficult to obtain larger sheets of monodomain graphene. However, graphene is the subject of intense study by many groups around the world and the quality of the sheets being produced is constantly improving. High quality sheets larger than a few tens of microns across are now routinely made in many laboratories. It is safe to predict that larger sheets will become available; in fact some groups have produced some rare millimeter-size grains [78, 161]. In the meantime, this study shows that presently available multidomain multilayers are already good enough to mount crystals on the order of 100 μm across. We are confident that the same procedures will work with larger crystals. This has also been confirmed with studies of graphene-wrapped crystals imaged in vacuum [153].

3.5.4 Conclusion

In this study, we showed that the background scatter is greatly decreased by the addition of a graphene multilayer in place of external hydration methods. We have demonstrated the first incorporation of graphene in macromolecular crystallography, in order to increase the signal-to-noise ratio of diffraction patterns from protein crystals. Although no attempt was made to eliminate air scatter in the path length from the collimator to the main beam stop in this proof-of-principle study, one can readily extrapolate how this may be done from previous sections. These proof-of-principle experiments open many obvious possibilities

for SMX, ranging from crystal supports to free standing-windows.

In principle, many additional sources of background could have been eliminated in this study to approach a background of zero. From operating in vacuum to removing the cryoloop, all sources of scatter external to the crystal and the graphene-wrap can be eliminated with clever engineering, however this was not required to show the enormous benefit of merely using graphene.

3.6 Summary

Only through careful elimination of any background scattering source external to the crystals, can the signal from increasingly small microcrystals be distinguished from the noise. What remains to be done is a careful tailoring of crystallography beamlines, where every piece of the pipeline is optimized using vacuum whenever possible and the thinest window and mounting materials necessary to keep the crystals hydrated. Although this is relevant to any light source, the need for this kind of beamline revolution is great at storage ring sources, where SMX has the capacity for strong growth.

In the next chapter, we will explore two growth and delivery devices which aim to provide quality microcrystal diffraction while minimizing their contribution to background scatter.

CHAPTER 4

PROTEIN CRYSTAL GROWTH AND DELIVERY DEVICES AT CHESS

4.1 Introduction

The serial microcrystallography experiment involves a large number of microcrystals, where the complete data set is produced from merging many single crystal data frames. Serial microcrystallography would benefit greatly from optimized systems that produce and deliver large numbers of homogeneously-sized crystals on demand directly to the beamline. The ideal device should be designed for high reproducibility and crystal segregation to ensure single crystal exposures, all the while minimizing material which would contribute background scatter. While there are many options for crystal growth and delivery to the beamline (some examples were mentioned in Sections 2.2 and 2.3), in this chapter we relate two methods being studied at CHESS with the aim of facilitating serial crystallography with small crystals.

Neither method as described is suited to handle micron-sized crystals; rather, the crystals involved are in the 5 to 50 μm range. Many protein systems provide crystals in this size range. Unfrozen crystals of this size are typically too small to each provide complete data sets, and thereby require data collection from multiple crystals. Thus, methods that facilitate the acquisition of these complete data sets would be useful. Further, the proof-of-principle methods described in this chapter leave significant room for future improvement with respect to speed, ease of crystal production and handling, and perhaps, reduction in crystal size.

The first method involves an emulsion-based device has been developed by the Fraden Lab from Brandeis University, in which nanoliter-sized droplets of protein solution are encapsulated in oil and stabilized by surfactant to produce one crystal per drop [79]. The collaboration described below explored the use of this technology to, ideally, grow a large number of small crystals in a microfluidic device that could also then be used to deliver the crystals one at a time into the x-ray beam without need to remove the crystals from the device. In the proof of principle experiment described here, crystals are grown in microdroplets within a microfluidic device. The droplets containing identically sized crystals were then transferred to a x-ray semi-transparent microfluidic chip, where diffraction data was measured at room temperature, one crystal at a time, *via* addressable locations within the chip. As proof of concept, a complete data set was obtained by merging single diffraction frames taken from different unoriented glucose isomerase crystals resolved to 2.1 Å at CHESS. The added benefit of precise microcontrol over the crystallization conditions to ensure consistent crystal sizing is a very attractive option for serial microcrystallography.

In the second method, a viscous jet injector is introduced which was specifically engineered to serially deliver microcrystals to the beamline while maintaining extremely low flow rates and minimizing external material [8]. In comparison to the early liquid jet injectors, the low flow rates reduce the chance of missing crystals as they transverse the beam and reduce total sample consumption. As it was designed to operate in-vacuum at XFELs, no additional external material is needed to deliver the crystals except for a low-scattering, hydrating gel column and a stabilizing air jacket.

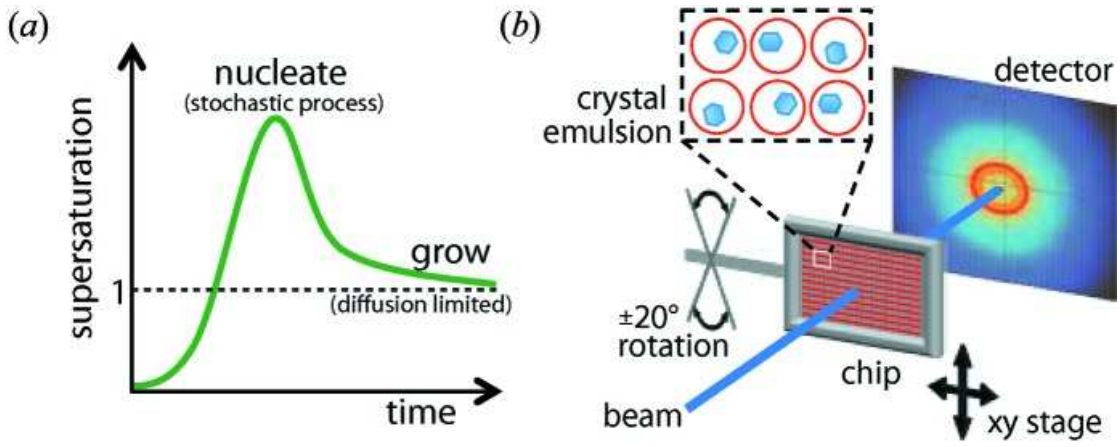


Figure 4.1: (a) An optimal crystallization trajectory increases supersaturation until just one crystal nucleates, then decreases supersaturation to prevent further nucleation while maintaining sufficient supersaturation to promote crystal growth. (b) Emulsion droplets with monodisperse crystals were stored in an x-ray semi-transparent microfluidic device. Sequentially collected diffraction frames from multiple individual crystals were merged to solve the protein structure. The chip could be translated in the x and y directions and rotated 20°. Reproduced from Heymann *et al.* [79].

4.2 Microfluidic chip: Brandeis Collaboration

The contents of this section have been published in IUCrJ by Michael Heymann, Achini Opthalage, Sathish Akella, Jennifer Wierman, Marian Szebenyi, Sol Gruner, and Seth Fraden [79].

The ideal crystallization procedure to produce protein crystals, illustrated in Figure 4.1(a), consists of slowly increasing the supersaturation of a protein solution until the moment that a single crystal is nucleated. Once the first nucleation event occurs, supersaturation is reduced enough to prevent further nucleation, while maintaining sufficient supersaturation to grow the crystal. Ideally, the growth conditions should be slow enough to allow for annealing of defects, and the procedure must be capable of producing crystals in large numbers and of identical size. Additionally, the technology to produce crystals must be simple

and inexpensive if serial microcrystallography is to be adopted by the structural biology community.

The challenge is to design such a method. Most crystallization methods (mentioned in Section 2.2) produce a series of kinetic supersaturation profiles that rise and fall as illustrated in Figure 4.1(a). However, both the time at which the supersaturation maximum occurs and the value of the supersaturation maximum are independent of the nucleation event. Nucleation is a random process at constant supersaturation, rendering it impossible to know *a priori* when to decrease supersaturation, which should coincide with the first nucleation event. One way to generate the ideal supersaturation profile would be to monitor the supersaturated solution with a technique, such as second harmonic generation (SHG) microscopy [166, 167], that is sensitive to the formation of small crystals and then, once the first crystal is detected, lower the supersaturation. However, this scheme will be cumbersome to implement in the high-throughput case of processing hundreds to thousands of samples. An alternative method is desired.

Microfluidically produced, monodisperse, emulsions have previously been used to produce drops of supersaturated protein solution in which each drop nucleates a single crystal [168, 169, 170, 171]. This situation could be ideal for serial microcrystallography for a number of reasons. Since only one crystal nucleates per drop, all the supersaturated protein in solution is delivered to a single crystal, making that crystal as large as possible. Microfluidic precision allows preparation of emulsion droplets with variations in size of a few percent only, even at high flow rates [172]. Furthermore, because of the small length scales in microfluidics, convection is suppressed and flows are laminar. Taken together,

these factors mean that processing proteins using microfluidics leads to crystals of a uniform size that are grown under identical conditions, which has the effect of creating crystals that have similar characteristics, such as unit cell and degree of disorder. Having identical crystals facilitates merging of diffraction data sets taken from different crystals.

In the microfluidic device described here, drops containing protein are produced and guided to 8000 storage sites on chip through surface tension [173, 174]. Next, supersaturation is increased to induce crystallization in such a way as to produce one crystal per drop. Finally, diffraction is collected from individual crystals within a thin-microfluidic chip and merged into data sets in order to solve the protein structure (Figure 4.1(b)).

Producing and diffracting from crystals in the same device eliminates the laborious and potentially damaging steps of looping and extracting the crystal from the mother liquor. Various microfluidic crystallization platforms compatible with *in situ* diffraction have been developed [175, 176, 177, 178]. However, these devices incorporated valves in the chip [175, 178], thus rendering them expensive to manufacture and difficult to operate. Other technologies are low throughput [177], or need a second round of scale-up to larger capillaries [176] to produce crystals large enough to collect diffraction data.

4.2.1 One crystal per drop through compartmentalization

The production of one crystal per drop is a result of a competition between two processes, nucleation and growth, in a confined volume. Both processes require supersaturation and therefore both nucleation and growth are non-equilibrium

processes. When the first nucleus forms inside the drop, it decreases the supersaturation in the surrounding protein solution as the crystal grows. If the rate of nucleation is low enough, then the growing crystal will consume enough of the protein in solution to decrease the supersaturation to the point where another nucleation event is improbable. Further nucleation is prevented if the time for a protein to diffuse across a drop is less than the time to nucleate a crystal [171]. Thus combining a small drop volume with the physics of nucleation and growth, generates negative feedback that acts to autonomously create the ideal dynamical supersaturation profile that produces one crystal per drop. Instead of having the negative feedback imposed externally, as in the cumbersome SHG microscopy scheme discussed previously, here the negative feedback is engineered into each drop; no external intervention is required. All the engineering goes into identifying the correct combination of diffusive flux, nucleation rate and drop volume for the emulsions. A theoretical argument and computer simulations describing the process leading to one crystal per drop in small volumes were performed by Heymann *et al.* and reported in [79].

4.2.2 Crystal emulsions

To yield identical crystals in sufficient quality and quantity for serial crystallography, a two step method is used. First the appropriate drop volume to nucleate one crystal per drop consistently is identified. For this emulsions were created in a batch process that yielded a polydisperse size distribution, ranging from a few microns to a few hundreds of microns in diameter (Figure 4.2(a-c)). Such a polydisperse emulsion allowed identification of the critical diameter for proper crystal nucleation in a single screening experiment. Next, microfluidics (Figure

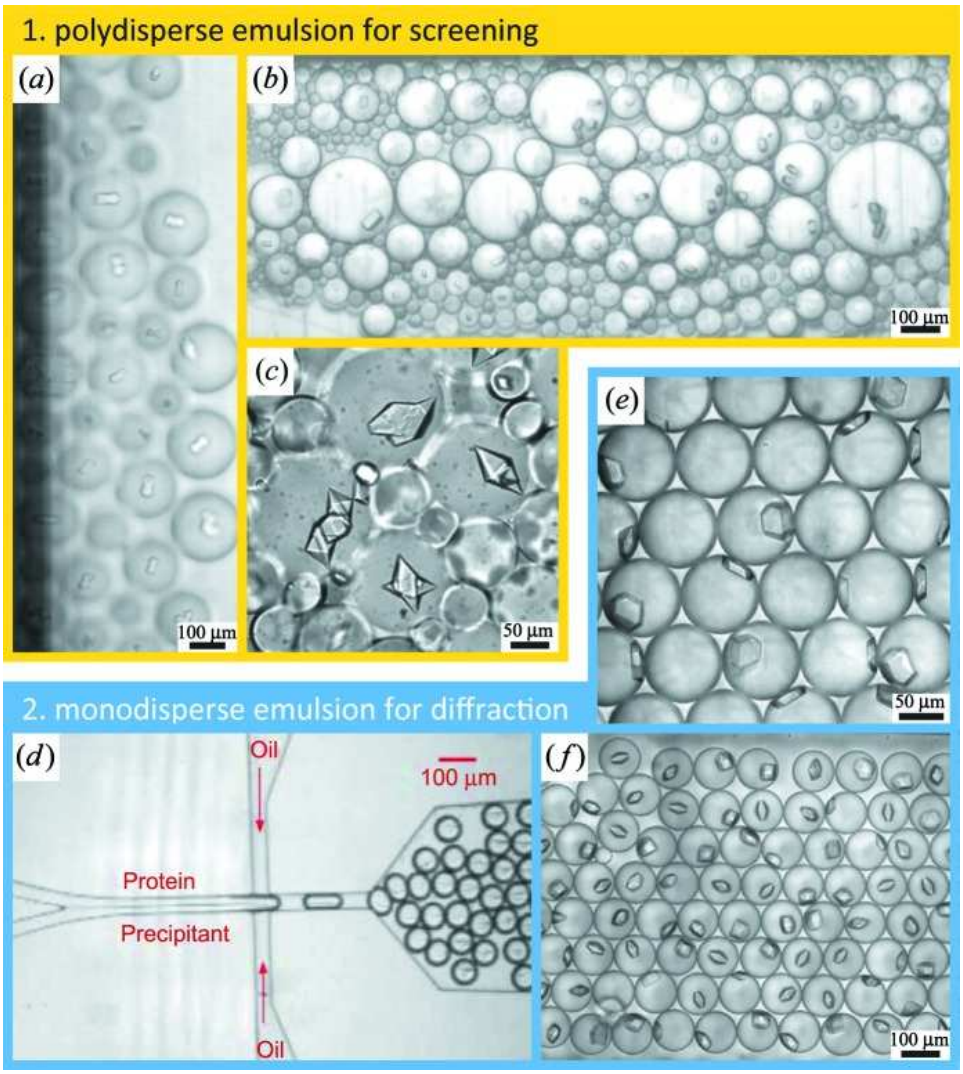


Figure 4.2: Protein crystallization in emulsion droplets stabilized by surfactant. Ideal drop sizes were first identified using polydisperse emulsion droplets. Monodisperse emulsions were used to produce identical crystals for diffraction experiments. Polydisperse emulsions are shown of (a) D1D2 heterodimer from human spliceosomal snRNP particle, (b) concanavalin A and (c) trypsin. (d) Protein and precipitant solutions were introduced in a co-flow geometry under laminar flow conditions that prevent mixing upstream of the nozzle where both solutions became encapsulated into emulsion droplets. Monodisperse emulsions are shown of (e) glucose isomerase and (f) lysozyme crystals. See Heymann *et al.* [79] for details. (a-c) are reproduced from Heymann *et al.* [79], (d-f) are reproduced from Akella *et al.* [179].

4.2(d)) were used to produce monodisperse emulsion droplets (Figure 4.2(e-f)) and grow identical crystals within the serial x-ray diffraction chip, as described in Section 4.2.4. For the purposes of this dissertation, a proof-of-principle full experimental sequence will only be reported for glucose isomerase, whereby crystals were grown in the serial diffraction chip, x-ray data were acquired, and the structure was solved.

Crystals were grown in emulsion droplets stabilized against coalescence with a 2% *(v/v)* solution of PFPE-PEG-PFPE surfactant E2K0660 in HFE7500 fluorinated oil (from 3M). The surfactant was synthesized as previously described [180]. A fluorinated oil and surfactant were chosen to minimize interactions with biological molecules. Fluorocarbon and hydrocarbon oils do not mix, nor are they miscible with water. In particular, the PFPE-PEG-PFPE surfactant in HFE7500 oil system has been shown to have excellent bio-compatibility [180, 181]. To confirm that it is compatible with protein crystallization, it was tested with five crystallization model proteins: Lysozyme, glucose isomerase, trypsin, concanavalin A, and D1D2, a sub-complex from the human snRNP spliceosome core particle (Figure 4.3). All five model proteins have previously been crystallized by vapor diffusion and a structure derived from x-ray crystallography deposited in the PDB.

In traditional vapor diffusion (VD) (discussed in Section 2.2), recipes have been optimized to nucleate only a few crystals per microliter. In contrast, emulsion droplets have volumes of a few pico- to nanoliters each. As the probability of nucleating a crystal is proportional to the sample volume, nucleation rates were increased by at least two orders of magnitude. Vapor phase and microbatch crystallization trials were produced around literature recipes and

optimized to nucleate crystal showers of appropriate density. When attempting to crystallize a novel protein target through screening crystallization conditions, such crystal showers are usually considered a first hit and the conditions are later refined extensively to grow the largest possible crystal. If instead the method presented here is used, the polydisperse emulsion screen can identify directly the conditions giving crystallites and crystal showers. This would eliminate the reverse engineering step of converting an optimized vapor phase recipe back to a recipe that grows crystal showers.

Polydisperse emulsions were then prepared by mixing 2 μL protein solution with 2 μL precipitant in a 150 μL PCR test-tube. Immediately after mixing, 30 μL 2% *(v/v)* solution of PFPE-PEG-PFPE surfactant (E2K0660) was added to HFE7500 fluorinated oil. Polydisperse emulsions were formed by gently agitating the vial by hand until droplets became too small to be resolved by eye. This procedure typically gave droplets ranging from a few microns to a few hundreds of microns in diameter (Figure 4.2(a-c)). Aqueous droplets were less dense then the immersing fluorinated oil, so droplets rose to the top of the vial within a minute. The emulsion was then loaded into rectangular glass capillaries (VitroTubes from VitroCom, Mountain Lakes, NJ, USA) and sealed with 5 Minute Epoxy to prevent evaporation. Crystallization was monitored over the course of a week. All compounds and proteins from commercial sources were used as received without further purification. To first order, preparing a polydisperse emulsion takes about the same time as preparing a hanging or sitting drop VD condition. Both require three pipetting steps and a final lidding or shaking operation.

Details of the specific crystallization conditions for lysozyme, glucose iso-

merase, trypsin, concanavalin A and D1D2 can be found in Heymann *et al.* [79]. All globular proteins, concanavalin, glucose isomerase and trypsin, crystallized readily in VD, microbatch, and the emulsion system. The heterodimer D1D2 formed crystals in vapor phase and the emulsion system only. In microbatch a thick protein skin grew at the droplet interphase potentially depleting all the protein from the drop. Thus, the PFPE-PEG-PFPE surfactant system is well suited to protect protein from absorbing to the fluoro oil-water interface and to stabilizing emulsions, making it ideal for crystallization trials. Future work should investigate compatibility of the surfactant with other proteins.

Finally, to yield identical crystals in sufficient quantity for serial crystallography, monodisperse emulsion droplets were produced with microfluidics in a co-flow geometry designed such that the protein solution and buffer do not mix in the laminar flow upstream of the appropriately sized dropmaker (Figure 4.2(d)). Protein and precipitant streams were pumped at equal flow rates of 300 μL per hour to co-encapsulate both in a 1:1 mixture. Upon droplet formation, mixing inside each droplet proceeds within less then a second due to recirculating flow that arises from shearing interactions of the fluid inside the drops with the stationary wall [182]. These monodisperse emulsion droplets were then injected into and incubated in the diffraction chip to grow crystals for the x-ray diffraction experiments.

Emulsion droplets can shrink by permeation of water vapor from the drops into the oil and through the thin, polymer-based chip and can be observed at a few percent per hour. To monitor crystallization, the emulsion droplets were either sealed into rectangular glass capillaries, or the chip was immersed in an oil bath to prevent water and oil evaporation.

4.2.3 X-ray semi-transparent chip fabrication

A detailed description of the fabrication of the microfluidic chips is given by Guha *et al.* [178] and Heymann *et al.* [79]. As an overview, microfluidic chips were created from curing thin films of PDMS molded on a photoresist master. The PDMS films were then sealed, which is colloquially referred to in the thermoplastic industry as lidding, by bonding either 25 μm COC or 8 μm Kapton foil to both sides of the thin PDMS slab containing the channels (Figure 4.3). Upon assembly the chip was surface treated with a fluorophilic coating to prevent protein interaction with the channel surface and to eliminate gas bubbles within the chip. The chip was then incubated at 363 K for at least 12 hours to evaporate the solvent away and accelerate chemical cross-linking between fluoropolymer and chip surface.

4.2.4 In-situ diffraction

The x-ray semi-transparent chip was mounted into a custom acrylic frame to collect diffraction data (Figure 4.4). Each frame was held together by screws to lock the chip into position and to minimize flow induced inside the chip from mechanically bending the thin foil chip. The acrylic frame is fashioned with ports allowing access into the foil-chip. To mount the frame-chip assembly in the synchrotron a stainless steel adapter was machined that a frame could be mounted onto using two screws (Figure 4.4(b)).

For the proof of principle experiment an x-ray semi-transparent chip was fabricated with the dropspot geometry [174] that can hold up to 8000 emulsion droplets in cavities with 150 μm diameter each (Figure 4.4(b-c)). The fluorinated

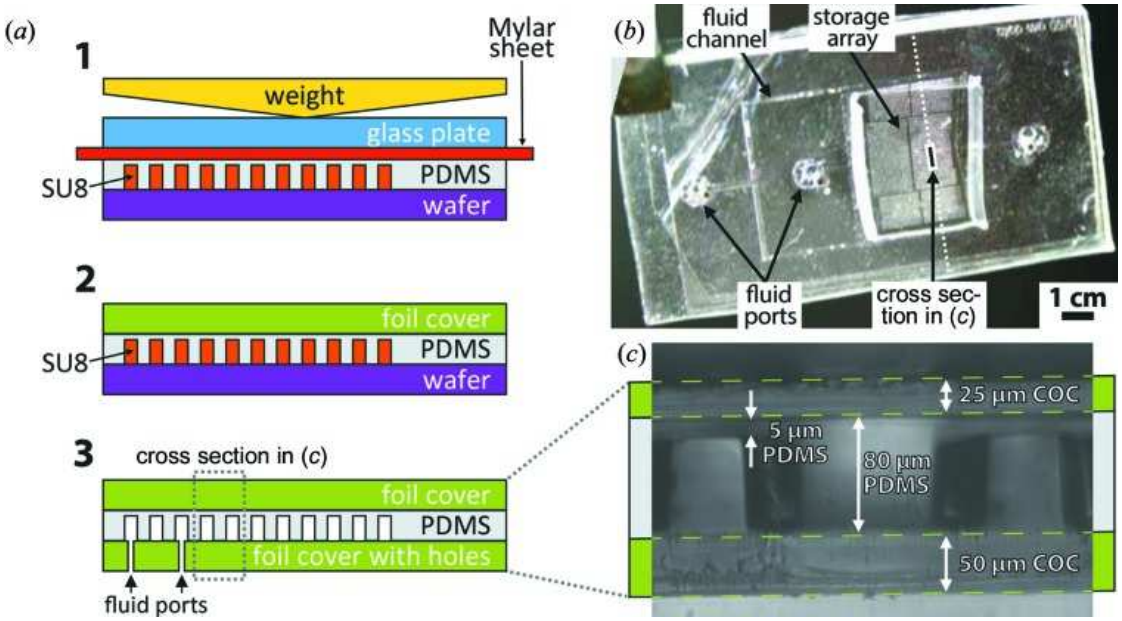


Figure 4.3: Brandeis microfluidic chip fabrication. (a) **1** PDMS resin was squeezed into a thin layer onto the photoresist (SU8) master. **2** After curing, a foil cover was bonded onto the featured PDMS using a silane coupling chemistry [183]. **3** The reinforced PDMS film was peeled off and the chip was lidded using another foil cover. (b) Top view and (c) cross section of a device made from COC foil and PDMS. The cross section in (c) was obtained by cutting the chip across the storage array into two and was imaged by placing the chip edge on onto the microscope stage to magnify the cut. The chip shown here had a 5mm-thick PDMS frame manifold for fluid interfacing where tubing could be directly inserted into the through holes in the PDMS. Reproduced from Heymann *et al.* [79].

oil has a density of 2 g/mL, while the water drops have a density of 1 g/mL. Thus there is a strong tendency for the drops to float to the top of the oil. Surface tension forces arrest droplets in a cavity and prevent them from locating to one side of the chip. A monodisperse ~110 µm diameter emulsion was produced of 30 mg/mL glucose isomerase, 100 mM Ammonium Sulfate, pH 7.3, 20 wt% PEG 10000 MW final concentration in a standard dropmaker (Figure 4.2(d)). Droplets exiting the dropmaker were immediately routed into the x-ray semi-transparent serial crystallography chip by plumbing the dropmaker outlet into the dropspot inlet (Figure 4.4(a)). After the dropspot chip was loaded, the

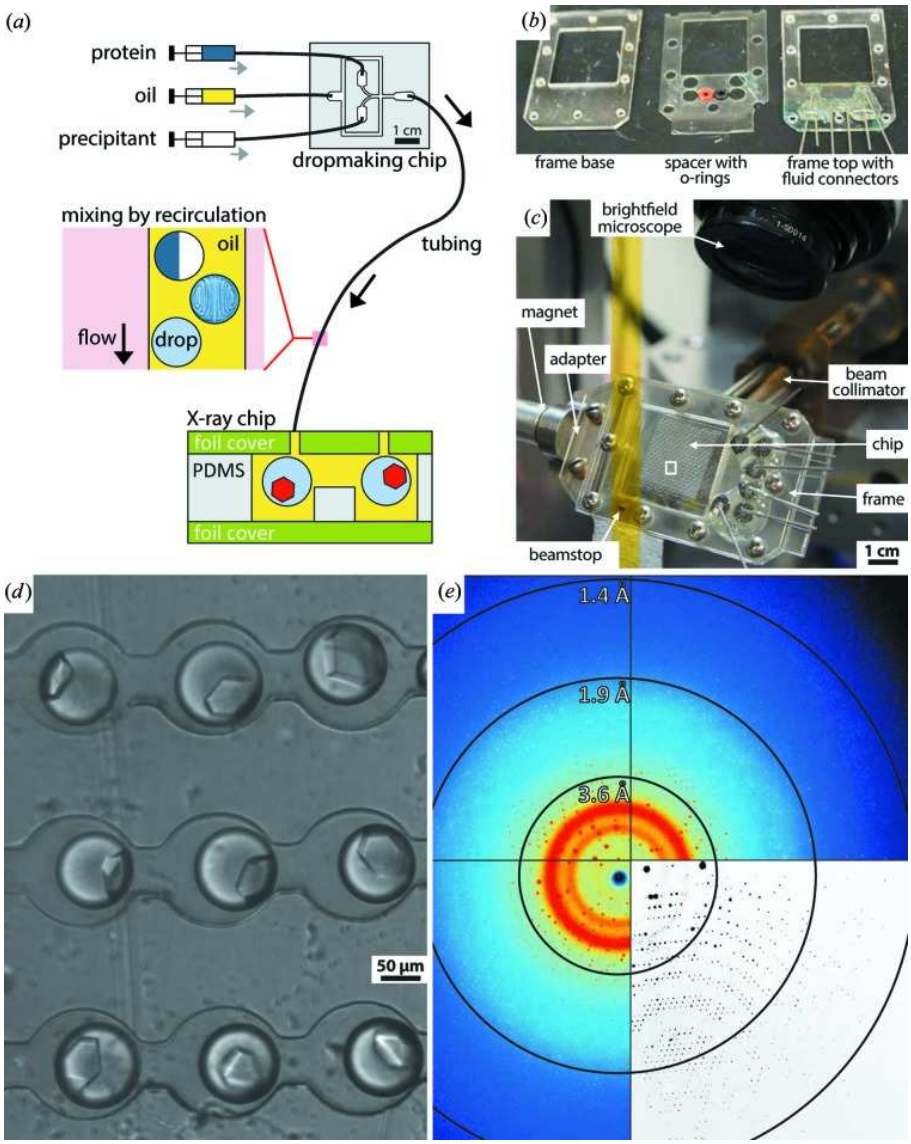


Figure 4.4: (a) Monodisperse emulsions were prepared using a dedicated drop-making chip as illustrated in Figure 4.2(d) and directly routed into the chip for serial crystallography storage. (b) A laser-cut frame held the x-ray semi-transparent chip with ports for access. (c) The x-ray semi-transparent chip mounted on the goniometer inside the Cornell CHESS F1 beamline. (d) Glucose isomerase crystals inside of the microfluidic device. Using a motorized stage, each crystal can be centered in the collimated x-ray beam. The beam is 100µm in diameter. (e) A representative diffraction pattern of a glucose isomerase crystal taken at room temperature from inside the chip. Crystals diffracted to 1.4 Å resolution with a mosaicity as low as 0.04°. The bottom-right quadrant shows the diffraction pattern after background subtraction, using the *ADXV* diffraction pattern visualization tool with subtract background option. Note the high level of background scatter, resulting from scatter from the microfluidic chip. Reproduced from Heymann *et al.* [79].

outlets were plugged except for one inlet where oil could enter the chip using hydrostatic pressure to compensate for oil evaporation from the chip. The chip was incubated at room temperature for three days and monitored for crystallization, before transferring into a water bath to prevent further evaporation. By then, most droplets had shrunken to about 90 μm diameter and more than 90% of them had nucleated a single crystal. Crystals grew to about 50 μm × 40 μm × 30 μm in size at room temperature (~ 298 K).

X-ray diffraction data were collected at CHESS, beamline F1 (λ = 0.9179Å, E = 13.508 keV, x-ray flux = 5.5 × 10^{10} x-rays per second), using a 100 μm monochromatic x-ray beam from a 24-pole wiggler. The chips were mounted at a distance of 200 mm from an ADSC Quantum 270 (Q270) detector, corresponding to a largest inscribed circle of resolution of 1.4 μm. The detector face was oriented perpendicular to the beam. For selected crystals within the chip, data sets were collected at room temperature (~ 298 K). Each recorded data set comprised ten frames, for a total of 10° oscillation. Each image consisted of a 5 second exposure with a 1° oscillation step size; a representative frame is shown in Figure 4.4(e). A total of 1520 images were collected from 152 glucose isomerase crystals in three different dropspot chips.

4.2.5 X-ray structure determination

The software *HKL-2000* was used to index, refine, integrate and scale each 10° data set [162] before merging. Parameters including unit-cell size, chi-squared values, resolution, mosaicity and completeness were evaluated for every partial data set during the indexing and scaling process. From these partial data

sets with 1520 frames total, 262 frames were selected from 72 crystals by rejecting frames with a mosaic spread higher than 0.1° and χ^2 values for x and y (corresponding to discrepancy between observed and predicted spot positions) above 2. Some frames were later rejected because of poor scaling statistics; the final data set included 248 frames.

Glucose isomerase crystals were determined to have a space group of $I222$, and diffracted to an average of 2 Å; an example image is shown in Figure 4.4(e). In some crystals, diffraction extended to 1.4 Å, with a mosaic spread of 0.04°.

Refinement statistics	
Space Group	$I222$
Unit Cell Parameters (Å)	a = 93.94, b = 99.47, c = 102.85
Resolution (Å)	49.7-2.09 (2.15-2.09)
Completeness (%)	93.2 (94)
# of independent reflections	26699 (2075)
Redundancy	8.2 (8.1)
R_{merge}	0.191 (0.686)
$\langle I \rangle / \langle \sigma(I) \rangle$	7.8 (4.1)
Mosaicity (°)	0.03-0.1

Table 4.1: Glucose isomerase crystal refinement statistics from Brandies Microfluidic Chip. Processing results of merging the 248 frames obtained from 72 glucose isomerase crystals. Values in parentheses refer to the highest resolution bin (2.15 - 2.09 Å).

The 248 selected frames were scaled together using *SCALEPACK* (HKL Research, Charlottesville, VA, USA) and merged with *Aimless* [184]. The limiting resolution of 2.09 Å was chosen as that at which $CC_{1/2}$ dropped below 0.5. Statistics are given in Table 4.1. The merged data set covered 93% of reciprocal space, suggesting that preferred orientation of the crystals was not a major problem. The glucose isomerase structure was readily solved by molecular replacement with *MOLREP* [132] using the structure previously determined at 1.90 Å reso-

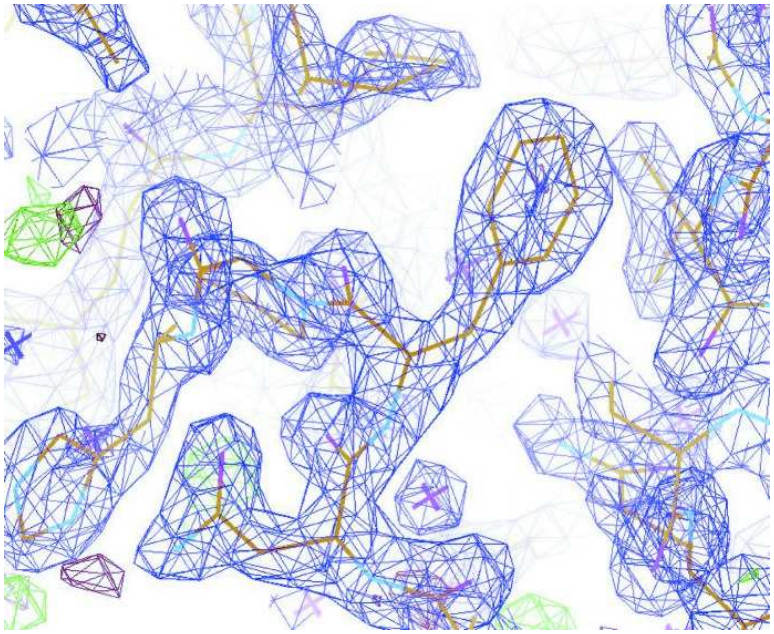


Figure 4.5: Electron density map of glucose isomerase. Part of the final refined structure showing the quality of the electron density map. The $2F_oF_c$ map is shown in purple, contoured at 2σ, while the F_oF_c map is shown in red (negative) and green (positive), contoured at 3σ. Reproduced from Heymann *et al.* [79].

lution (PDB entry 8XIA [185]), with water molecules removed. Prior to refinement, 5% of the reflections were randomly flagged for R_{free} analysis [186].

Structure refinement was carried out through multiple iterations of *REFMAC* [187], refining atomic coordinates and isotropic B factors. $2F_o - F_c$ and $F_o - F_c$ electron density maps were generated after each refinement step, and further refinement was carried out by manual inspection using *Coot* [135]. In the refinement process, two disordered N-terminal residues were removed, as well as a bound sugar molecule present in the model but not in the crystal, and 124 water molecules were added. Final refinement gave R_{cryst} and R_{free} of 0.144 and 0.174, respectively. Complete processing statistics are given in Table 3. Figure 4.5 shows the quality of the final refined structure.

4.2.6 Results and conclusion

Here, a technology is presented that optimizes the kinetics of crystallization, eliminates crystal handling, eliminates cryoprotection and simplifies collection of diffraction data for structural biology. Processing methods for protein crystallization have been developed to follow the ideal kinetic pathway of slowly increasing supersaturation until a single crystal nucleates and then reducing supersaturation so that one crystal grows slowly to allow annealing of defects. Sample volume is not a thermodynamic variable in phase equilibrium, but since crystallization is a non-equilibrium process, volume plays a key role in determining the kinetics of crystallization. A combination of simulation, theory and experiment suggesting that selecting the appropriate droplet diameter, D, guarantees that only one crystal per drop will form when the drop volume, V, is smaller than $\sim (D/J)^{n/(2+n)}$, for a nucleation rate, J, in n dimensions. The critical drop diameter for a particular crystallization condition in a single experiment is determined using a polydisperse emulsion with droplets ranging from a few micrometers to a few hundreds of micrometers in size. These polydisperse emulsions can be made with ease within seconds using only a pipette and a test tube. The probability of crystallization is proportional to drop volume. As the drops are of order 1 nL, which are smaller drops than employed by other methods, the nucleation rates and supersaturation that are seen are higher than usual. It remains to be seen how such high nucleation and growth rates impact crystal quality. Future studies are needed to access the quality of protein crystal structure determination as a function of crystal size, nucleation rate and crystal growth rate to determine the optimal crystal size and crystallization conditions for serial crystallography.

Employing these kinetic processing methods, monodisperse crystals were grown compartmentalized in emulsion droplets, with one crystal per drop. Monodisperse, microfluidically produced drops of supersaturated protein solutions were stored on chip and slowly concentrated as water permeated through the thin-foil chip. One single crystal per drop was nucleated and grown on-chip in identical conditions. The chip for nucleating crystals was thin enough to be x-ray semi-transparent and diffraction patterns were collected from these crystals on-chip at room temperature. The structure of glucose isomerase was solved and refined at 2.09 Å resolution, to an R_{cryst}/R_{free} of 0.144/0.174, using merged diffraction datasets from 72 crystals of about 50 μm × 40 μm × 30 μm in size.

Diffraction from room temperature crystals stored on the chip in which they were nucleated and grown has many advantages over traditional off-chip cryoprotected crystals. On-chip diffraction means the crystals are not removed from their mother liquor, thus avoiding a process that can lead to dehydration and osmotic shock of the crystals and the generation of stress and strain. Room temperature diffraction eliminates the laborious step of cryoprotection and has the additional effect of lowering the mosaicity, as cryoprotection generates stresses due to changing solvent conditions and temperature induced volume changes. This chip can be inexpensively mass produced and is simple to operate without the need of controlling valves.

The long term vision is to create a chip that uses temperature and concentration gradients to discover optimal crystal growth conditions [173] and minimizes background contribution. While the background scattering from this particular chip is relatively high (Figure 4.4), reducing the thickness of window material within the chip will reduce the excess scatter seen in the diffraction

image. Next, crystals would be grown at the optimal conditions to create a stream of tiny crystals that would be serially conveyed to a part of the chip with ultra-thin windows for in-situ diffraction. Possibilities include exploring windows made from materials such as ultrathin silicon nitride [188] or graphene [78, 139]. Ways need to be developed to reduce in-beam volumes of the fluids surrounding crystals.

Similar devices, as mentioned in Section 2.2, are being engineered in other groups. For example, the Perry Group [78] has optimized their microfluidic chips to incorporate graphene as a water barrier and have gathered data at CHESS in recent months. Further collaboration in developing this technology is expected at CHESS. As all of these devices continue to use polymers to add structural integrity to graphene, it remains to be seen just how thin these devices can be engineered to minimize background scatter.

4.3 Viscous jet injector: ASU device

In this section, we explore the use of crystal injectors developed for XFEL SFX for SMX at storage rings. This was the result of a collaboration with John Spence, Uwe Weierstall, Daniel James, Garrett Nelson and Chelsie Conrad.

The gas dynamic virtual nozzle (GDVN), introduced the concept of flowing crystals within a hydrating stream across the XFEL x-ray beam [2, 12, 83, 84]. The GDVN succeeded in delivering crystals to the beam rapidly, yet so rapidly that much of the sample passes between the x-ray pulses; thus many crystals are missed as they flow through the interaction point between the x-ray pulses (see Figure 1.2). As a measure of sample consumption, the actual number of crystals

hit with the GDVN at XFELs are estimated to be less than a percent of the total number of crystals used. This is in part a consequence of the GDVN flow rates of $\sim$ 10 μL per minute, and in part the pulsed nature of the LCLS XFEL (120 Hz). While a more continuous probe, such as a storage ring source, would solve the problem of pulses missing crystals, the GDVN still consumes tens of milligrams of sample before enough data is recorded by a detector for a structure solution. This imposes a demand for large sample volumes. A simple solution would be to slow the flow rate, which reduces the sample consumption, and eases the burden of collecting larger volumes of protein crystal solution. The viscous lipidic cubic phase jet injector, or LCP injector, engineered at Arizona State University (ASU) provides this solution.

With a goal of room temperature serial microcrystallography at CHESS, MacCHESS (Macromolecular Facility at CHESS) acquired the viscous gel injector from Arizona State University in February 2015 for use as a microcrystal delivery system. The injector uses a more viscous suspension for the microcrystals than the GDVN, and slows the flow rate to $\sim$ 0.05 μL to 2 μL per minute, which allows for 10-100 ms exposures and sample consumptions of less than a milligram [8]. The increase in viscosity of the medium extruded required several design changes from the GDVN, including using the gas stream as a stabilizing mechanism, rather than a focusing mechanism.

The in-air injector is made of four stainless steel parts: a nozzle body, an isolated reservoir, a hydraulic stage body and a piston, as seen in Figure 4.6. An additional gas line feed through part can also be added for vacuum compatibility. The piston sits inside the hydraulic stage body, and an external pump provides hydraulic pressure to the backside of the piston. Compressed Teflon spheres

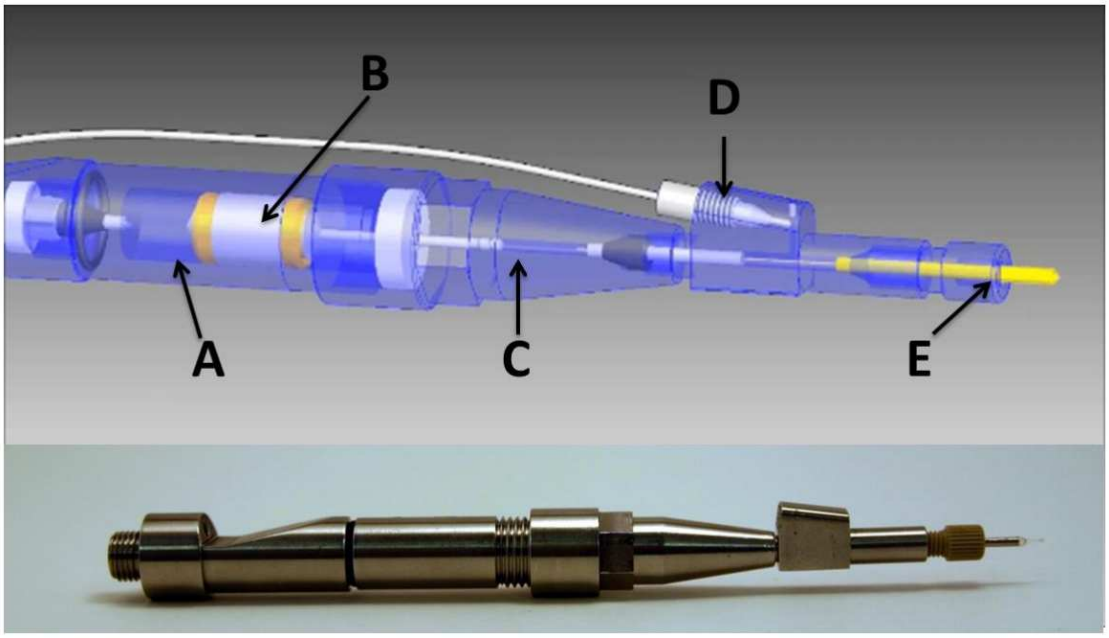


Figure 4.6: A schematic and figure of the viscous jet injector from Arizona State University. (A) A hydraulic stage body which is back pressured from a constant flow rate pump. (B) Pressure is provided by the piston which is sealed from the (C) reservoir by two compressed Teflon spheres (not shown). (D) A gas inlet is routed into the main nozzle body, which is also fitted with replenishable nozzle fittings (E). In addition, a vacuum fitting (attached to hydraulic body at lower left in picture) can be added for in-vacuum compatibility. Reproduced from James [189].

create a seal from the piston and hydraulic stage body to the isolated sample reservoir. The piston provides a pressure amplification to the reservoir equal to the ratio of the front and back cross sectional areas of the piston, $A_f/A_b = 34$, specific to the injector acquired by MacCHESS. This factor not only amplifies the pressure to the reservoir, but also divides the flow rate from the hydraulic pump to the reservoir by the same factor. The sample within the reservoir is then piped through a capillary to the exit of the nozzle, where a gas sheath stabilizes the extruding gel column for x-ray imaging. The only high pressure fitting within the entire jet injector is a ferrule which seals the reservoir to the capillary, which eliminates high pressure constraints on the rest of the injector.

4.3.1 Assembly of the injector system at CHESS

The next section will cover an overview of the set up and operation of the viscous jet injector in the context of operation at CHESS. Please see Daniel James' dissertation entitled *Injection Methods and Instrumentation for Serial X-ray Free Electron Laser Experiments* [189] for an excellent in-depth explanation of the viscous jet injector protocol.

The viscous jet injector is driven by a constant flow rate hydraulic pump (Shimadzu LC-20AD), which can be controlled remotely from outside the experimental hutch. Feed lines connect the pump to the backside of the injector and provide the necessary hydraulic pressure with purified water to drive the piston inside the injector. We connect a Tee and syringe to the hydraulic supply line for manual control of the injector when initial adjustments need to be made, but it is shunted off when the injector is in operation. Any air is purged within the liquid lines to the injector and within the hydraulic piston chamber up to the backside of the piston, as bubbles within the lines add non-linear pressure changes. For the stabilizing air jacket, nitrogen or (preferably) helium gas is supplied to the nozzle head *via* tubing connected from compressed gas cylinders. Regulators control gas supply, as the rate of gas flow will determine the stability of the gel column. The rate depends on the viscosity of the gel itself, and is adjusted per sample loading based on the stability of the column. Both gas and liquid lines are kept at short lengths to reduce resistance within the system. Once the piston is seated in the hydraulic body, the body is installed in the hutch. The remainder of the injector, the reservoir and the nozzle, require more setup, and will be installed later.

In all CHESS experiments with the injector, the body of the injector is

mounted onto the beamline in a vertical position above the x-ray beam with two axis of fine adjustment perpendicular to the beam. The first setups at CHESS beamline F1 held the body of the injector vertically with an optical post clamp to a horizontal beam mounted on a high-precision air-bearing stage for alignment to the x-ray beam. Subsequently, the system was redesigned for setup at beamlines A2 and G3 with a custom-machined support in an effort to stabilize the entire setup and integrate it with a microscopic camera.

Lastly, a microscope camera with adequate illumination on the injector is installed to monitor the gel stream in real-time. This enables visual monitoring during remote control of the injector from outside the experimental hutch. Once the sample-loaded injector is set up inside the hutch, the only need to re-enter is to reset the system after the sample is finished or there is a blockage. Once the hydraulic body has been secured, the detached reservoir of the injector can be loaded with sample, fixed to the hydraulic body and the nozzle assembled. A brief description of this process follows.

Loading the reservoir

Loading the reservoir requires first loading two Teflon spheres into the bore of the reservoir and setting them to the correct position for loading sample. The Teflon balls deform slightly upon insertion, as the diameter of the bore is smaller than the diameter of the Teflon spheres to provide a proper seal. A gapping tool is used to ensure the correct positioning of the Teflon spheres inside the bore of the reservoir. The spheres are positioned at the front of the reservoir to avoid air gaps between the sample and the spheres. A specialized loading needle is threaded into the reservoir and seals the reservoir to a syringe loaded with the

sample. The volume of the reservoir is 20 μL, so we fill same volume of sample into the reservoir from the syringe. Any additional volume will eject the Teflon spheres out the back of the reservoir and break the seal. The reservoir remains attached to the loading needle until the nozzle is ready for full assembly.

Nozzle assembly

One of the problems of the initial design of the viscous jet injector was an asymmetric curling of the gel column after extrusion. Adding a gas sheath similar to the design of the GDVN mostly solved this problem (more on this later). The newer design follows the same principle as the GDVN, with a coned fused silica capillary (Molex/Polymicro TSP020375) inside a slightly melted square capillary (Friedrich and Dimmock Inc. BST-040-10). In the current viscous jet injector design, the coned capillary tip defines and extrudes the gel column, while the square capillary acts as a gas aperture with the melted end directing gas flow around the gel column for stability. In studies performed at CHESS, we focus on using a 20 μm diameter inner capillary, though we have the capability of switching to 5, 10, and 50 μm diameter inner capillary. The tips of the capillaries must be polished down to a cone, so that the protrusion of the capillary is in a lower shear region of the gas flow just outside the end of the gas aperture. An apparatus was specially machined in the Gruner lab to hold and rotate the silica capillary while polishing on a Allied Multiprep polisher to be within the ASU designated limits of 15-20°. Next, a 1 cm piece of square capillary for the gas aperture is glued to one end of a 2.5 cm piece of stainless steel tubing. The exposed end of the square glass capillary is melted using a cigarette lighter, until an inserted coned capillary can protrude approximately 200 μm from the

end. Both the polished coned capillary and the melted gas aperture are made in lab, and need replenishing often. Details of the fabrication of these pieces are found in [189].

The stainless steel tubing is then seated into a 10-32 coned port fitting for the end of the nozzle. The coned capillary is fitted with a Teflon adapter sleeve and a sealing ferrule, and inserted into the stainless steel nozzle body. The fitting with the gas aperture is threaded into place on the nozzle body, and the length of the coned capillary adjusted until the conical tip protrudes from the front of the gas aperture. Excess capillary is cut off and the nozzle body may be threaded onto the reservoir for complete assembly. The nozzle body is tightened to the reservoir using wrenches to provide the proper seal with the ferrule.

In-air operation

The reservoir is screwed onto the secured hydraulic stage and the gas line attached to the gas port on the nozzle body. Next, we align the tip of the coned capillary using the microscope camera, ensuring that the x-ray-interaction region is also in view. The gas flow is turned on and we advance the injector's piston to contact the Teflon spheres using the Tee-attached syringe. The injector, now ready for use, can be controlled outside the experimental hutch *via* the pump and remote control programming provided by Shimadzu, the pump manufacturer.

4.3.2 Pump timing

For storage ring sources, the x-ray beam is ostensibly an effectively constant source and the exposures are limited by the detector frame rate. Let us assume that the microcrystals will only survive for a single exposure of the detector, and that there is only one crystal within the illuminated gel volume of the beam (this is idealized for a radiation limiting exposure). Assume then that the optimal flow rate of the gel, Q, will expose fresh gel volume for every frame of the detector. Therefore, the detector frame rate sets the ideal flow rate of the gel:

$$Q = \pi * b * \left(\frac{d_{cap}}{2}\right)^2 (d_{beam} \times f_{det}) \tag{4.1}$$

commonly described in μL per minute, where d_{cap} (μm) is the inner capillary diameter of the injector, d_{beam} (μm) is the diameter of the beam, and f_{det} (s^{-1}) is the frame rate of the detector. Slower flow rates with similar detector frame rates will result in more recorded exposure of the same volume (*i.e.*, multiple recorded exposures of the same crystal), as d_{beam} can be substituted for whatever gel-displacement between detector frames is desired. b is merely a conversion factor for μm^3 per second to μL per minute, 6×10^{-8} $\left(\frac{\mu L}{min} \times \frac{s}{\mu m^3}\right)$.

4.3.3 Crystal preparation in agarose

Three conditions must be met with the selection of a crystal medium: it should not compromise crystal quality, it should form a continuous column flow and it should contribute minimal background scattering. Initially, LCP was proposed as the choice medium [8], as it was already established as a preferred viscous growth medium for membrane protein crystals [8, 70, 77]. Examples of mediums in recent studies include oil-based grease [88], petroleum jelly [51], agarose

[89], and high molecular weight PEG [90]. Agarose is compatible with a wide range of proteins including soluble and membrane proteins, and in 2015 Conrad *et al.* [89] suggested using agarose as a low-background scattering medium for the viscous jet injector. Since we are highly aware of the impact of background scatter and appreciate the flexibility of agarose as a medium, we chose to use agarose gel for testing the viscous jet injector at CHESS. Crystals can be either grown in agarose or transferred to agarose using a method of mixing proposed by Cheng *et al.* [190]. The latter, as our preferred method, involves coupled two syringes, one with agarose made from the mother liquor of the crystals, and one with concentrated microcrystals in mother liquor. The syringes are mixed until an even distribution of crystals is achieved. All agarose-crystal mixture is left in one syringe, to which the loading needle is attached for transfer to the injector's reservoir.

For setups at CHESS, we grew lysozyme crystals in large 1 mL batches at 277 K based on recipes from Falkner *et al.* [66]. Various concentrations of buffer (*i.e.*, 15-18% *(w/v)* NaCl with 6% *(v/v)* PEG 2000), protein solution (4.5-8 mg/mL) and precipitant were prepared at pH 3.0-4.0 to provide a range of microcrystal size (sub-micron to approximately 10 micron) and densities. After growth, crystals were loosely spun down to a pellet for transferring to agarose gel *via* mixing as per Cheng *et al.* [190].

Following preparations of Conrad *et al.* [89], ultralow-gelling-temperature agarose (Sigma-Aldrich, catalog No. A5030) was dissolved in a mixture of 30% *(v/v)* and mother liquor buffer to a concentration of 7% *(w/v)* and submerged in a boiling water bath for 30 minutes. Once dissolved, 20 µL of the warm agarose mixture was drawn into a warmed syringe and then allowed to equilibrate to

room temperature before mixing with 5 μL of concentrated protein crystals in solution. This ratio can be adjusted, though the final concentration of agarose should be close to 5.6% *(w/v)* for a stable gel stream, per Conrad *et al.* [89].

Ideally only single crystals occupy the illuminated volume of the beam, but crystal density within the gel can be adjusted to create the proper spacing of crystals between exposures [12]. From knowing crystal size and ratios of gel to protein solution prior to mixing, an approximate number density of crystals within the agarose gel can be estimated of crystals per unit volume. Concentration of crystals within the sample together with flow rate, Q, can predict the probability for a single crystal being within the intersection volume of the beam and the sample stream. Higher concentrations of crystals run the risk of clogging the injector, and having multiple crystals within the illuminated beam volume. But lower concentrations run the risk of having fewer crystal hits per unit volume. The objective is to maximize single crystal hit rates while avoiding blockages from over concentration of crystals.

4.3.4 Experimental setup

The jet injector was installed at CHESS on beamline F1 (λ = 0.9773Å, E = 12.68 keV, x-ray flux = 1.4 $\times$ 10^9 x-rays per second), using a 100 μm monochromatic x-ray beam from a 24-pole wiggler. The injector was mounted in-air vertically at a distance of 300 mm from an ADSC Quantum 270 (Q270) detector. The detector face was oriented perpendicular to the beam and data sets were collected at room temperature (~ 298 K). A microscopic camera was mounted at 45° to monitor flow quality. Lysozyme crystals of ~1-5 μm were mixed with agarose

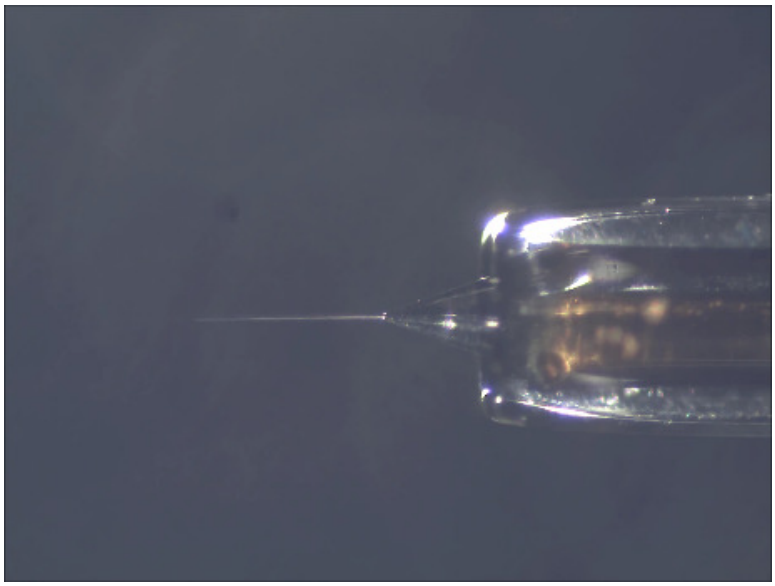


Figure 4.7: An example of a stable 20 μm agarose extrusion. The injector is mounted horizontally off-beam, so that visualization is improved. An air column of helium is used to stabilize the column.

gel to achieve a 5.6% *(w/v)* agarose reservoir solution. The injector was prepped as described above and a syringe pump operated at 1.6 $\times 10^{-4}$μL per minute to provide hydraulic pressure to the piston. The injector was centered on the beam, where the interaction point was about 100 μm from the tip of the inner capillary. Exposures were recorded separately for 1, 10 and 15 seconds as the injector extruded gel across the beam. No other background reduction measures were taken to avoid air scatter.

4.3.5 Results

All of the exposures with a stable column of gel (an example is shown in Figure 4.7) showed no visible Bragg diffraction. A slight change in diffuse background was the only indication that the x-rays were interacting with the gel.

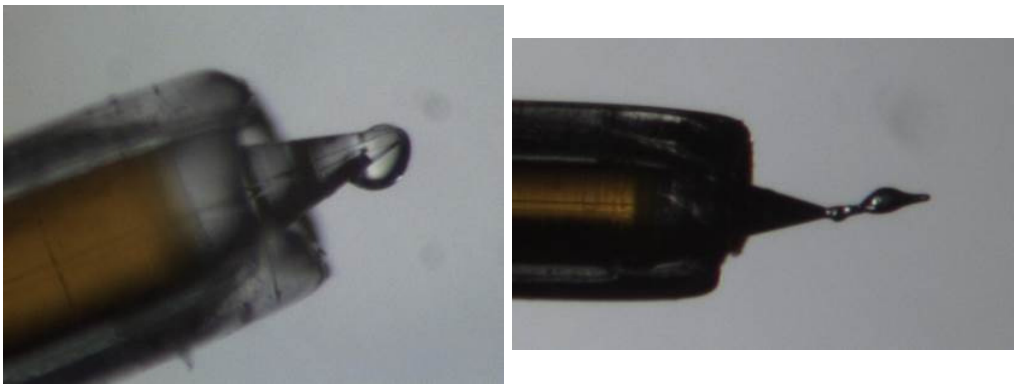


Figure 4.8: Two examples of problematic streams with the viscous jet injector. Due to variations in mixing, concentration, or crystal density, the gel column would form spherical blobs or asymmetrical columns near or in the x-ray interaction region, and continue growing until they either spontaneously disappeared, or repetitively created new spherical blobs. An increase in background signal could be seen in exposure images when the blobs where within the x-ray interaction point. Usually, this required cleaning with a damp Kim wipe to remove the excess gel.

As the injector operated, there were instances when the column would create a spherical shape of gel near the tip of the inner capillary, or "blub" (some examples of problematic streams are shown in Figure 4.8). When this occurred, the background scatter from the gel would increase diffusely around 4.5 Å in the exposures, but no crystal diffraction could be detected. Instabilities would occur frequently, and require cleaning the tip with a damp Kim wipe. The injector also clogged occasionally, where the stream would stop until either the clog passed and the stream resumed, or the pressure build up would force the inner capillary through the outer capillary, breaking the nozzle tip. In the former case, there was no need to disrupt data collection, however, in the latter, a complete reset was required.

Some consideration is needed for crystal size, column size (dictated by the inner capillary size) and buffer concentrations, since both the crystal concentration and size within the gel can cause injector blockages. Others have suggested

that maintaining a column diameter four or five times that of the crystal diameter reduces the majority of blockages, and our recent observations agree. However, Nogly *et al.* [53] and Botha *et al.* [51] report using crystal sizes just under the diameter of the inner capillary with success. More optimization is needed with crystal size, gel viscosity and concentration for a good determination of this parameter. The intention, now that the viscous jet injector exists at CHESS and has been initially set up, is to optimize its operation as a crystal delivery system. Clearly, much work remains to be done to determine the minimal crystal size that can be used with this system.

4.3.6 Conclusions

From initial testing with the viscous jet injector, several important lessons were learned. The viscous jet injector allows rapid delivery of microcrystals to the beam, while maintaining a relatively low-background compared to many other delivery devices. Comparable devices would need total device thicknesses of less than 20 μm to achieve similar background levels. Through adjusting flow rates and detector frame rates per crystal sample system, the injector can even optimize single exposures to a maximum tolerable dose per crystal if needed. Lastly, the injector is compatible with vacuum, which will further reduce the background scatter of the system.

While a theoretical flow rate, Q, was explained in previous sections, the actual flow rate of the injector varies somewhat randomly over time due to heterogeneity of the gel, *e.g.*, small clogs, air bubbles or pockets of unmixed media which change the density of the gel. In addition, the tip required cleaning from

time to time, as the gel column can destabilize and form globules at the tip which then disrupt the stream. Both of these complications led to a need for an operator to constantly monitor the injector for clogs and accumulation of gel on the capillary tip. There is also much setup needed to optimize the stream per sample, from adjusting concentrations, to monitoring gel and gas flow for a continual stream. However, once conditions are established per sample, the injector does deliver microcrystals in a steady stream for exposure at rates which are ideal for storage ring experiments (see flow rates above).

As mentioned above, the recommended ratio of column size to crystal size is ~5:1, which sets a fairly high amount of background material. Unfortunately, the background contribution from such a large column can easily obscure microcrystals less than a few microns across. More study is needed with viscous jet injectors that reduce the column size to less than a few microns and maintain stability.

Optimizing a beamline at CHESS for serial microcrystallography using the viscous jet injector is currently in progress. As known from previous chapters, reducing sources of background scatter such that the primary source of x-rays is from the crystal itself is the preferred approach for SMX. Implementing this at CHESS requires a complete overhaul of the typical MacCHESS beamline, including installing a sample chamber for either vacuum or helium compatible with the injector.

4.4 Summary

This chapter has covered two very different devices which have explored the delivery of microcrystals into the beam at CHESS for reliable serial microcrystallography. With modifications needed to minimize mounting material thickness or integrate *in vacuo* operation, both devices show promise for low-background serial microcrystallography.

Even if these devices are optimized to a point where they contribute a minimum amount of background scatter, techniques are needed to deal with data from microcrystals which do not generate discernible Bragg peaks within any single frame. This is especially true at storage ring sources, where radiation damage limits the amount of information attainable from small crystals. In the next chapter, a method which aims to provide just such a technique will be introduced.

CHAPTER 5

EXPAND-MAXIMIZE-COMPRESS ALGORITHM

5.1 Introduction

The overall pipeline of protein crystallography is straightforward once diffraction data of suitable quality are collected. From a data set of diffraction images, crystallographers index Bragg peaks in individual images, calculate (*hkl*) intensities across the set of images and (with some phasing) solve the structure of the protein. For over half a century, this crystallography pipeline has worked with great success, and remains largely unchanged even for serial microcrystallography. But the pipeline relies heavily on the first step after recording the image - indexing the individual images, *i.e.*, assigning images an orientation of the crystal with respect to the beam based on the Bragg peak pattern. When indexing fails, images are considered useless and the pipeline breaks down.

Consider an experiment where microcrystals are delivered to the beam in random orientations, and only single images are recorded before radiation damage renders the crystal unusable. Now consider that the inherently weak diffraction from the microcrystals produces datasets which are so severely Poisson noise limited that no single image contains enough information for indexing (indexing fails). This is not unheard of, as these 'sparse' images can be found within microcrystallography datasets from storage ring sources; *e.g.*, Gati *et al.* [11] reported indexing only 2,233 out of 28,800 (7%) images with microcrystal diffraction and Stellato *et al.* [33] reported indexing only 40,233 images out of over 1,490,000 (2.7%) images with microcrystal diffraction. The remaining 93-97% of images contained Bragg diffraction, but were determined to be not 'in-

dexable', and they are thrown out! To be specific, the discarded images either contained more than one diffraction pattern (multiple crystals) or did not contain enough Bragg diffraction for indexing algorithms to process the image. In the latter case, the single crystal still contributes Bragg diffraction in the image and so it is still useful for structure solution, if it could be assigned an orientation. Intuitively, one might believe that sparse images can never be merged into complete data sets. The challenge is to access the information in the sparse images through a redundancy in a large number of measurements sufficient for a unique reconstruction.

A solution to this challenge began as a method proposed for single particle imaging at XFELs, where particles in random orientation diffracted from the ultra-bright, ultra-short XFEL pulse to produce millions of Poisson-limited images. Loh *et al.* [43] proposed the Expand-Maximize-Compress (*EMC*) algorithm to simultaneously assign orientations of the individual images and recreate the 3D intensity distribution which the orientations sampled given enough measurements are available. This has proven very successful in reconstructing structures from sparse data frames of several systems [15, 44, 112, 191]. When applied to protein crystallography, the *EMC* algorithm can reconstruct the 3D reciprocal space intensity distribution from sparse diffraction images of protein crystals in unknown orientation (discussed in Chapter 6) [17]. Eventually, the hope is to apply the *EMC* algorithm to datasets in serial microcrystallography which current indexing algorithms would otherwise fail to process.

In this chapter, we will walk through the evolution of proof-of-principle experiments with the *EMC* algorithm, where the datasets are characterized by sparse data images with unknown orientation. We begin by describing the gen-

eral principle behind the *EMC* algorithm and then detail three studies of increasing geometric complexity: a 2D shadow radiography study to reconstruct the shadow of a randomly rotating lead mask [112], a 3D shadow radiography study to reconstruct a 3D structure of a rotating plastic figure [15], and a crystallographic study to reconstruct a 3D reciprocal space pattern from a rotating inorganic crystal [44]. To generate the necessary sparse data frames, the studies used a very weak x-ray source and fast framing detectors developed by the Gruner Group at Cornell University. This chapter sets a foundation for the study in Chapter 6, details the first protein crystallography experiment with *EMC* [17]. The goal is to make a convincing argument that the study of serial microcrystallography in the sparse data regime is possible with the application of the *EMC* algorithm.

5.2 Expectation maximization algorithm

The *EMC* algorithm is based on the principle of expectation-maximization, an iterative technique for finding a maximum likelihood estimate for some parameter of a statistical model [192]. In the general context of *EMC*, the 'model' is the spatial intensity distribution recorded in the detector data frames and the 'parameter' is an orientation of the specimen giving rise to a given data frame. *EMC* is a three-step iteration which begins with a guesstimate model. In the first step, the model is expanded to expected average photon counts within individual data frames, which we consider 'views'. The views are compared to the real data frames, and a probability distribution is assigned to each real frame for all views. In the second step, *EMC* maximizes the likelihood of the views generating those probabilities, and the views are updated. In the last step, the views

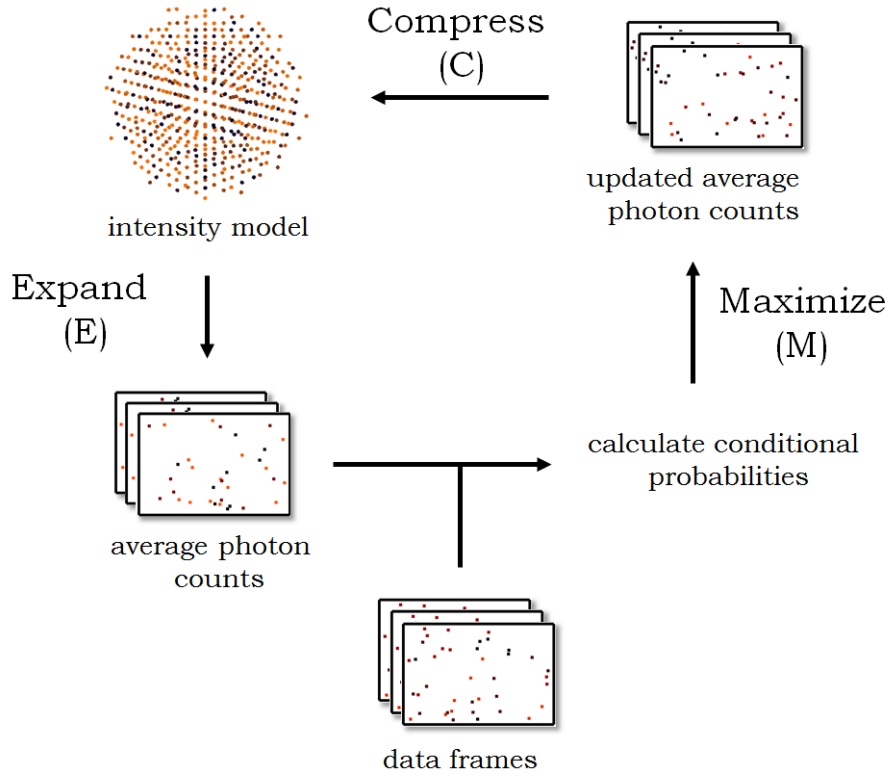


Figure 5.1: A schematic of the *EMC* algorithm for intensities in 3D space. It consists of three steps: Expand (*E*) the intensity guess (or model) to average photons counts as seen by the detector, Maximize (*M*) the probabilities of the data frames compared to the average photon counts and update the photon counts, and Compress (*C*) the updated photon counts into a new intensity model. The algorithm cycles until it converges upon a solution which agrees with the data frames.

are compressed to form an updated model and *EMC* repeats until the views converge with the real data.

Consider a large set of sparse x-ray images generated from illuminating an object sampled at random orientations, such as what would be recorded in a SMX data set. Assume the object samples a uniform distribution of orientations, such that Poisson statistics are applicable and photon counts are independent samples of intensity at each pixel of the detector. Next, assume the frames are sparse enough (perhaps from tiny crystals) that photon counts per frame are not

sufficient to identify the object's orientation. Since the average intensity, W, of a pixel follows Poisson statistics, then the probability of recording K photons is given by

$$P(K) = \frac{W^K e^{-W}}{K!}. \tag{5.1}$$

This can be modified to describe the probability of recording photons within a data frame as follows. No information is known about the object (including its orientation), so a random estimate of the sampled intensity distribution, $W(\mathbf{q})$, can serve as the initial model for *EMC* (see Figure 5.1). $W(\mathbf{q})$ represents the intensity of x-ray scatter at a given location, $\mathbf{q}$, in the space being observed. For example, for a crystal, W is the magnitude of reciprocal space at reciprocal space position, $\mathbf{q}$. A discrete number of sample space views, Ω_j, is established for model orientations j which are separated by some sampling angle distribution $\delta\mathbf{q}$ (this is dependent on the rotational geometry of the experimental setup and will be explained in each following study. Below, we use $\mathbf{q}$ to represent the position of $\delta\mathbf{q}$).

The expansion step samples slices of $W(\mathbf{q})$ for object orientations Ω_j. Intensity slices are arrays W_{ij} of average photon counts at pixel i within a frame when the object has orientation Ω_j. The probability of the real data frame k being generated by the object in view Ω_j and recording K_{ik} photons at pixel i, is given by

$$P_{jk}(W) = \frac{\prod_i W_{ij}^{K_{ik}} e^{-W_{ij}}}{\sum_j \left(\prod_i W_{ij}^{K_{ik}} e^{-W_{ij}} \right)}. \tag{5.2}$$

This is simply the probability that a data frame is generated by a specific model view. In short, each real data frame is given probability distributions for every model view.

In the maximization step, these probabilities $P_{jk}(W)$ update the average photon counts W_{ij} (or intensity slices) to maximize the likelihood of generating the real data frames, where the log-likelihood:

$$log[Q(W'_{ij})] = \sum_i \sum_j \sum_k [P_{jk}(W)(K_{ik} log[W'_{ij}] - W'_{ij})] \tag{5.3}$$

can be simplified to:

$$W_{ij} \rightarrow W'_{ij} = \frac{\sum_k P_{jk}(W) K_{ik}}{\sum_k P_{jk}(W)}, \tag{5.4}$$

This is a weighted mean of expected photon counts according to the probability distribution $P_{jk}(W)$ in the current model, and generates the updated average photon counts, W'_{ij} for the next iteration.

The compression step subsequently maps the updated W slices back to an updated intensity model $W'(\mathbf{q})$, which ensures consistency of intensity slices in the next round. The cycle continues until the model converges upon a single view to within some reasonable end. Using this scheme, the *EMC* algorithm searches for the most probable intensity distribution that is consistent with all the real data frames. Placing constraints such as limiting $\mathbf{q}$ to a single rotation axis typically aid in shrinking the number of iterations needed for convergence.

EMC is powerful tool to use large sparse data sets in reconstructing structure from otherwise 'useless' data frames. In the context of microcrystallography, *EMC* may provide the necessary catalyst to obtain information from Bragg diffraction images which indexing algorithms fail to process. The next sections follow studies which provide proof-of-principle examples of the application of *EMC*, which all share the common characteristics of SMX in sparse data frames and random, unknown orientation.

5.3 Two-dimensional shadow radiography experiment

This section is modified from the text of "Solving structure with sparse randomly-oriented x-ray data" published in Optics Express by Hugh Philipp, Kartik Ayyer, Mark Tate, Veit Elser and Sol Gruner [112]. It is not my own work, but rather is presented here, by author permission, as of a review and background for Section 5.5.

In the following experiment, Philipp *et al.* [112] demonstrate that a simplified version of *EMC* algorithm is capable of reconstructing a 2D image from randomly oriented sparse data frames. In this shadow radiography-*EMC* study, the model is a 2D disk of random intensities, and the parameter is the orientations of the 2D sparse data frames about the rotational axis. With an average of 2.5 photons per frame, the algorithm successfully reconstructs the shape of a rotating lead mask placed over a uniformly illuminated low-fluence x-ray beam. Although this experiment is the simplest of the proof-of-principle experiments, the strikingly low number of photons needed per frame in order to have a convergent solution was an encouraging result for subsequent studies with *EMC*.

5.3.1 Experimental setup

A 2D mask was cut from an x-ray opaque lead sheet as seen in Figures 5.2 and 5.3(a) and then mounted on a rotating stage. A low-power copper anode x-ray tube generated 8 keV Cu K_α x-rays used to evenly illuminate the mask. The rotating stage and mask were placed within the beam-path directly in front of a single-chip 194×185 pixel array detector (CS-PAD [111]), such that the mask

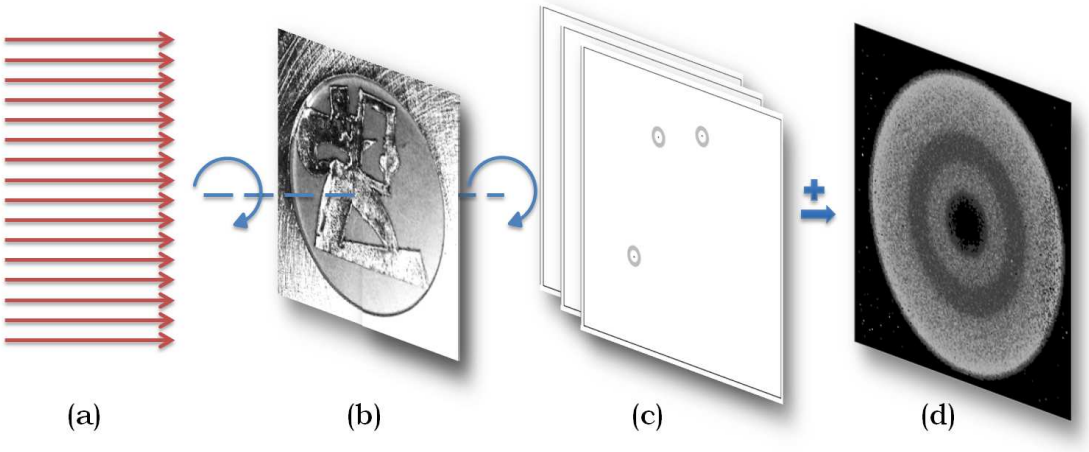


Figure 5.2: Schematic of the experimental setup of 2D shadow radiography. (a) Incoming x-rays fully illuminate a lead mask (b) which is rotating about an axis parallel to the x-ray beam. A detector records frames (c) with an average of 2.5 photons per frame. When summed (d), the 450,000 data frames used later in reconstructing the mask show a uniform distribution about the rotation axis.

was centered on the face of the detector chip. The axis of rotation was along the x-ray beam direction. To generate a static image, the detector collected 432 consecutive frames without rotating the mask, which were then summed to make the image in Figures 5.2 and 5.3(b). The image represents a control for the system and demonstrates the opacity of the mask to x-rays.

For sparse randomly oriented data sets, the mask was rotated continuously within the beam while the detector collected images at 100 Hz. The image processing did not utilize information about the rotation angle or orientation of the mask, and were treated independently during processing. Additional thresholding was needed to remove noise from the detector frames and convert from an integrated value at each pixel to photon counts for further analysis. Summed frames refer to adding thresholded frames. Two separate data sets were recorded with either low-fluence x-rays or very-low-fluence x-rays (determined by the x-ray tube current). The data sets produced photon hit rates of

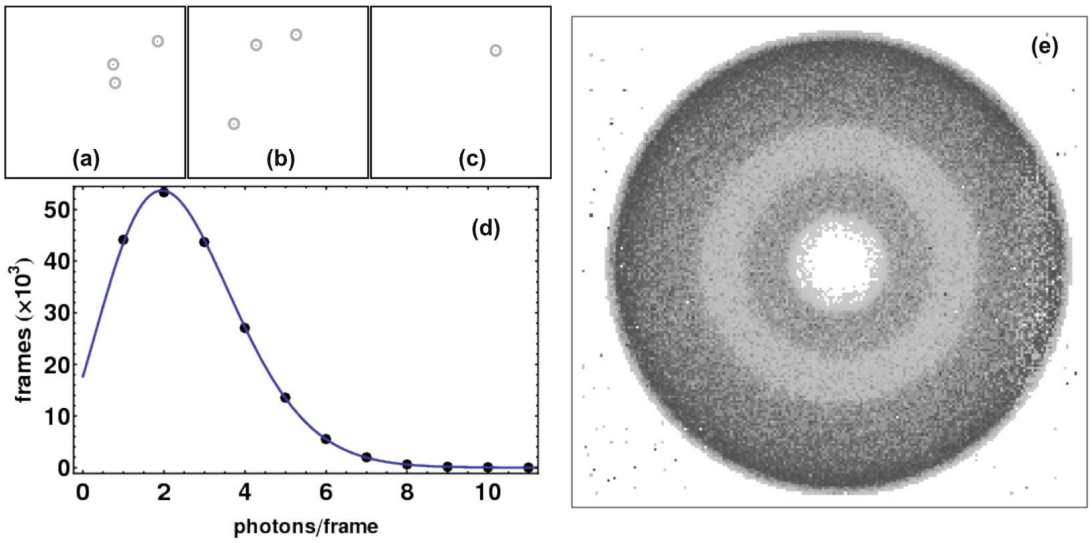


Figure 5.3: (a-c) Three example frames of sparse data photon hits from the 2.5 photons-per-frame data set. X-ray photon hits are circled for better visual identification. (d) Photon occupancy histogram compared with the Poisson distribution. (e) Powder pattern of all thresholded frames summed together within the 2.5 photons-per-frame data set. The smooth pattern indicates good angular distribution throughout orientations. Reproduced from Philipp *et al.* [112].

approximately 11.5 photons per frame and 2.5 photons per frame, respectively. Figure 5.3(a-c) shows three examples of images taken with the very-low fluence source. More details on thresholding and signal-to-noise can be found in results published by Philipp *et al.* [112].

5.3.2 Initial guess

For the initial model $W(\mathbf{q})$, *EMC* used a 2D disk with the diameter of the mask. Every pixel was randomly assigned a value in the range [0,1]. This model was the input for the first iteration of *EMC*.

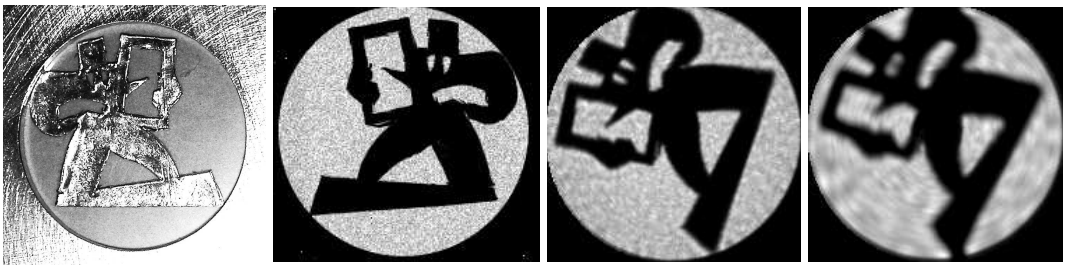
5.3.3 Rotation group subset

Since the data frames were constrained to rotation about an axis, the algorithm needed only sample angular distributions in one dimension. The model was given 180 discrete equally spaced 2° rotation steps, constrained about one axis to generate average photon counts W_{ij}.

5.3.4 Results and conclusion

As an example, Figure 5.3(e) shows the summation of all frames from one data set into a single image. The rotational smearing confirms a uniform distribution as a function of the rotation angles. Since *EMC* does not have prior information of the frame orientation, this smearing also confirms the random orientation needed for a proof-of-principle sparse data set. To confirm Poisson statistics assumed in the *EMC* maximization step, a histogram of the photon occupancy per-frame was plotted for the 2.5-photon-per-frame data set and is shown in Figure 5.3(d).

The reconstructed images are shown in Figure 5.4(c) and Figure 5.4(d). Both data sets, the 2.5- and 11.5- photon-per-frame, acquired 1.2 million photons total over their respective frames count total. Figure 5.4(d) was reconstructed using 450,000 data frames, which averaged 2.5 photons per frame and, in comparison, Figure 5.4(c) was reconstructed using 100,000 data frames, which averaged 11.5 photons per frame. The two data sets differ in several characteristics; for example, the mask image is much sharper in the 11.5-photons-per-frame reconstruction, which agrees intuitively with having comparatively more data within each frame. Also, the number of iterations *EMC* needed for the model to con-



(a) (b) (c) (d)

Figure 5.4: (a) The 2D lead x-ray mask mounted within an aperture in an aluminum disk. This was placed in front of the detector on a rotating stage. (b) A static x-ray image of the pattern collected as 432 individual frames with approximately 0.2 photon per pixel per frame. (c) Reconstructed 2D image from randomly-oriented data having an average 11.5 photons/frame and 1.2 million recorded photons. (d) Reconstructed from randomly-oriented data having an average 2.5 photons/frame and 1.2 million recorded photons. Though these images look aligned, the final image alignment depends on the initial model and is arbitrary. Reproduced from Philipp *et al.* [112]

verge differs dramatically, as the 2.5-photon-per-frame set required 220 cycles, while the 11.5-photon-per-frame only required 49 iterations.

This study provided the first experimental results of a 2D radiographic reconstruction from extremely sparse, randomly oriented data frames. These results encouraged Philip *et al.* [112, 15] to continue more advanced applications of *EMC*. The next section describes the first application of *EMC* to a more complex, 3D system and the resulting tomographic reconstruction.

5.4 Three-dimensional shadow radiography experiment

This section is modified from the text of "Real-space x-ray tomographic reconstruction of randomly oriented objects with sparse data frames" published in Optics Express by Kartik Ayyer, Hugh Philipp, Mark Tate, Veit Elser and Sol Gruner [15]. It is not my own work, but rather is presented here, by author

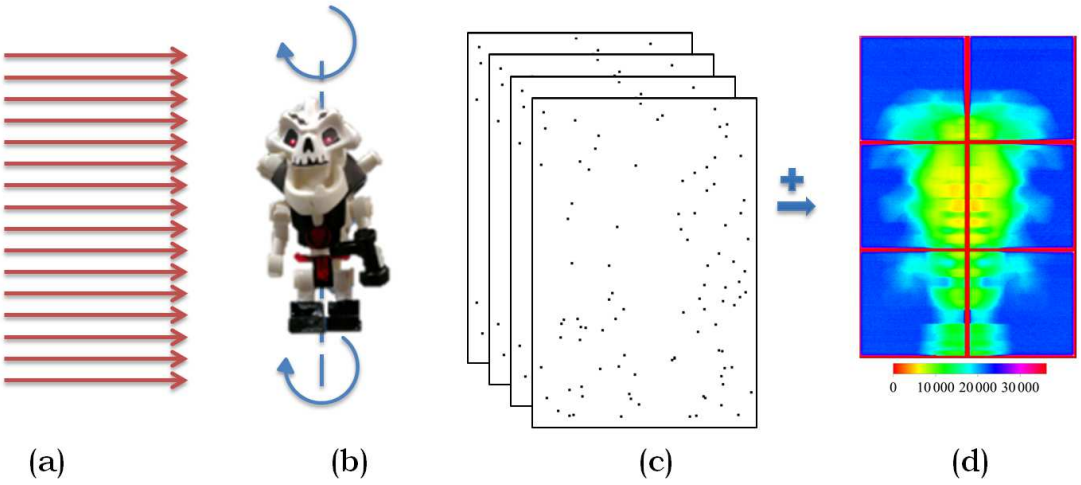


Figure 5.5: Schematic of the experimental setup of 3D shadow radiography. (a) Incoming x-rays fully illuminate a figure (b) rotating about an axis perpendicular to the incoming x-rays. An MM-PAD (not shown) detector records frames (c) with an average of 99.5 photons per frame. When summed (d), the 15.6 million data frames used later in reconstructing the figure show a uniform distribution about the rotation axis with appropriate angle-averaged shadowing from the rotating figure.

permission, as of a review and background for Section 5.5.

To extend *EMC* to a 3D study, x-ray transmission data was collected from a fully-illuminated 5 cm plastic figure rotating about an axis parallel to the detector face. The resulting *EMC* application generated a tomographic reconstruction of the 3D attenuation from the plastic figure. For *EMC* in this study, the intensity model, $W(\mathbf{q})$, is a 3D attenuation distribution in reciprocal space, and the parameter is the orientations of the 2D sparse data frames about the rotational axis. With an average signal level of $\sim 10^{-3} - 10^{-2}$ photons per pixel, Ayyer *et al.* [15] demonstrated a successful reconstruction of the 3D real-space tomographic projection. This provided additional support and extension of the 2D study in Section 5.3 which brings application of the *EMC* algorithm one step closer to serial microcrystallography.

5.4.1 Experimental setup

The plastic figure (Figures 5.5 and 5.6(a)) was attached to a rotation stage and a low-power molybdenum anode x-ray tube generated 17.4 keV Mo K_α x-rays to evenly illuminate the figure (schematically shown in Figure 5.5). The figure was placed in the beam path directly in front of a 3×2 tiled MM-PAD [42]. Each chip has 128×128 pixels measuring 150×150 μm. The MM-PAD is an analog integrating detector with a signal to noise radio for Mo K_α x-rays of ~12. Hence, thresholding these sparse data frames was used to remove detector noise. These thresholded frames were then used for the analysis as follows. Summed frames refer to adding thresholded frames. The rotation stage (and figure) were set to continuously rotate at 2° per second while the detector recorded 4 ms exposures with an average signal of $\sim 10^{-3}$ photons per pixel per frame. A total of 15.6 million frames were recorded in batches of 1,000, between which variable time delays were inserted to ensure an unbiased sequence of frames. Additional data sets were generated by summing 2, 4, and 10 consecutive frames within the 1000 frame batches. With a rotation of only 8×10^{-3} degrees between frames, it follows that the summed frames give identical orientations of the object, even for the 10-frame summed set which only rotates 0.08°.

From the angle-averaged pattern shown in Figure 5.7, the symmetric coverage about the axis of rotation suggests that the figure sampled a range of rotational orientations. The number of known rotations within each dataset established that the figure sampled an uniform distribution of orientations, however no orientation information was passed to the algorithm. Due to gaps between and around the detector chips and defective pixels, masks were applied to omit the empty spaces from the reconstruction (shown in red bands in Figure 5.7).

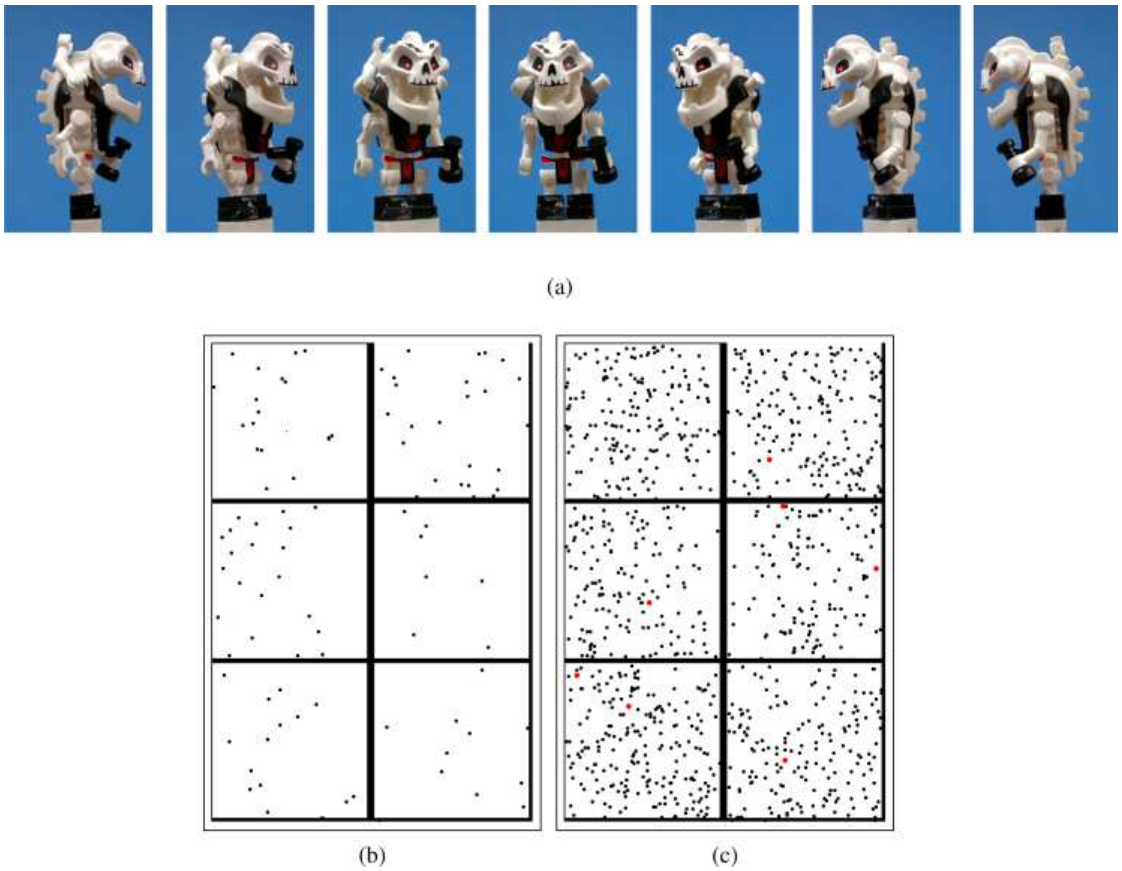


Figure 5.6: (a) Shows images of the rotated plastic figure used as the object in the 3D shadow radiography experiment. (b) Example of a data frame taken with the MM-PAD with 9.1×10^{-4} photons per pixel per frame. (c) Image of 10 summed sequential data frames. The summed frames have a combined 1025 photons with any overlapping photons highlighted in red. Pixels in (b) and (c) are enhanced for visibility. Reproduced from Ayyer *et al.* [15].

5.4.2 Tomographic-*EMC* adaptation

The images recorded in this study are snapshot attenuation projections of the figure in some random orientation. To transform the projections into slices of a 3D figure requires a modified Fourier transform, so it was convenient to iterate with a Fourier space model. This slight addition in complexity for *EMC* requires both the expansion and compression steps to Fourier transform back and forth from the 2D attenuation frames to the 3D Fourier model $W(\mathbf{q})$. Specifically, the

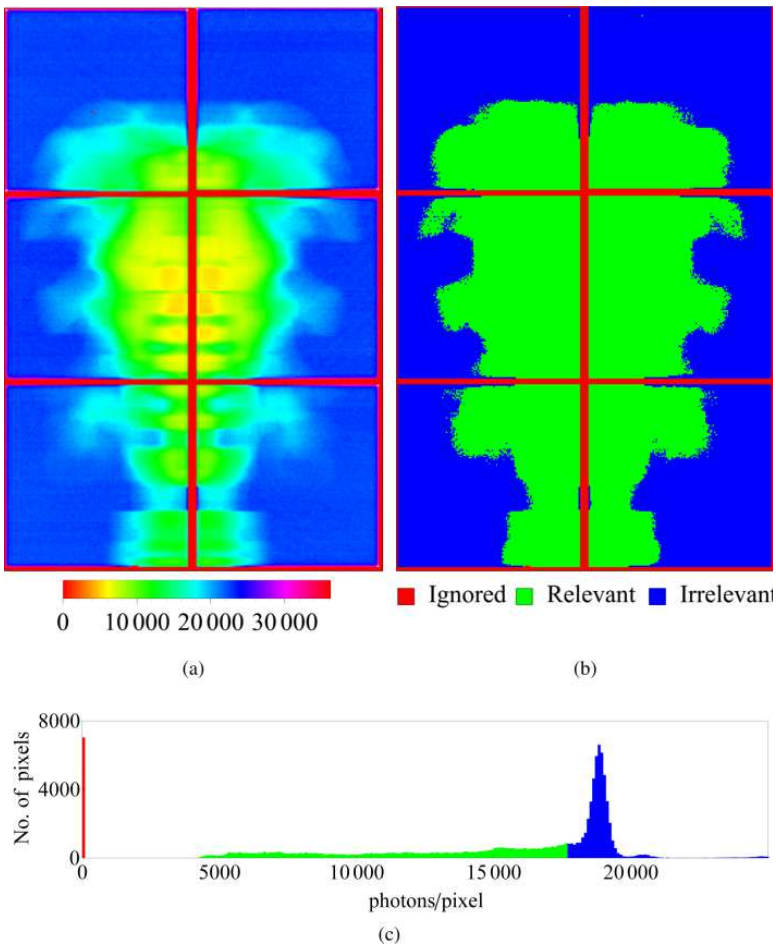


Figure 5.7: Angle-averaged pattern, detector mask and photon count histogram. (a) Shows an angle-averaged summation of data frames along with the masking of detector gaps. (b) Shows specific masking for the relevant and irrelevant areas of the detector. The photons recorded in the blue regions do not interact with the object, and contribute no information to the algorithm. (c) Shows a histogram of photon counts per pixel. The cutoff for relevant photon counts is 17,800, corresponding to the transition from green to blue in the histogram. This histogram was used to develop the mask shown in (b). Reproduced from Ayyer *et al.* [15].

expansion step required three additional steps: linear interpolation of the 3D model to slices, a Fourier transform from the slices to attenuated projections, and finally an exponential scale factor to generate the model intensity attenuation patterns (average photon counts). After maximizing the average photon counts, compression reverses these steps. The adaptation of *EMC* to tomogra-

phy is specific to this study only and the technical details of the conversion lie outside the necessary context of this dissertation. For those details, please see Kartik Ayyer's dissertation on Reconstructing Images From Sparse Data [16].

5.4.3 Initial guess

To make a 3D model, it was convenient to limit the spatial sampling to regions where photon counts could possibly provide structural information, *i.e.*, pixels of the detector where photons have the possibility of being attenuated by the figure during rotation. This histogram (Figure 5.7(c)), was used to decide cutoffs for the mask. From the angle averaged image shown in Figure 5.7, pixels outside of this region (Figure 5.7(b)) were set to zero (zeroed pixels are in blue). Rotating the pixel mask about the rotation axis limited the voxels of interest to those within the 3D mold (within the 3D green region). For the initial model $W(\mathbf{q})$, voxels within the mold were given a random number from [0,1] those outside are assigned zero. The array is Fourier transformed to create the initial model.

5.4.4 Rotation group subset

In the general form, *EMC* can be used to sample rotations about all three axes of the model. For this specific study, the angle averaged image shows a constraint about a single axis (otherwise it would resemble a sphere). To speed convergence, a restriction was placed on the sampling of the initial model to rotation about one axis in Fourier space for many uniformly-spaced orientations.

	Total photons/frame	# of frames (in millions)	Relevant photons/frame	Absorbed photons/frame
1	99.5	15.6	37	3.9
2	198.7	7.8	73.8	7.8
4	397.2	3.9	147.6	15.6
10	991.6	1.6	368.4	39.1

Table 5.1: Table showing properties of various data sets generated by summing 1, 2, 4, and 10 successive frames. Reproduced from Ayyer *et al.* [15].

5.4.5 Results and conclusion

Four data sets were generated from the same 15.6 million frames for a comparison of *EMC*'s resolution with average number of photons per frame. Each data set contained the same total number of photons, 1.56×10^9, with varying number of photons per frame dependent on the summed frame size. Table 5.1 lists the characteristics of the original, 2-, 4- and 10-frame summed data sets. For unsummed data frames, the mean number of photons per frame was 99.5, though only 37 were recorded on the relevant unmasked portion of the detector as seen in Figure 5.7. By comparison, summing 10 frames generates only 368.4 relevant photons per frame. A static image at high flux was taken as a control and it shown in the bottom row of Figure 5.8.

From Figure 5.8, the reconstructed tomographies increase in resolution and contrast as the mean number of photons per frame increases. This is intuitive, as the amount of information per frame increases. Compared, the four cases showcase the significance of information within a frame, as they all share the same initial data set. The difference is the number of photons per frame that *EMC* can use for reconstruction. The lowest resolution data, 99.5 mean photons-per-frame, has few distinguishing features, and converges to an overly-symmetric, and somewhat diffuse shape. The highest resolution data set, from 10 summed

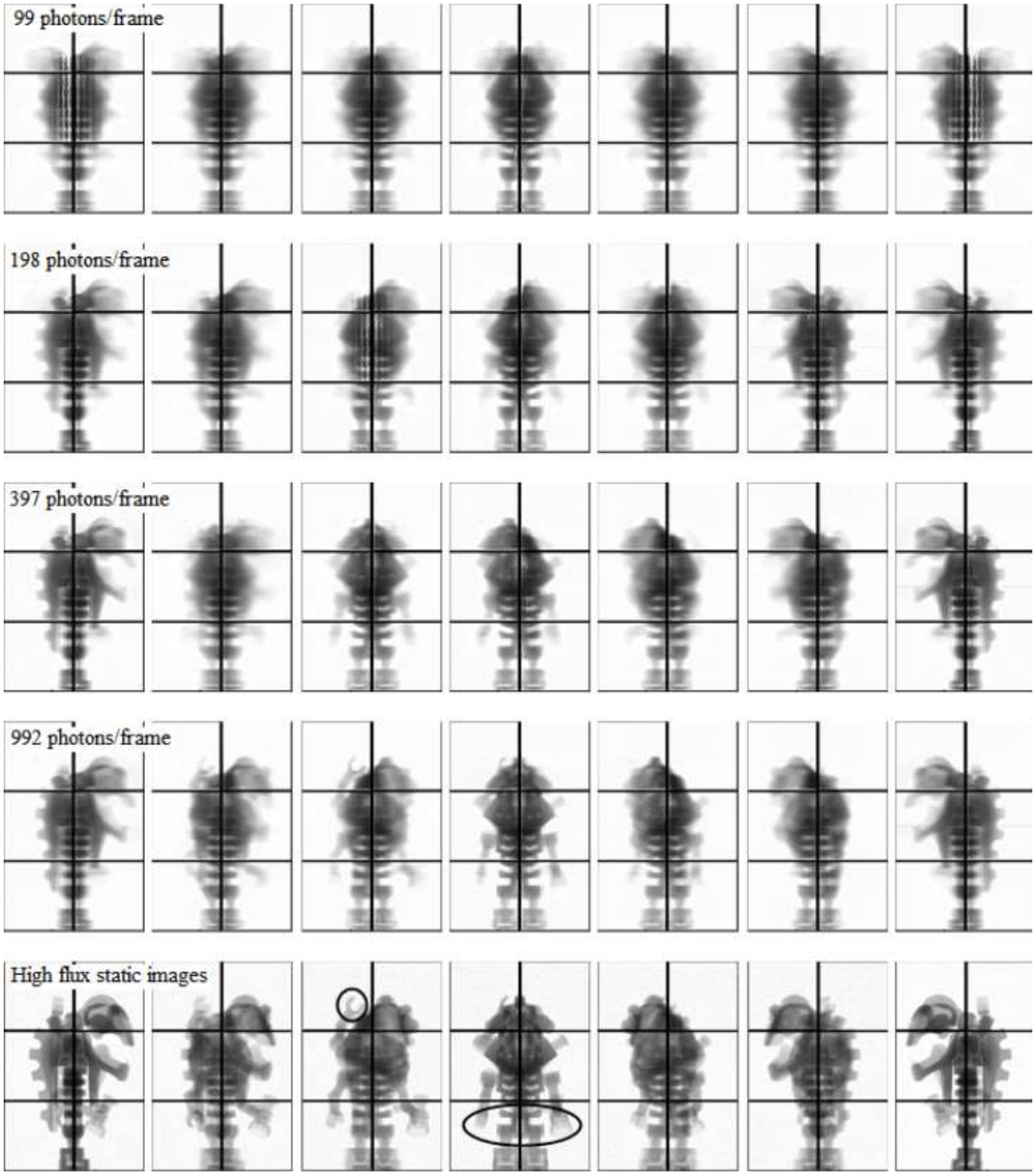


Figure 5.8: Comparisons of the reconstructed tomographies for the unsummed, 2-, 4-, and 10-frames summed with a high fluence static image (bottom row, some fine features are circled). The unsummed dataset averaged 99 photons per frame, and the summed averaged 198, 397, and 992 photons per photons per frame. Reproduced from Ayyer *et al.* [15].

frames, gives 991.6 photons per frame, and the figure has a decidedly asymmetric shape with one arm raised, along with other distinguishing features not seen with fewer x-rays per summed frame data sets. From these results, the inability of the 99.5 photons-per-frame set to form the asymmetric shape of the figure suggests that there is a minimum number of photons per frame needed to determine structure to high resolution. This is no surprise, as it has been seen in single molecule diffraction imaging [43, 193].

Ayyer *et al.* [15] provided proof that an extension of the 2D *EMC* study of Section 5.3 can be applied to randomly-oriented sparse radiographic data and successfully reconstruct the 3D density of an object. As yet another step towards crystallography, it is just shy of reconstructing randomly oriented diffraction data. The next section describes in detail a crystallographic experiment and the application of *EMC*.

5.5 Inorganic crystallography experiment

The contents of this section have been published in IUCrJ by Kartik Ayyer, Hugh Philipp, Mark Tate, Jennifer Wierman, Veit Elser and Sol Gruner [44].

In the first test of *EMC* with crystallographic data, an inorganic crystal was determined to be a sufficient sample to avoid the effects of radiation damage and the need for hydration. The lack of mounting and hydrating solution had the added benefit of not contributing to excess background scatter. To simulate the sparse data frame conditions for a 1 μm^3 crystal at a storage ring source, an experiment was performed with a large 400 μm crystal using the same standard laboratory x-ray source as Section 5.4. Each frame was acquired with the crystal

in an arbitrary orientation around a single rotation axis. Even with an average signal level as low as 48 photons per frame (4.8×10^{-4} photons per pixel), we demonstrated successful recovery of orientation about the axis of rotation and reconstruction of 3D intensities. We compared our reconstruction with a high-fluence data set where the orientations were recorded. We also examined the effect of background scatter on the quality of the reconstruction and the ability to recover orientations. The ability of the *EMC* algorithm to recover the 3D intensities from sparse data, albeit from a large crystal, is an important step towards the development of *EMC*-based serial protein crystallography.

5.5.1 Experimental setup

The sample studied was a 400 μm sized small-molecule crystal with chemical formula $C_{78}H_{120}Mo_2N_6O$ (mol. wt. 1.35 kDa). It was mounted on the end of a glass fiber attached to a goniometer head, which allowed the crystal to be centered on the rotation axis. A rotation stage (Newport URS100BPP) was set to rotate continuously at 0.1° per second during data collection. Although the angle of rotation was known for each frame, it was not passed to the reconstruction algorithm. The crystal was illuminated by a Molybdenum K_α beam generated by a Rigaku RU-3HR rotating anode set to 30 kV and 40 mA. Filtering was done using 200 μm of Zirconium foil to increase the fraction of K_α radiation. The x-rays were focused to a spot of size 0.5×0.5 mm using Ni-coated Franks mirrors 1m from the sample, with a beam convergence of 1mrad and 10^6photons per second. The data were recorded at a distance of 37 mm from the sample using the same MM-PAD detector as Section 5.4 [42]. Two data sets were collected, one with low fluence at 10 ms per frame and the other with high fluence at 500

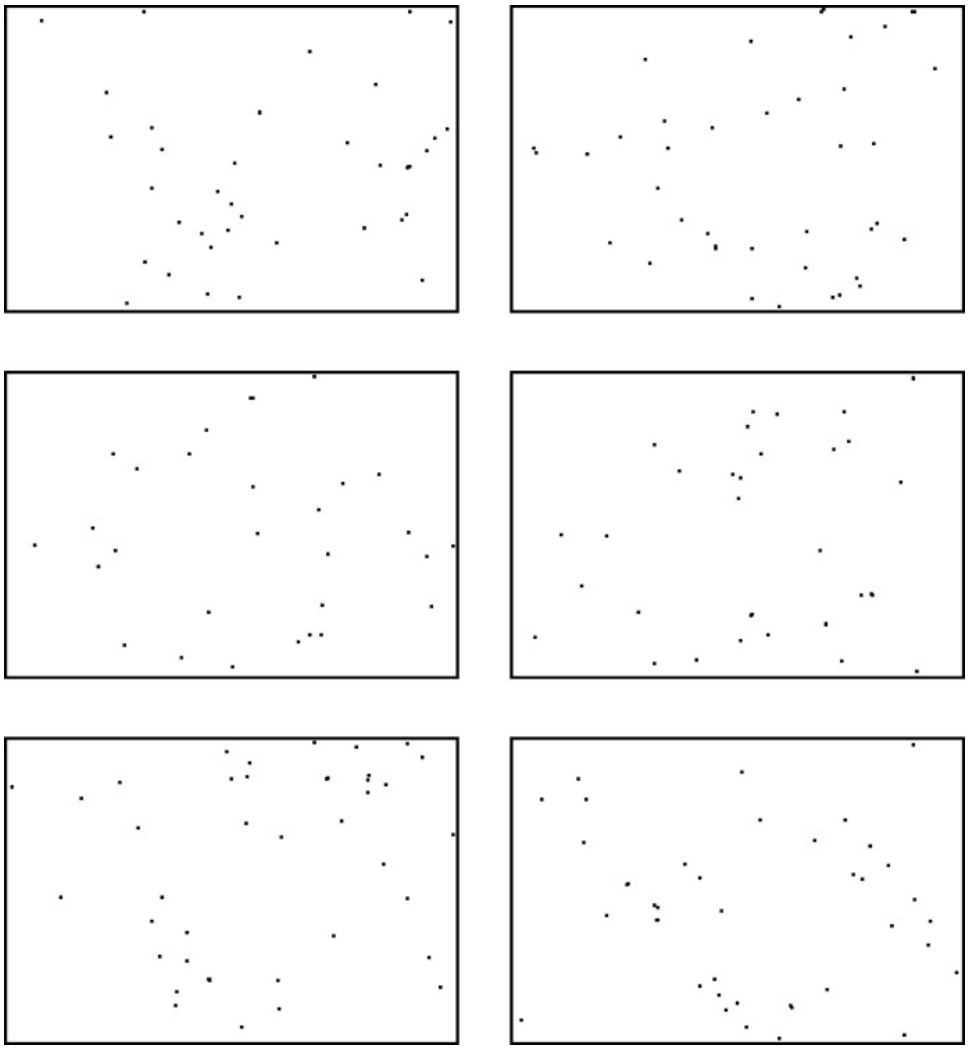


Figure 5.9: Six successive data frames obtained by summing 15 successive low-flux frames together. Each summed frame has 48 photons on average. The locations of the photons have been emphasized to improve visibility. Although the crystal rotates slightly between these frames (0.015°), this rotation is below the instrumental divergence in the apparatus. At high fluence per frame, each of these frames would look similar. The observed difference between these frames and the lack of any discernible Bragg pattern show the data to be well within the sparse-data regime.

Summed frames	# of frames	photons/frame	Iterations to converge
1	4,321,197	3.22	–
10	434,420	32.00	–
15	290,541	47.85	2200
100	44,221	314.41	400
200	22,321	622.88	250

Table 5.2: Table showing properties of various data sets generated by summing successive frames. Before summing, the frames were in batches of 1000 contiguous frames with gaps. There were also some rejected frames which had a very high photon count caused by incorrect detector offsets. The number of iterations required for convergence depends upon the random start, so the numbers given here are approximate and are used to highlight the trend.

ms per frame. The low-fluence data set was taken in groups of 1000 consecutive frames, with a time delay between sets to allow the frames to be written to disk.

The data were then thresholded and photon counts were obtained using a procedure similar to that employed in Section 5.4. In the low-fluence data set there were 4.3 million frames with an average of 3.2 photons per frame. Since the crystal rotated only 10^{-3} degrees between two successive frames, multiple data sets were prepared by combining successive frames within a batch. Table 5.2 lists the details of the different data sets. Figure 5.9 shows the first six frames from the 15 summed frames data set.

Figure 5.10 shows the angle-averaged pattern obtained by summing over all data frames. The radial streaks near Bragg spots are caused by the polychromaticity of the beam. These arcs near the rotation axis are caused by the these streaks intersecting the curved Ewald sphere. Since the exact shape of the arc is very sensitive to the rotation axis, a region of the rotation axis was not used in the calculation for P_{jk}.

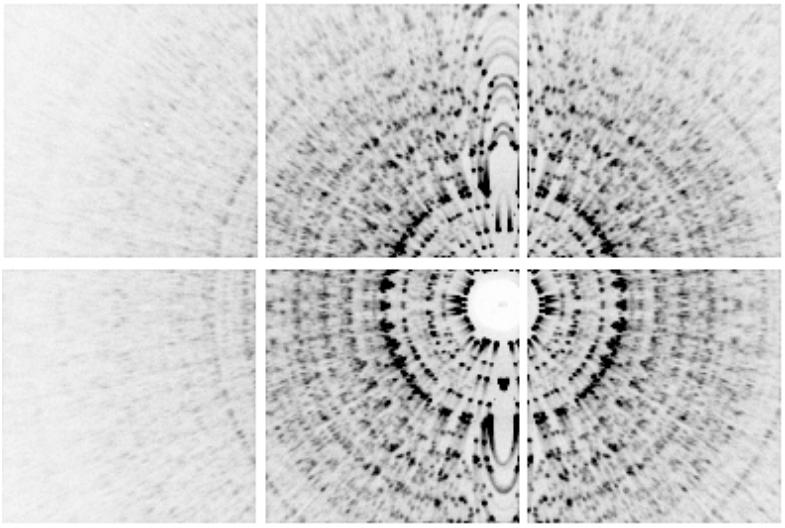


Figure 5.10: Angle-averaged pattern produced by summing over all frames in the low-fluence data set. The direct beam goes through the center of the beamstop and the rotation axis is vertical. Note the radial streaks caused by polychromaticity of the beam due to Bremmstrahlung. These streaks form arcs near the vertical rotation axis due to the curvature of the Ewald sphere. The white gaps in the images are the spaces between the six detector tiles in the 2×3 tiled array of the MM-PAD detector.

5.5.2 *hkl*-space *EMC* adaptation

In a crystallographic-*EMC* study, the model needed to be a reciprocal space distribution of Bragg peaks, and the parameter was the orientations of the sparse data frames. We used a slightly modified version of the *EMC* algorithm [43] to iteratively assign orientations and reconstruct the 3D intensity distribution. One feature of this technique was that all regions of reciprocal space are treated equally. No particular preference was given to reciprocal lattice points. This was in contrast to the approach taken by indexing algorithms which emphasize the Bragg spots to the extent of ignoring the diffuse scattering. In the case of sparse data, most Bragg spot positions contribute no photons and some of the photons could be from non-Bragg background, making this approach impractical.

One modification to the traditional algorithm was in the choice of space for the 3D intensity distribution. The standard choice is Fourier space, where the slice representing the detector plane is the surface of a sphere passing through the origin (the Ewald sphere). Here, for reasons explained below, the best choice was *hkl*-space where the three axes represent the fractional coordinates with respect to the reciprocal unit cell. Thus, the reciprocal lattice points laid on a cubic grid with an integer spacing regardless of unit cell parameters. Unless the crystal has cubic symmetry, the detector pixels no longer lie on the surface of sphere, but along some other surface. The pixel coordinates in this space could be calculated by using the basis vectors to determine the scaling and rotational transformation to the Ewald sphere surface.

With our choice of *hkl* space, the Bragg peaks all mapped to a cubic grid. Thus, while interpolating in the expand and compress steps, symmetry-related Bragg peaks saw the same environment, which would not necessarily have been the case if the basis vectors did not lie along the grid axes. This was important, because the maximize step is sensitive to the slight variations among these peaks caused by interpolation errors. These errors were not a major factor in the case of non-Bragg intensity distributions where the variation was smooth on a one-pixel scale.

To define such a space, one needed to know the unit-cell vectors. These could be determined from the (high-signal) angle-averaged (powder pattern) data, if not already known.

5.5.3 Initial guess

In this case, once the mapping of pixels to *hkl* space has been defined, the model was constructed by placing a 3D Gaussian of random height at each lattice point. This cubic grid of random intensities was used as the input for the first iteration, $W(\mathbf{q})$. Other than this initial guess and the use of *hkl* space, no other crystallographic symmetry was assumed during the reconstruction.

5.5.4 Rotation group subset

In general, the set of views, Ω_j, was generated by sampling the 3D rotation group uniformly. This was done with the help of unit quaternions. However, in cases where the crystal is rotated about a single axis and the relative orientation of the axis with respect to the crystal basis vectors is known, one can sample angles about just that axis. This was done in this experiment, where the axis was determined from the high fluence dataset using the *XDS* software [124].

5.5.5 Results and conclusion

The crystal analyzed had a body-centered cubic (b.c.c.) unit cell with a lattice constant of 21.47Å in space group $I43d$. The unit-cell parameters were taken as a given and used to generate the mapping to *hkl* space (Section 5.5.2). The high fluence data set with known orientations was used to generate a reference 3D intensity model. This was compared with the reconstructions from different low flux data sets by comparing the Patterson maps, which were generated as follows. First, the intensities were integrated in a small sphere about every (*hkl*)

point. The 3D (hkl) grid of intensities was then inverse Fourier transformed to generate the electron-density auto-correlation function, which is the Patterson map. Figure 5.12 shows the comparison of the maps for one particular data set (48 photons-per-frame). For a quantitative comparison, we have calculated an R factor between the Fourier amplitudes of the low- and high-fluence reconstructions ($R_{high-low}$), shown in Figure 5.11. This quantity was calculated as

$$R_{high-low} = \frac{\sum |F_{hkl,high} - |F_{hkl,low}|}{\sum |F_{hkl,high}|}, \tag{5.5}$$

where $F_{hkl} = (I_{hkl})^{1/2}$ and I_{hkl} refers to the integrated intensity at Bragg peak (h, k, l). We observe that, at low resolutions, this R factor is near to 0.14. At high $q = (h^2 + k^2 + l^2)^{1/2}$, the lack of good statistics in the high-fluence data set leads to a large value. This is illustrated Figure 5.11 by the dashed line, which represents the mean number of photons per peak in the given resolution shell.

Dependence on photons per frame

As mentioned in Section 5.5.1, the crystal rotated 10^{-3} degrees over one frame. This allowed us to sum successive thresholded frames as they came from almost identical orientations. Using this method, we could study the effect of number of photons per frame on reconstruction quality while keeping other parameters the same. One effect of decreasing the number of photons per frame was that it took more iterations to reach convergence. For less than 48 photons per frame, the reconstruction did not converge at all. Above this threshold value, the reconstruction quality was independent of number of photons per frame. However, it took many more iterations, as can be seen in Figure 5.13 and Table 5.2. This is consistent with the observations in simulations with speckle intensity patterns [43]. The threshold value itself is lower because of the different distribution

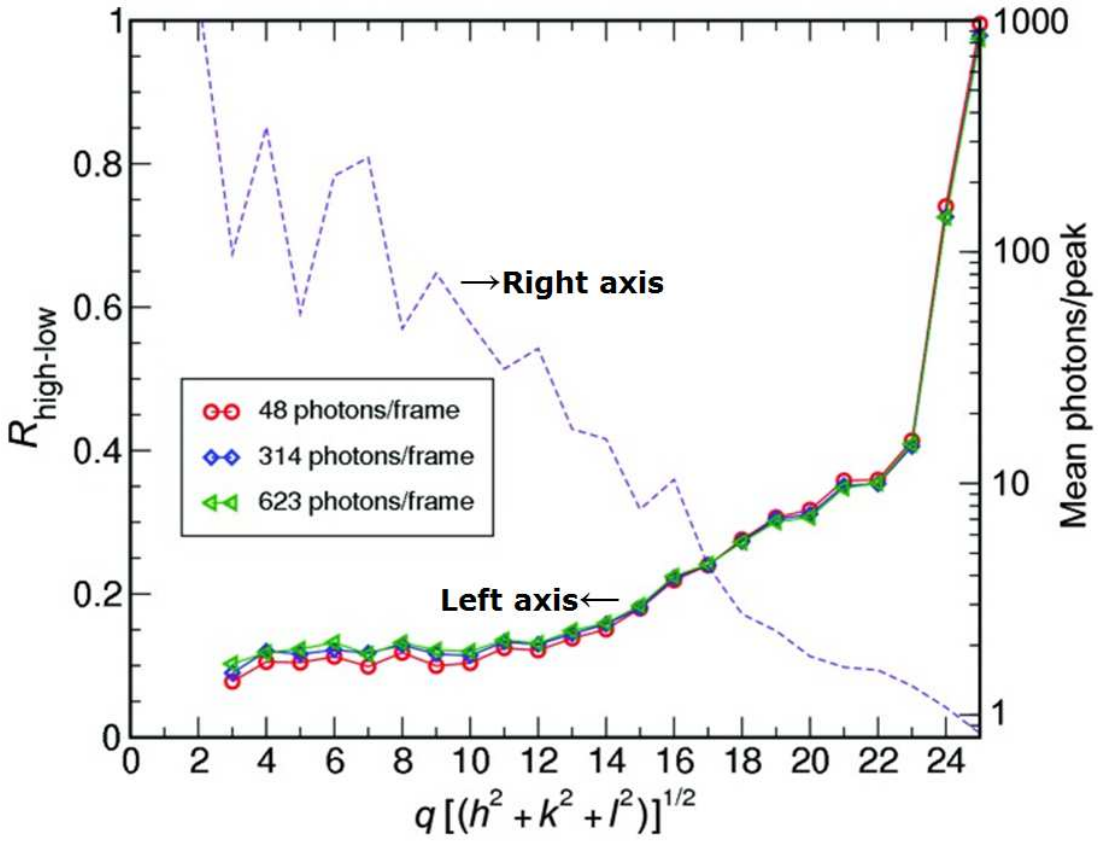


Figure 5.11: Plot showing $R_{high-low}$ as a function of spatial frequency q. This quantity is calculated by comparing a low-fluence reconstruction with the high-fluence assembly using the known orientations. The three most intense summed data sets mentioned in Table 5.2 have been used to plot. The dashed line represents the mean number of photons per Bragg peak as a function of q for the high-fluence data set.

of the intensity in this case (concentrated in Bragg peaks as opposed to large, smooth speckles). While the situation might be affected by other factors, such as beam monochromacity and crystal homogeneity, the rate of convergence will still be principally determined by the number of photons per frame.

In this study, we have shown that the 3D diffraction intensity distribution can be calculated from a large number of sparse data frames, each with unknown orientation. This result bodes well for the possibility of performing se-

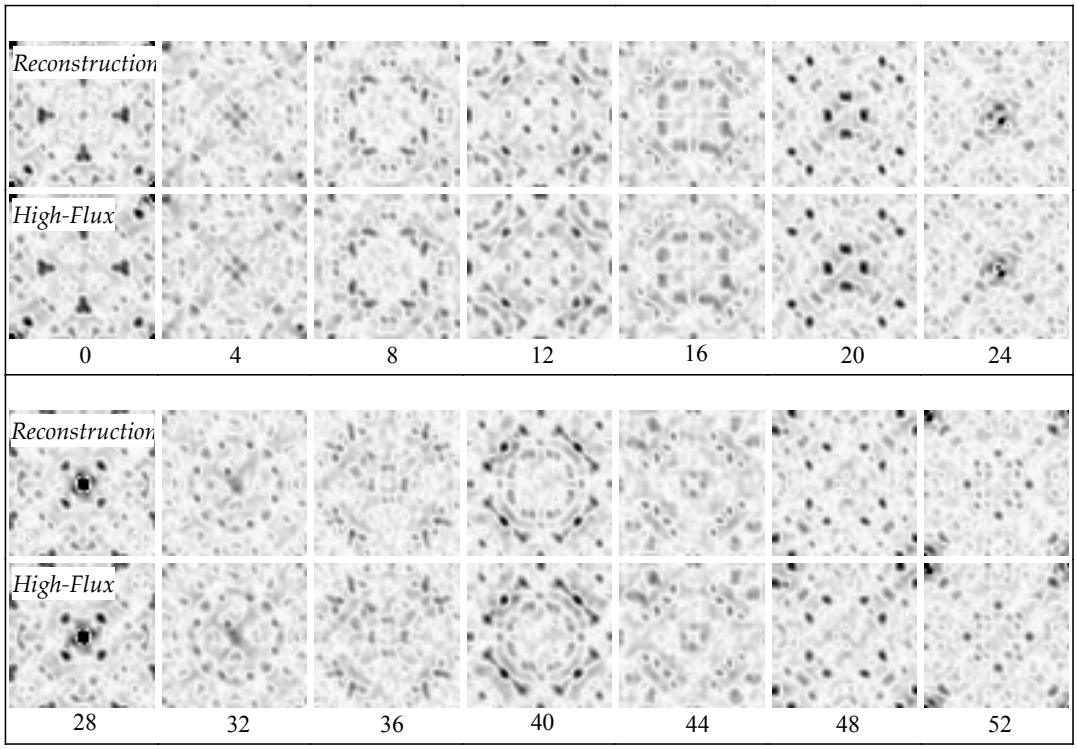


Figure 5.12: Comparison of slices through the 3D Patterson maps generated from the high-fluence data set of known orientation and a low-fluence (48 photons-per-frame) reconstruction. The map was $53 \times 53 \times 53$ voxels in size and every forth slice is shown here, with the slice number shown below each pair.

rial crystallography with micron-sized or smaller microcrystals at synchrotron sources. Complications will no doubt arise when the method is applied to protein crystals of decreasing size and an additional degree of rotational freedom, as will be seen in future studies. Still, it is promising that successful reconstruction of the 3D intensities has been shown for a signal level as low as 48 photons per frame.

One feature of the serial crystallography experiment not replicated here was the collection of data from all orientations in three dimensions. Reconstruction from the full rotation group was studied in simulations in Loh *et al.* [43] for ape-

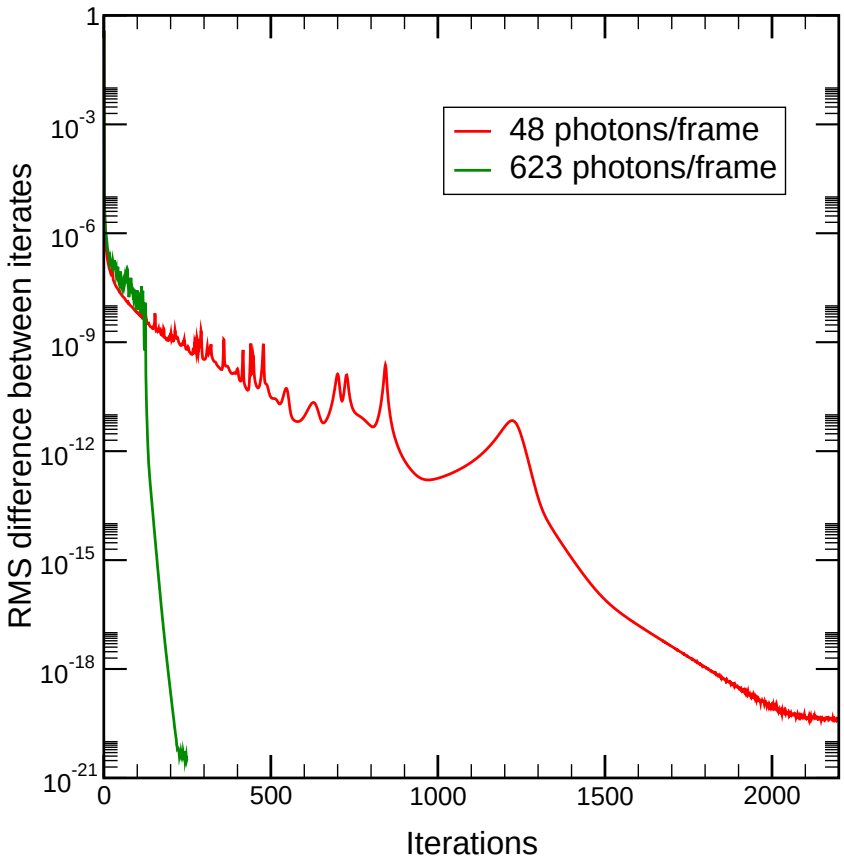


Figure 5.13: A plot showing the difference between successive iterates as a function of iteration number. The units on the logarithmic vertical are arbitrary; for reference, the lower limit corresponds to the floating-point precision of the computation. Two data sets are shown, with 15 summed frames (48photons-per-frame) and 200 summed frames (623photons-per-frame). The sparser data set takes much longer to converge and the slope of the curve in the last few iterations is strongly related to the number of signal photons per frame.

riodic structures with speckle intensity distributions. There, it was shown that the number of photons per frame required for successful reconstruction grows only logarithmically with number of orientational samples. Although, the total number of photons required for a complete data set with good signal-to-noise ratio and good resolution will be higher than what was collected here, with the fluence available at third-generation x-ray sources one should be able to collect a complete data set of similarly sparse nature with micron-sized crystals.

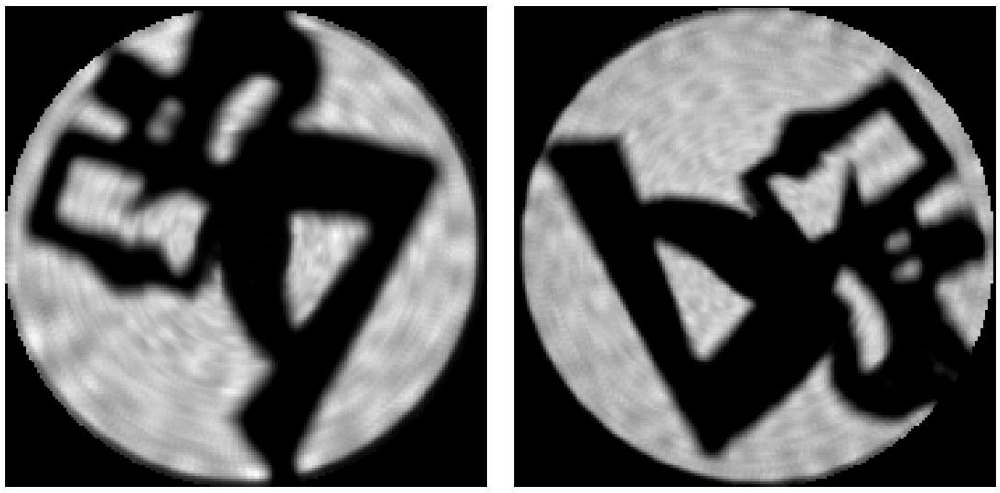
5.6 Effects of additional background and information reduction

Stepping back one moment, Chapter 3 provided motivation for reducing background in serial microcrystallography experiments. So far, this chapter has introduced three experiments with very little background contribution in comparison any experiment performed with a hydrated protein sample. To approach the reality of background scatter affecting reconstruction convergence, it is beneficial to explore the effect of adding background to two of the above studies.

5.6.1 Two dimensional shadow radiography

From the rotating mask reconstruction, imagine a situation where the beam path is exposed to air and gas molecules interact with the x-ray beam to provide incoherent, uncorrelated photons to the detector in the experimental setup in Section 5.3. Philipp *et al.* [112] simulated this by adding a uniform distribution of imitation photon counts to the data sets in a signal-to-noise ratio of 1. The process was repeated several times with similar results.

The additional noise caused a loss in resolution for the example reconstruction compared to the data sets without noise. Adding a signal-to-noise ratio of 1 to the 11.5 photons-per-frame data set generated reconstruction resolutions similar to that of a zero-background 3 photons-per-frame reconstruction. Figure 5.14 shows a comparison between the original zero-background 2.5 photons-per-frame reconstruction and the artificially noisy 23 photons-per-frame reconstruction. The 23 photons-per-frame dataset now resembles the original zero-



(a) (b)

Figure 5.14: Effect of background on 2D reconstruction quality. (a) 2.5 photons-per-frame data set reconstruction with no added background. This is also seen in Figure 5.4(d). (b) 11.5 photons-per-frame data set reconstruction with added signal-to-noise of 1. An average of 11.5 photons of Poisson distributed background were added per frame "by hand". The added noise also raised the platform background level, which was subtracted in this image to facilitate comparison to (a). The resolution of the reconstruction is similar between the two images, even though the original 11.5 photons-per-frame data contains more photons generated from the mask (Figure 5.4(c)). The new 23 photons-per-frame data is obscured by the background noise. Compare the "eyes" in (b) to the "eyes" in Figure 5.4(c) to see the degradation in image quality. Reproduced from Philip *et al.* [112].

background 2.5 photons-per-frame in both resolution and number of iterations to converge (as seen in Figure 5.14). Even though the 23 photons-per-frame dataset contains more orientation information of the mask on a per frame basis, the background signal degrades the reconstruction. The result is unsurprising, but reveals a real limitation for reconstructions in future studies in protein crystallography. This also provides substantial motivation to reduce any sources of background scatter to that inherent in the crystal itself, as mentioned in Chapter 3.

5.6.2 Inorganic crystallography

Returning to the crystallographic study, the large non-hydrated crystal provided relatively little background scattering compared to the Bragg spots. To study the effect of uniform background on the quality of the reconstruction, additional simulated photons counts were added, with a Poisson distribution of uniform mean, to each data frame of the 314 photons-per-frame data set. Except in the cases of extreme background, there is no effect on the orientation recovery. The weak highest-resolution peaks are lost as they are drowned out by the noise in the background. This is an unavoidable aspect of crystallography in the

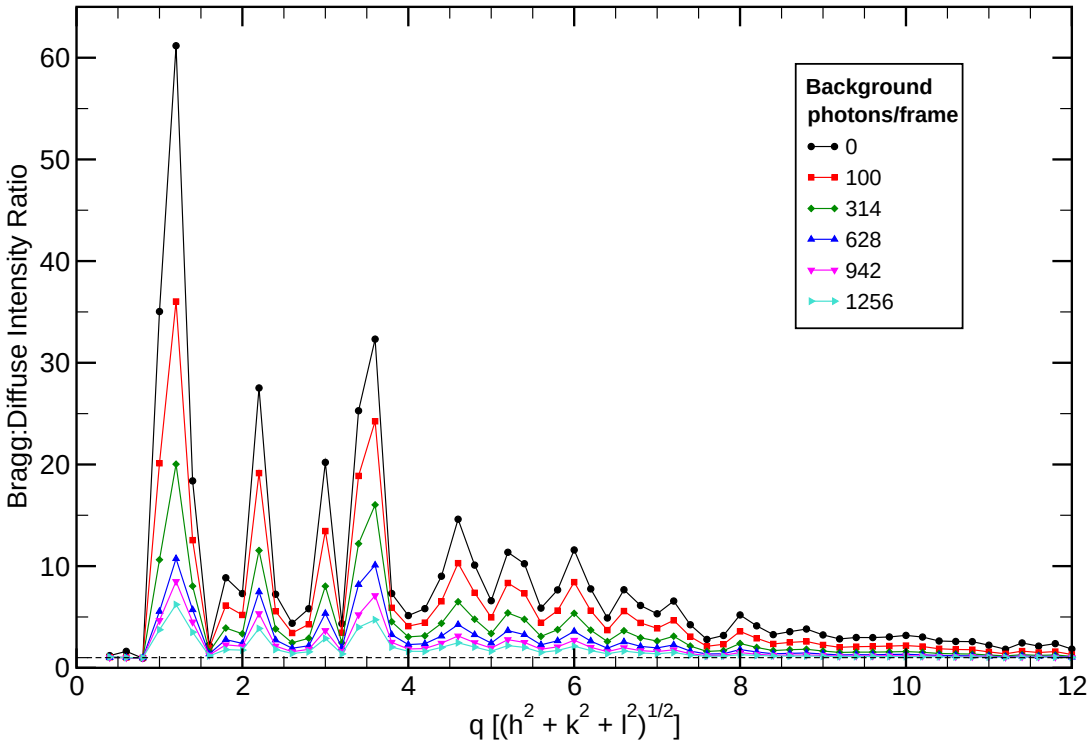


Figure 5.15: Plot of Bragg to diffuse intensity ratio vs. q for various amounts of additional background photons/frame. A high ratio indicates that the orientations have been correctly identified and most of the intensity is in Bragg peaks. There were 314 photons-per-frame in the base data set. Even with 400% background, the low resolution peaks could be resolved as seen the 1256 photons-per-frame plot.

presence of background.

To demonstrate the successful recovery of orientation at different background levels, the ratio of average intensity per voxel in the neighborhood of a Bragg point to the average intensity in the diffuse region is plotted versus reciprocal length, q, in Figure 5.15. If this ratio is close to 1, the Bragg peaks do not stand out over the background. As the plot shows, even with high background, the strong, low-resolution peaks are successfully recovered. However, as expected, the weak, high-resolution peaks are lost.

We observed that the addition of relatively high levels of uniform background (400%) does not affect orientation recovery. This is important, as some base level of background scattering is unavoidable with protein crystals due to at least the solvent internal to the crystal. As is the case in all protein crystallography experiments, the background reduces the resolution as higher order peaks are drowned out. Reducing the background requires minimizing the amount of material in the beam path. Fortunately, it is possible to reduce the background to insignificant levels by appropriate x-ray optics, vacuum paths and graphene windows surrounding the crystal stream as mentioned in Chapter 3.

5.7 Summary

This chapter, highlighted three proof-of-principle experiments with *EMC*. Each experiment showcased a reconstruction of original geometries from large data sets characterized by sparse data frames and unknown orientation. The exercise grew increasingly more complex with each study and revealed new challenges

for the *EMC* algorithm, including how the addition of background affects the quality of reconstruction. Most importantly, all three experiments suggest the feasibility of using the *EMC* algorithm for protein crystallography in the sparse data regime, which is featured in the next chapter, Protein Crystallography with *EMC*.

CHAPTER 6

PROTEIN CRYSTALLOGRAPHY WITH *EMC*

The contents of this section have been published in IUCrJ by Jennifer Wierman, Ti-Yen Lan, Mark Tate, Hugh Philipp, Veit Elser and Sol Gruner [17].

6.1 Introduction

As seen in previous chapters, the goal of serial microcrystallography is to acquire complete diffraction data sets by merging data from a succession of tiny crystals, the total volume of which is practically comparable to that of a single large crystal. Not only do smaller crystals yield fewer diffracted x-rays, but the number of diffracted x-rays in a given exposure from a microcrystal is limited by the dose that can be tolerated before classic radiation damage compromises the diffraction [30, 194]. Ultimately, a sufficiently small crystal size is reached that the number of photons diffracted per frame is too small to resolve Bragg peaks and the data frame is considered sparse. Intuitively, one might believe that sparse exposures can never be merged into complete data sets. However as shown in Chapter 5, this has already been shown not to be the case for non-protein structures [15, 44, 112]. In the following, it is demonstrated this is also the case for a protein crystal.

In this proof-of-principle study, we collected 8 million sparse frames from a 400 μm sized hen egg white lysozyme (HEWL) crystal rotating around a single axis with a relatively dim laboratory x-ray source and the fast-framing Mixed-Mode Pixel Array Detector (MM-PAD) [42] to simulate frames collected from

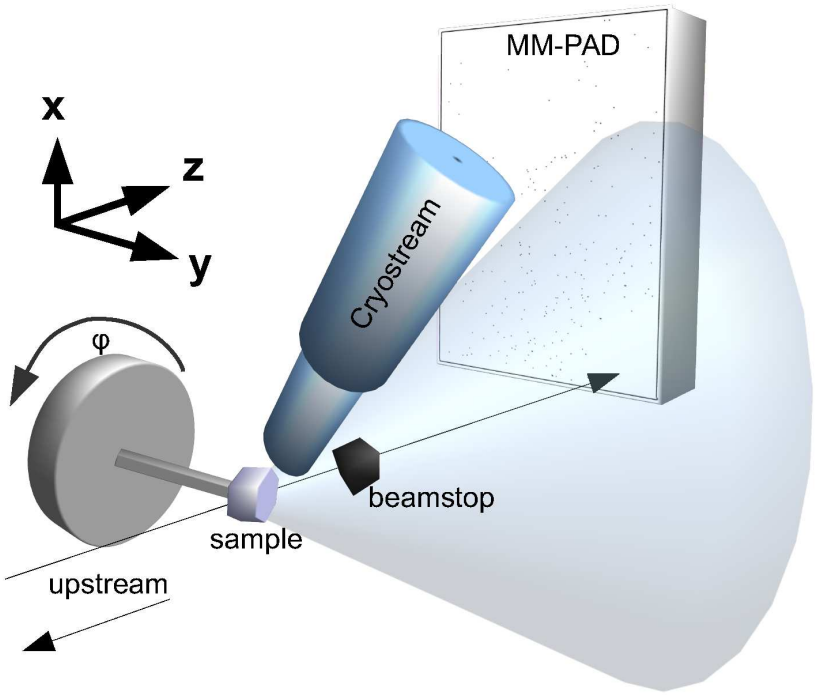


Figure 6.1: A simplified schematic of the experimental setup with the x-ray beam originating from the left side of the image along the z-axis. It illuminates the crystal rotating about y-axis (or $\hat{\phi}$), perpendicular to the beam axis. The main beam is then blocked by a beamstop. The diffracted photons are recorded with the MM-PAD. A cryostream, in blue, cools and maintains the crystal at 100K. The figure is not drawn to scale.

microcrystals at storage ring sources. Each frame consists of only ~200 photons on average (Figure 6.2), the majority of which were from background scatter. With only the prior knowledge of the unit cell parameters and rotation axis orientation, we successfully reconstructed the 3D Bragg intensities of the crystal. The algorithm made no assumptions on the crystal symmetry and was not given the angle of each frame about the rotation axis. The reconstructed intensities were of sufficient quality for a molecular replacement phasing algorithm to solve the structure to 1.5 Å resolution.

6.2 Experimental setup

As shown schematically in Figure 6.1, the physical setup of this study involved rotating a large protein crystal while continuously cooling to mitigate radiation damage. The axis of rotation was set to be perpendicular to the beam axis during data collection, The sample was illuminated by a Cu K_α x-ray beam (λ = 1.54 Å) generated from a Cu rotating anode set to 40 kV and 50 mA (Rigaku RU-H3R). The x-ray beam, with 10^7 photons/s, was monochromatized and focused to a ~0.5 mm × 0.5 mm spot at the sample using Ni-coated Franks mirrors placed 1 m from the sample. The beam had a divergence of 1 mrad. The center of the beam was placed in one corner of the MM-PAD's active area to record the highest possible resolution, which was approximately 1.3 Å. A pin-diode beamstop was used to keep the direct beam from striking the detector while recording the incident intensity.

6.2.1 Sample preparation

Lyophilized lysozyme powder from hen egg whites (Sigma, Saint Louis, Mo, USA) was used for crystallization by dissolving in deionized water at 50 mg/ml, without further purification. Crystals were grown at 293 K in 6 μL droplets by the hanging drop diffusion method with a 50% buffer solution of 1.0 M sodium chloride plus 0.1 M sodium acetate at pH 4.5 with 20% PEG. Crystals were then retrieved from the droplets after maximum growth after a few days with a Hampton Research Cryoloop. Crystals were then mounted on a goniometer and flash cooled under an Oxford Cryosystem Cryostream and kept at 100 K for data collection. By cryocooling a single macro-crystal we mimicked an experi-

ment with multiple micro-crystals that diffract identically.

6.2.2 Data collection

A single HEWL crystal approximately 400 μm sized was mounted on the goniometer and set continuously rotating on a rotation stage (Newport URS100) at 0.05° per second. Sparse data frames were ensured by simply reducing the exposure time per frame to a sufficiently short duration. The MM-PAD was at a distance of 33 mm from the rotating sample and recorded frames with a 10 ms exposure time, providing a 0.0005° oscillation angle per frame.

The data frames were then thresholded and photon counts were obtained using a procedure similar to that employed by Ayyer *et al.* [15, 44]. A dataset of 8.8 million frames, which corresponds to 12 full revolutions of the crystal, and with an average of ~200 photons per frame was then passed to the *EMC* algorithm. Although the orientation of each data frame was known, this information was not used by the *EMC* procedure.

6.3 Orientation recovery with *EMC*

The *EMC* algorithm developed by Loh *et al.* [43] and described in Chapter 5 was used to iteratively assemble the non-oriented, shot-noise limited frames into a 3D intensity map.

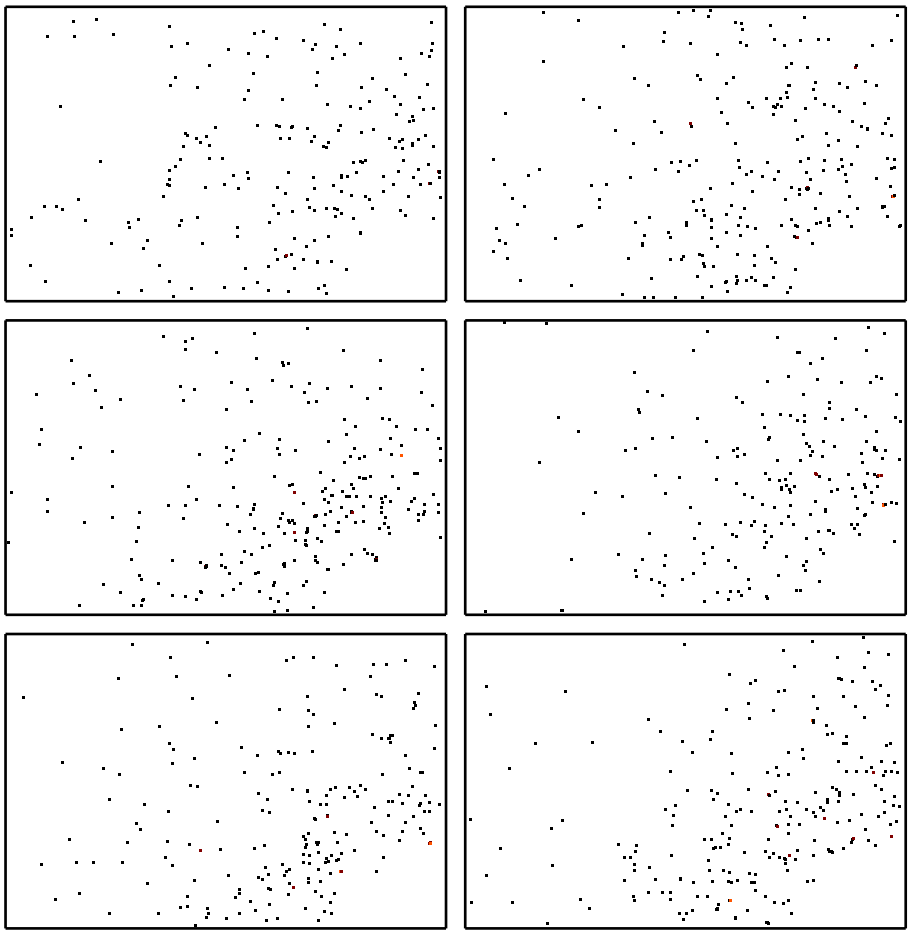


Figure 6.2: Random selection of six data frames (393 × 262 pixels). The direct beam is incident normally at the lower right region of the detector, which is blocked by the beamstop. Resolution at the upper left corner is 1.3 Å. Each frame consists of only ~200 photons on average, and the maximal photon count per pixel in these frames is 3. The size of pixels is smaller than the rendered photons in this image, which are enlarged for visual clarity.

6.3.1 Initial guess

To test the robustness of the *EMC* algorithm, we assumed that the parameters of the tetragonal unit cell were only known roughly, as might be the case, for example, from a diffraction powder pattern. The initial intensity estimate was seeded by placing small 3D Gaussian peaks of random height at each predicted

Bragg position. No symmetry such as Friedel pairs or systematic absences was imposed in this process.

6.3.2 Rotation group subset

Similarly to Section 5.5, the experimental setup only allowed orientation sampling within a small rotation subspace (*e.g.*, one rotational axis). One can expect difficulty in searching for a solution within the whole rotation space, unless the constraint imposed by the measurement is strong, which is not the case in the sparse regime. Therefore, we confined ourselves to a uniform distribution of 1D rotations about the rotation axis in this study. We note that crystals generally will have random orientations over all 3D rotations in serial crystallography. This broader rotation angle space will be explored in future studies. Since frames were taken sequentially while rotating, we merged the first revolution into bins of width one degree to retrieve the rotation axis orientation with the *XDS* package [124].

6.4 Crystallographic-*EMC* integration

The *EMC* algorithm reconstructs the total scattered intensity, including the diffuse background scatter which should be subtracted from the Bragg peak intensities. In addition, the Bragg peaks do not necessarily fall perfectly onto any *a priori* lattice. To determine precise values of the unit cell parameters, a 3D version of the peak segmentation algorithm described in Zhang *et al.* [116] was used. The algorithm proceeds for several iterations, and each iteration refines

the segmentation from the previous iteration. The segmentation is a classification of voxels into signal or background based on a standard score. The standard score $z(W)$ of a voxel with intensity value W is computed as

$$z(W) = \frac{W - \mu}{\sigma}, \tag{6.1}$$

where μ and σ are respectively the mean and standard deviation of the voxels in a surrounding $n \times n \times n$ voxel cube. Voxels with standard score above a particular threshold γ are classified as signal. This procedure is repeated three more times with the difference that the μ and σ computation only includes the voxels classified as background in the previous iteration. For good-quality segmentation of the Bragg peaks, we increased γ from 1.0 to 3.0 in successive iterations. For a candidate set of unit cell parameters, we computed the total intensity of segmented peaks lying within ellipsoids centered on the corresponding Bragg positions. The ellipsoid volume was a small fraction of the unit cell, with principal axes consistent with the tetragonal cell. The unit cell parameters giving the greatest total intensity were taken as the refined values.

Using the refined unit cell parameters, we determined the Bragg peak intensities using the following integration procedure: The voxels closest to any particular Bragg peak position were assigned to be background or signal based on whether they were within an ellipsoid centered on the peak. These ellipsoids were similar to those used in parameter refinement but larger. The mean of the background voxels were then subtracted from each signal voxel before being summed to give the intensity for each reflection. Partial peaks, such as those adjacent to boundary, detector gaps, or the beamstop region, were rejected.

Reconstruction		**Refinement**	
Space Group	$P4_32_12$	# of atoms	1963
Unit Cell Parameters (Å)	a,b = 77.5, c = 36.2	R factor	0.2823
Resolution (Å)	54.801-1.497	R_{free}	0.3199
Completeness (%)	92.01	Rms bond length (Å)	0.0192
# of reflections	16,056	Rms bond angle (°)	0.1200

Table 6.1: Refinement statistics for lysozyme with intensities from the *EMC* algorithm.

6.5 Model building and refinement

The reconstructed intensities and subsequent structure factors were fed into *MOLREP* [134] from the *CCP4* suite to produce a molecular replacement solution using several published coordinates (lysozyme entries 193L, 1FLQ and 2LZM) from the PDB as a starting model. The solutions were refined through 20 iterations in *REFMAC* [134] with both rigid-body and restrained refinement, and rebuilt in *Coot* [135], with cyclical refinement. Refinement statistics are shown for 193L in Table 6.1, as the final molecular replacement solution used 193L as the model for phasing. As a negative control, we used lysozyme entry 1FLQ, a mutant of lysozyme with all alanines substituted for glycine. While *MOLREP* provided a solution for our reconstructed data with phases from 1FLQ, the refined map is much less ordered and fits poorly with the original model of lysozyme. As another negative control, we used lysozyme entry 2LZM, a human lysozyme with a similar structure yet 40 more residues than HEWL, for molecular replacement. Here, *MOLREP* did not provide a phased solution for our reconstruction.

6.6 Results and conclusions

6.6.1 Validation of reconstruction

As a check, the reconstructed intensity distribution in reciprocal space was compared to the actual intensity distribution. The actual (*i.e.*, "reference") distribution could be recovered because the orientation of each frame was known, even though this information was not used in the *EMC* reconstruction. Several slices

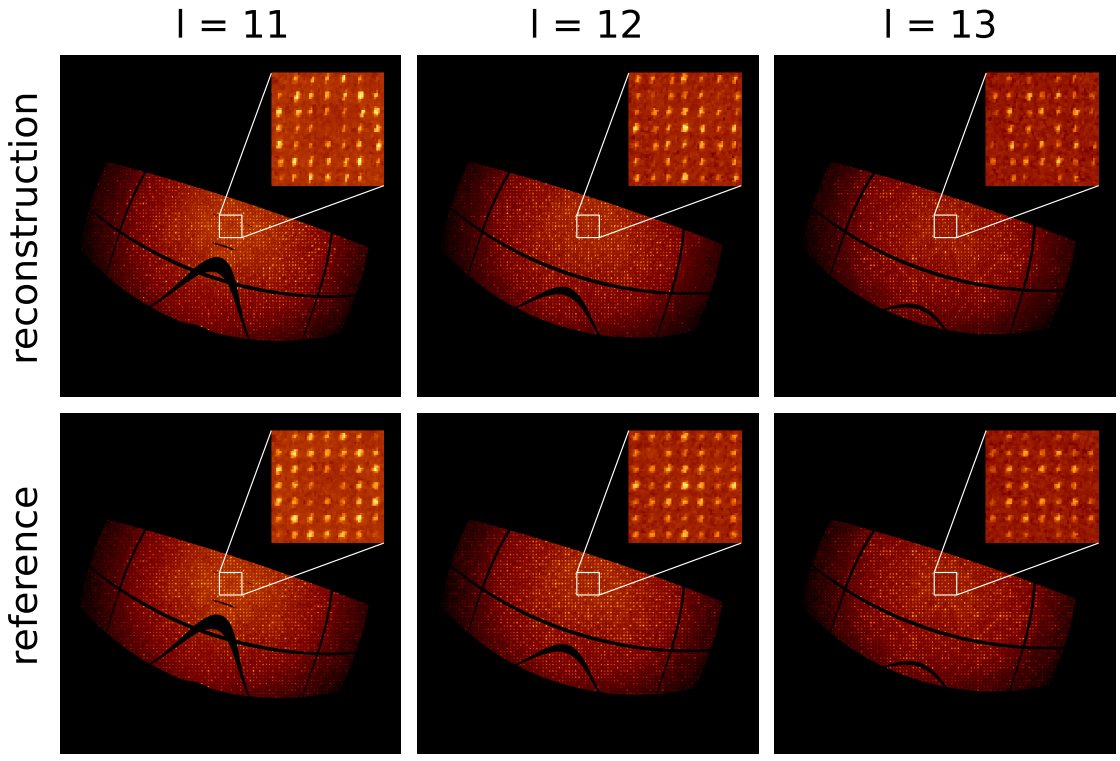


Figure 6.3: Slices of the reconstructed and reference intensity maps in the hk plane at constant values of l. Even without imposing symmetry when seeding the initial intensity estimate, the reconstructed intensity obeys the reflection condition $00l : l = 4n$ required by the space group symmetry $P4_32_12$ of the HEWL crystal (see insets). The mapping into reciprocal space transforms the detector gaps [42] into curves, shown as black lines and arcs.

of the reconstructed intensity and reference intensity perpendicular to the l-axis are shown in Figure 6.3. We checked that the reconstructed intensity obeys the reflection conditions $00l : l = 4n$ and $h00 : h = 2n$ required by the space group symmetry $P4_32_12$ of the HEWL crystal [195]. This suggests a successful orientation recovery because no symmetry was imposed when we seeded the initial intensity estimate.

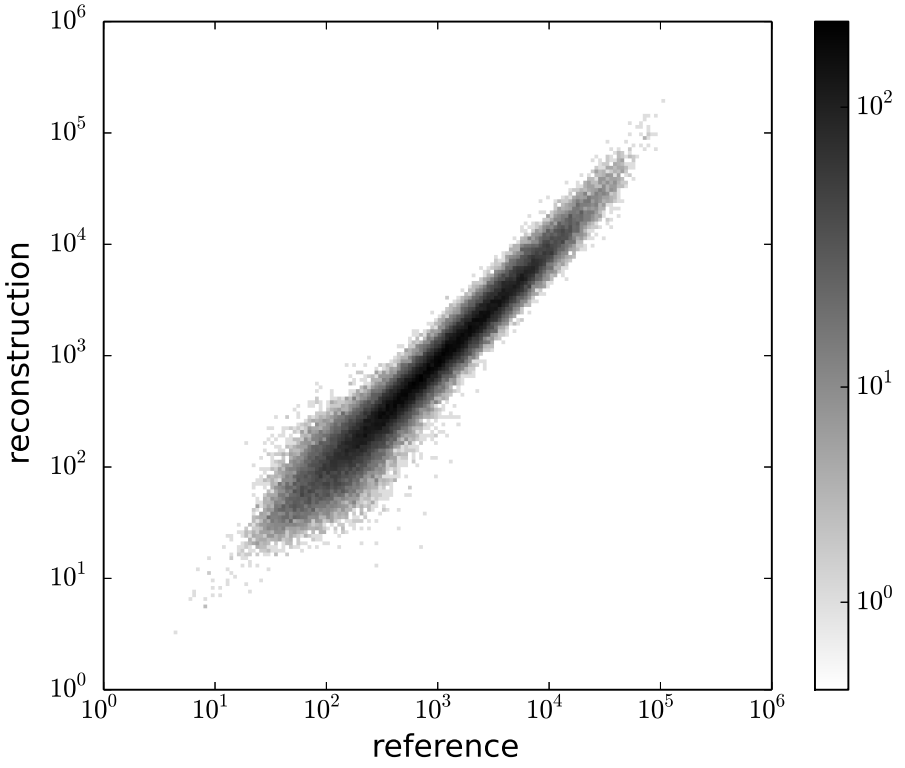


Figure 6.4: Scatter plot comparing the reconstructed Bragg peak intensities with the reference intensities. The reflections collapse well onto the diagonal, which indicates an accurate reconstruction of Bragg peak intensities.

A more direct justification involves comparing the integrated reflections. Figure 6.4 shows the scatter plot comparing the reconstructed intensities with the reference intensities. The reflections collapse well onto the diagonal, which indicates that orientations of most frames were recovered by the *EMC* algo-

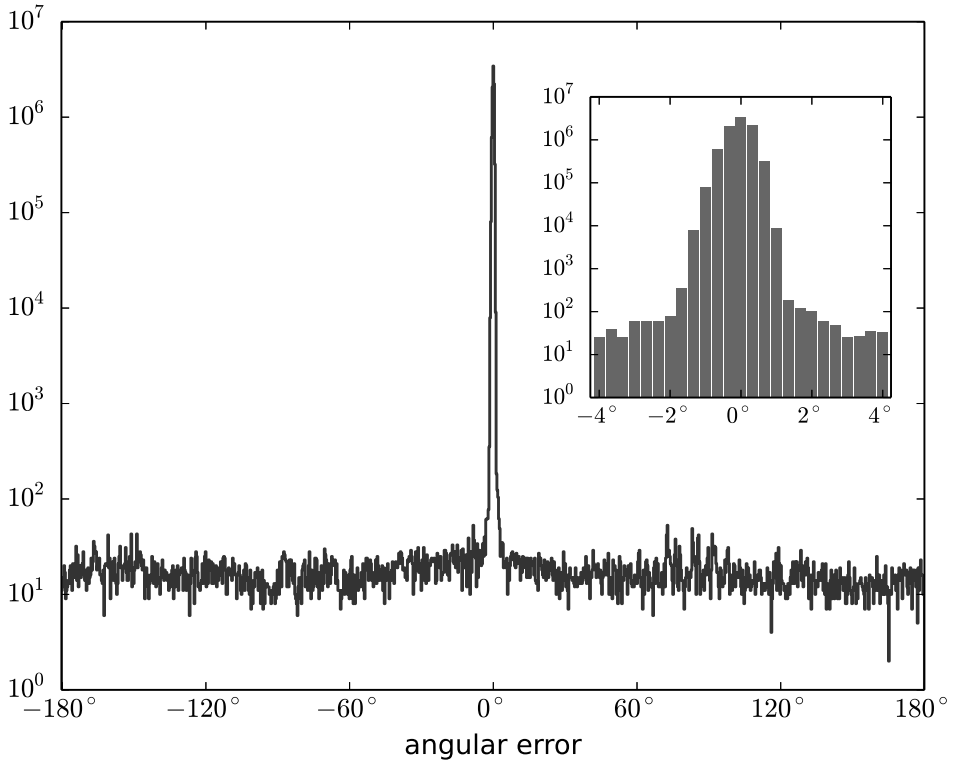


Figure 6.5: Histogram of the difference between most probable orientations of frames and actual orientations, expressed in degrees about the rotation axis. The *EMC* algorithm correctly assigned 99.7% of the frames within 1°, as shown in the inset. This validates proper orientation assignment from the reconstruction.

rithm. We expect that the distribution of reflections in the scatter plot becomes broader as the average photon count per frame decreases, because this reduces the information for orientation recovery.

The difference between the most probable orientation of each frame assigned by the *EMC* algorithm and its actual orientation is shown in Figure 6.5 as a histogram of 1D rotations about the rotation axis. We found that 99.7% of the frames were assigned to the correct orientation within 1°. We suspect that the outliers are due to abnormally low signal-to-noise ratio in some frames, perhaps caused by extra background scatter from the cryoloop or an orientation with

few reflections. This motivates the necessity of background reduction in future experiments, specifically in the case of small or weakly diffracting crystals.

6.6.2 Validation of structure

The structure we built from the *EMC* reconstructed intensities (Figure 6.6) agrees with the published structure of lysozyme, 193L. The rms difference when all of the C_α atoms of the two structures are superimposed is 0.27 Å, which could be attributed to differing solvent content during crystallization and water placement during refinement between the deposited model and our initial crystal. With a completeness of 92.01%, 16,056 independent reflections, an R factor of 0.28 and R_{free} equal to 0.32, our structure determined from reconstructed sparse data compares favorably with structures obtained by more conventional means.

6.6.3 Validation of sparse data

From the reconstructed intensity map, we were able to identify photons which contributed to Bragg peaks within each frame. This allowed us to determine that on average as many as 80% of the counts were background photons scattered from the main x-ray beam hitting air, solution around and within the crystal, upstream optics, *etc.* Looking back at Figure 6.2, it is impossible to tell which 20% of the photons belong to the Bragg peaks. This lack of sensitivity to background is special to crystal data sets and consistent with the findings of Ayyer *et al.* [44].

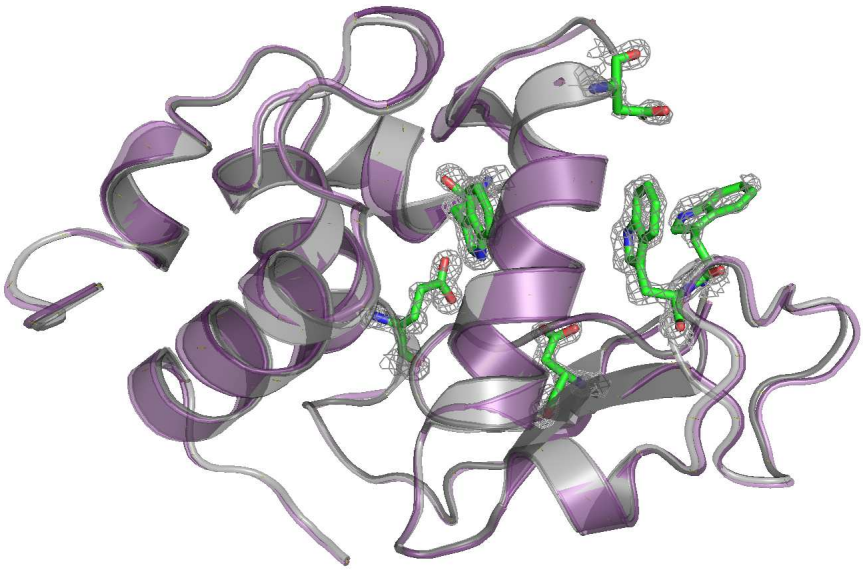


Figure 6.6: Structure of the reconstructed protein (gray) compared with the model 193L (purple) used in molecular replacement. Comparison of higher resolution features (active sites) are rendered in green sticks (model structure) and gray mesh (reconstruction). When all of the C_α atoms of the two structures are superimposed, the rms difference is only 0.27 Å. This could arise from differing solvent content during crystallization and water placement during refinement between the deposited model and our initial crystal.

6.6.4 Computational details

We performed the reconstruction on a single machine (Intel Xeon E5-2640 at 2.00 GHz with 128 GB RAM running Scientific Linux) using 16 cores. The estimates of unit cell parameters were $a = b = 77.0$ Å, $c = 36.0$ Å, and the reconstruction used data up to a resolution of 2.0 Å, with only 195 photons per frame on average. We used a reciprocal lattice grid with voxel size $a^*/7$, which corresponds to $543 \times 543 \times 543$ voxels. The sampled rotations consisted of 1080 uniformly distributed rotations about the rotation axis. The reconstruction ran for 30 iterations, and each iteration took 1.3 hours on average. Convergence was monitored by the rms change of the 3D intensities, which was found to be insensitive

to the choice of random seeds for the initial intensities. Based on the converged intensity at 2.0 Å resolution, the probability distribution $P_{jk}(W)$ was calculated and fed into Equation 5.2 to incorporate data up to resolution 1.3 Å. For this we used a finer reciprocal grid of size $a^*/9$ ($939 \times 939 \times 939$ voxels), to mitigate peak overlaps. The resulting intensity was rescaled so that its sum equals the total number of recorded photons over all the frames; this is what we call the reconstructed intensity. $n = 15$ was used for the size of the cubic window in peak segmentation, as described in Section 6.4. The Bragg peak intensities were integrated using the refined unit cell parameters $a = b = 77.52$ Å, $c = 36.23$ Å.

6.7 Summary

It was shown experimentally that a series of non-oriented, sparse diffraction frames from a protein crystal rotating about a single rotation axis can be assembled into a 3D intensity with the aid of the *EMC* algorithm. Validation of reconstruction is supported by the recovery of symmetries which were absent in the initial seeding process, the consistency of integrated reflections with the reference intensity and the comparison of most probable orientations of frames with actual orientations. Moreover, we have demonstrated that the protein structure can be solved by phasing the integrated reflections of the reconstruction through molecular replacement. This result suggests that the indexability of each frame *per se* does not necessarily limit the structure determination in serial crystallography.

In fact, this study may relax many limitations in serial crystallography imposed by indexability of frames: *i.e.*, size of the crystal, brilliance of x-ray source,

or radiation sensitivity. With minor modifications, one can envision a serial microcrystallography experiment performed at room temperature at storage ring sources within microfluidic chips [79], or from gel injectors [8, 53] from crystals smaller than currently being investigated. Several features are still needed to make the experiment more realistic for serial crystallography. One is the sampling of the entire rotation group, in which the constraint for solution convergence shall be stronger because of the larger redundancy among frames. Another is the need to determine that many crystals of varied diffraction quality can be merged using *EMC*. The computation time, which scales with the product of the number of rotations and the number of frames, is expected to grow rapidly at the same time, so further optimizations are necessary. Also, background reduction, such as the usage of a graphene window [139], is desirable when taking data from multiple small crystals. In the next chapter, proof-of-principle experiments are outlined where the entire rotation group is sampled, and data is collected from multiple crystals.

CHAPTER 7

SUMMARY AND FUTURE DIRECTIONS

It is well known that many biologically relevant proteins do not readily form large enough crystals for traditional crystallography. The advent of SFX at XFELs inspired new techniques for studying smaller crystals than ever before, but as XFELs are limited in experimental availability, one looks to other sources for the development of serial microcrystallography. This dissertation has explored the technology available to perform room temperature serial microcrystallography at storage ring sources. SMX should be feasible at storage ring sources. This is true even from crystals that are so small that single indexable exposures cannot be obtained, provided further development of the *EMC* algorithm, optimized delivery systems, and improved background reduction at beamlines.

7.1 Requirements for serial microcrystallography

Compounded by the size of microcrystals, the limitations on serial microcrystallography at room temperature include radiation damage imposing short crystal diffraction lifetimes, and extraneous background scatter obscuring the already weak microcrystal diffraction signal. From these limitations, new methods and techniques are absolutely necessary to obtain structures if serial microcrystallography is to be a viable approach at storage ring sources. There are three areas which require rethinking and optimization when compared to what has been achieved so far in microcrystallography: beamline station improvements, sample chamber improvements and software algorithm integration.

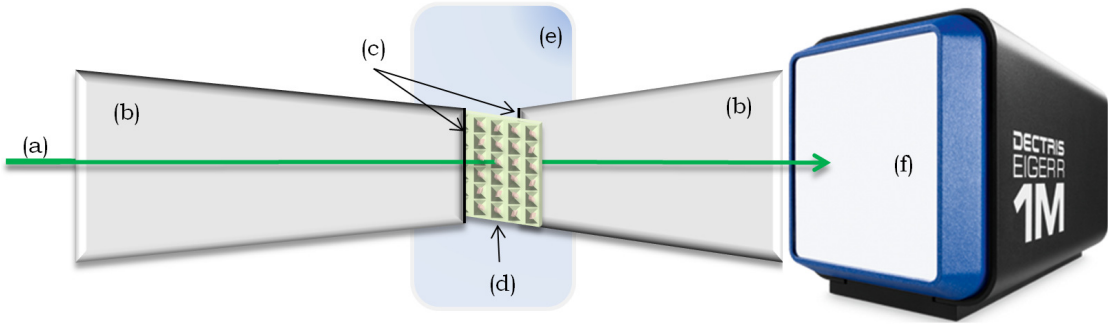


Figure 7.1: Schematic of serial microcrystallography with a microfluidic chip. An intense, collimated x-ray microbeam (a) is held in vacuum (b) to reduce excess air scatter. Ultra thin windows (c), perhaps of graphene, separate the vacuum beam path from microcrystal delivery system (d), such as the microfluidic chip from Oghbaey *et al.* [100], inside an evacuated sample chamber (e). An EIGER detector (f) collects the diffraction data.

7.1.1 Beamline improvements

Serial microcrystallography requires an intense, collimated microbeam free from any scatter produced by optics, edges, windows or slits. The consummate beamline should be completely in vacuum from the source to the detector, and therefore have zero background. Although this is ideal, it is unreasonable for a room temperature biological system like a protein crystal to survive in vacuum. Therefore, we require some sort of partition with x-ray transparent windows of the sample from the beam path (seen in Figure 7.1). The feasibility of using atomically-thin sheets of graphene to reduce background scatter as a crystal mount was shown, and as they have been shown to be gas-tight [150, 151], graphene windows show promise as an attractive option for window material. Ultra-thin silicon nitride membranes could also function as a separation as long as they are fractions of the width of the microcrystal. Otherwise, there should be little to no material other than the crystal in the beam path.

Some beamlines are beginning to upgrade to systems with less background,

such as P11 at PETRA III [60], BL32XU at SPring-8 [154], ID13 [196] and ID23 at the ESRF [197], but much remains to be done in order to have effectively zero background at any beamline.

7.1.2 Sample chamber improvements

The sample chamber does not necessarily require a large chamber, so much as a method to seal the microcrystal from vacuum. In fact, the less material, the more distinct from background the signal is from the crystal. The entire sample environment should be optimized to isolate the crystal diffraction. Characteristics of the ideal delivery device would include incorporating extremely thin substrates for sample windows, such as graphene, and optimizing the hit rate through concentrating the crystals in addressable locations (as seen in Figure 7.1).

This dissertation explored two microcrystallography devices: a microfluidic chip and a viscous jet injector. Both of these devices contribute more background scatter than is acceptable for the ideal serial microcrystallography experiment. With the viscous jet injector described in Section 4.3, the present limitation on the column diameter, a ~5:1 ratio, to avoid sample blockage sets a higher threshold to background than desired, as the gel will contribute approximately 10 times more background scatter than the crystal. Further, stable columns with diameters on the order of a micron have yet to be demonstrated. The microfluidic chips described in Chapters 2 and 4 contain window materials far too thick for microcrystallography. No one has succeeded so far in creating a sample environment where a microcrystal is the primary source of background

scatter. Sui *et al.* [78] and Oghbaey *et al.* [100] have microfluidic devices that may better approach the ideal device, though it is yet to be determined if this is enough for crystals smaller than a few microns.

7.1.3 Software algorithm improvements

The final improvement needed for serial microcrystallography is diffraction data processing from microcrystals. Already, programs have been developed which deal with large numbers of single-crystal images, such as *Cheetah* [122], and *CrystFEL* [129], as long as every individual image can be indexed. From micron-sized or smaller crystals at room temperature, there may not be enough diffraction to determine its orientation through indexing Bragg peaks. When microcrystal diffraction is weak enough that Bragg peaks are no longer visible, a relaxation to the requirement of indexing individual data frames is suggested in order for a structure determination. The major focus of this dissertation has been proof-of-principle experiments which show that indexing data frames on a per-frame basis is not necessary for a structure solution in protein crystallography when reciprocal space intensities can be reconstructed using the *EMC* algorithm.

The *EMC* algorithm has much potential to revolutionize serial microcrystallography at storage ring sources. While large fractions of microcrystal diffraction data are currently thrown out as too "sparse" [11, 33], the *EMC* algorithm might be able to use the sparse data to reconstruct the 3D intensity distribution, given enough diffraction images of microcrystals in random orientations.

7.2 The future

What has yet to be done is to reconstruct a protein structure from a full rotation data set. Next attempts with relatively large crystals will be with the Miller group from University of Toronto, where we plan to reconstruct the structure of a significant protein using their ultra-thin microfluidic chip [100] at CHESS beamline G3. In the meantime, G3 will be optimized for serial microcrystallography using microfocusing monocapillaries to create an intense microbeam and building sample chambers with ultra-thin windows.

There is also a need to show that a computationally practical *EMC* approach can be devised to solve structures of a series of sparse-data crystals in arbitrary orientations. This work, too, is ongoing within the Gruner and Elser research groups.

In conclusion, serial microcrystallography at storage ring sources is advancing very rapidly. There are many technical problems that will have to be overcome but, as far as we can tell, there do not appear to be "show-stoppers".

BIBLIOGRAPHY

[1] M. S. Hunter and P. Fromme, "Toward structure determination using membrane–protein nanocrystals and microcrystals," *Methods*, vol. 55, no. 4, pp. 387–404, 2011.

[2] S. Boutet, L. Lomb, G. J. Williams, T. R. Barends, A. Aquila, R. B. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman *et al.*, "High-resolution protein structure determination by serial femtosecond crystallography," *Science*, vol. 337, no. 6092, pp. 362–364, 2012.

[3] J. L. Smith, R. F. Fischetti, and M. Yamamoto, "Micro-crystallography comes of age," *Current Opinion in Structural Biology*, vol. 22, pp. 602–612, 2012.

[4] I. Nederlof, Y. W. Li, M. van Heel, and J. P. Abrahams, "Imaging protein three-dimensional nanocrystals with cryo-EM," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 5, pp. 852–859, 2013.

[5] I. Nederlof, E. van Genderen, Y.-W. Li, and J. P. Abrahams, "A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 7, pp. 1223–1230, 2013.

[6] M. S. Hunter, B. Segelke, M. Messerschmidt, G. J. Williams, N. A. Zatsepin, A. Barty, W. H. Benner, D. B. Carlson, M. Coleman, A. Graf *et al.*, "Fixed-target protein serial microcrystallography with an x-ray free electron laser," *Scientific Reports*, vol. 4, 2014.

[7] C. Kupitz, S. Basu, I. Grotjohann, R. Fromme, N. A. Zatsepin, K. N. Rendek, M. S. Hunter, R. L. Shoeman, T. A. White, D. Wang *et al.*, "Serial time-resolved crystallography of photosystem II using a femtosecond x-ray laser," *Nature*, vol. 513, no. 7517, pp. 261–265, 2014.

[8] U. Weierstall, D. James, C. Wang, T. A. White, D. Wang, W. Liu, J. C. Spence, R. B. Doak, G. Nelson, P. Fromme *et al.*, "Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography," *Nature Communications*, vol. 5, 2014.

[9] S. Quevillon-Cheruel, D. Liger, N. Leulliot, M. Graille, A. Poupon, I. L. De la Sierra-Gallay, C.-Z. Zhou, B. Collinet, J. Janin, and H. Van Tilbeurgh,

"The Paris-Sud yeast structural genomics pilot-project: from structure to function," *Biochimie*, vol. 86, no. 9, pp. 617–623, 2004.

[10] L. C. Johansson, D. Arnlund, T. A. White, G. Katona, D. P. DePonte, U. Weierstall, R. B. Doak, R. L. Shoeman, L. Lomb, E. Malmerberg *et al.*, "Lipidic phase membrane protein serial femtosecond crystallography," *Nature Methods*, vol. 9, no. 3, pp. 263–265, 2012.

[11] C. Gati, G. Bourenkov, M. Klinge, D. Rehders, F. Stellato, D. Oberthür, O. Yefanov, B. P. Sommer, S. Mogk, M. Duszenko *et al.*, "Serial crystallography on in vivo grown microcrystals using synchrotron radiation," *IUCrJ*, vol. 1, no. 2, pp. 87–94, 2014.

[12] Chapman, H. N., et al., "Femtosecond X-ray protein nanocrystallography," *Nature*, vol. 470, pp. 73–78, 2011.

[13] J. M. Holton and K. A. Frankel, "The minimum crystal size needed for a complete diffraction data set," *Acta Crystallographica Section D*, vol. 66, pp. 393–408, 2010.

[14] S. M. Gruner and E. E. Lattman, "Biostructural Science Inspired by Next-Generation X-Ray Sources," *Annu Rev Biophys*, no. February, pp. 1–19, 2015.

[15] K. Ayyer, H. T. Philipp, M. W. Tate, V. Elser, and S. M. Gruner, "Real-space x-ray tomographic reconstruction of randomly oriented objects with sparse data frames," *Opt. Express*, vol. 22, no. 3, pp. 2403–2413, Feb 2014.

[16] K. S. Ayyer, "Reconstructing Images From Sparse Data," Ph.D. dissertation, Cornell University, 2014.

[17] J. L. Wierman, T. Y. Lan, M. W. Tate, H. T. Philipp, V. Elser, and S. M. Gruner, "Protein crystal structure from non-oriented, single-axis sparse X-ray data," *IUCrJ*, vol. 3, pp. 43–50, 2016.

[18] M. Warkentin, R. Badeau, J. Hopkins, and R. E. Thorne, "Dark progression reveals slow timescales for radiation damage between t= 180 and 240 K," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 9, pp. 792–803, 2011.

[19] R. B. G. Ravelli and E. F. Garman, "Radiation damage in macromolecular

crystallography," *Current Opinion in Structural Biology*, vol. 16, pp. 624–629, 2006.

[20] J. M. Holton, "A beginner's guide to radiation damage," *Journal of Synchrotron Radiation*, vol. 16, pp. 133–142, 2010.

[21] E. Garman, "Cool data: quantity AND quality," *Acta Crystallographica Section D: Biological Crystallography*, vol. 55, no. 10, pp. 1641–1653, 1999.

[22] ——, "Cool crystals: macromolecular cryocrystallography and radiation damage," *Current Opinion in Structural Biology*, vol. 13, no. 5, pp. 545–551, 2003.

[23] M. Warkentin, V. Stanislavskaia, K. Hammes, and R. E. Thorne, "Cryocrystallography in capillaries: critical glycerol concentrations and cooling rates," *Journal of Applied Crystallography*, vol. 41, no. 4, pp. 791–797, 2008.

[24] D. A. Keedy, L. R. Kenner, M. Warkentin, R. Woldeyes, M. C. Thompson, A. S. Brewster, A. H. Van Benschoten, E. L. Baxter, J. B. Hopkins, M. Uervirojnangkoorn, S. E. McPhillps, J. Song, R. Alonso-Mori, J. M. Holton, W. I. Weis, A. T. Brunger, S. M. Soltis, H. Lemke, A. Gonzalez, N. K. Sauter, A. E. Cohen, H. van den Bedem, R. E. Thorne, and J. Fraser, "Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography," *eLife*, vol. 4, p. e07574, 2015.

[25] J. S. Fraser, H. van den Bedem, A. J. Samelson, P. T. Lang, J. M. Holton, N. Echols, and T. Alber, "Accessing protein conformational ensembles using room-temperature x-ray crystallography," *Proceedings of the National Academy of Sciences*, vol. 108, no. 39, pp. 16 247–16 252, 2011.

[26] R. J. Southworth-Davies, M. A. Medina, I. Carmichael, and E. F. Garman, "Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography," *Structure*, vol. 15, pp. 1531–1541, 2007.

[27] C. G. Darwin, "Xxxiv. the theory of x-ray reflexion," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 27, no. 158, pp. 315–333, 1914.

[28] J. Drenth, *Principles of Protein X-Ray Crystallography*. New York: Springer, 1999.

[29] E. H. Snell and J. R. Helliwell, "Macromolecular crystallization in microgravity," *Reports on Progress in Physics*, vol. 68, no. 4, p. 799, 2005.

[30] C. Nave and E. F. Garman, "Towards an understanding of radiation damage in cryocooled macromolecular crystals," *Journal of Synchrotron Radiation*, vol. 12, no. 3, pp. 257–260, 2005.

[31] R. Sanishvili, D. W. Yoder, S. B. Pothineni, G. Rosenbaum, S. Xu, S. Vogt, S. Stepanov, O. A. Makarov, S. Corcoran, R. Benn *et al.*, "Radiation damage in protein crystals is reduced with a micron-sized x-ray beam," *Proceedings of the National Academy of Sciences*, vol. 108, no. 15, pp. 6127–6132, 2011.

[32] P. Sliz, S. C. Harrison, and G. Rosenbaum, "How does radiation damage in protein crystals depend on x-ray dose?" *Structure*, vol. 11, no. 1, pp. 13–19, 2003.

[33] F. Stellato, D. Oberthür, M. Liang, R. Bean, C. Gati, O. Yefanov, A. Barty, A. Burkhardt, P. Fischer, L. Galli *et al.*, "Room-temperature macromolecular serial crystallography using synchrotron radiation," *IUCrJ*, vol. 1, no. 4, pp. 204–212, 2014.

[34] W. C. Röntgen, "Ueber eine neue art von strahlen," *Annalen der Physik*, vol. 300, no. 1, pp. 1–11, 1898.

[35] G. Shen and Y. Wang, "High-pressure apparatus integrated with synchrotron radiation," *Reviews in Mineralogy and Geochemistry*, vol. 78, no. 1, pp. 745–777, 2014.

[36] R. I., G. G., and M. S., "X-ray beams with high coherence," *New Journal of Physics*, vol. 12, 2010.

[37] D. Elektronen-Synchrotron DESY. (2007) What is sr, how is it generated and what are its properties? [Online]. Available: http://photon-science.desy.de/research/studentsteaching/primers

[38] S. W. Cornaby, "A handbook of x-ray single-bounce monocapillary optics, including optical design and synchrotron applications," Ph.D. dissertation, Cornell University, USA, 2008.

[39] R. Bonifacio, C. Pellegrini, and L. Narducci, "Collective instabilities and high-gain regime in a free electron laser," *Optics Communications*, vol. 50, no. 6, pp. 373–378, 1984.

[40] T. Inagaki, T. Tanaka, N. Adumi, T. Fukui, T. Hara, R. Kinjo, H. Maesaka, Y. Otake, K. Togawa, Y. Inubushi *et al.*, "Stable generation of high power self-seeded XFEL at SACLA," *Proceedings of IPAC2014, June*, pp. 15–20, 2014.

[41] R. Huang and D. H. Bilderback, "Single-bounce monocapillaries for focusing synchrotron radiation: modeling, measurements and theoretical limits," *Journal of Synchrotron Radiation*, vol. 13, pp. 74–84, 2006.

[42] M. W. Tate, D. Chamberlain, K. S. Green, H. T. Philipp, P. Purohit, C. Strohman, and S. M. Gruner, "A medium-format, mixed-mode pixel array detector for kilohertz x-ray imaging," *Journal of Physics: Conference Series*, vol. 425, no. 062004, 2013.

[43] N.-T. D. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," *Phys. Rev. E*, vol. 80, p. 026705, Aug 2009.

[44] K. Ayyer, H. T. Philipp, M. W. Tate, J. L. Wierman, V. Elser, and S. M. Gruner, "Determination of crystallographic intensities from sparse data," *IUCrJ*, vol. 2, no. 1, pp. 29–34, Jan 2015.

[45] N. Coquelle, A. S. Brewster, U. Kapp, A. Shilova, B. Weinhausen, M. Burghammer, and J.-P. Colletier, "Raster-scanning serial protein crystallography using micro-and nano-focused synchrotron beams," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 5, pp. 1184–1196, 2015.

[46] K. Hirata, K. Shinzawa-Itoh, N. Yano, S. Takemura, K. Kato, M. Hatanaka, K. Muramoto, T. Kawahara, T. Tsukihara, E. Yamashita *et al.*, "Determination of damage-free crystal structure of an x-ray-sensitive protein using an XFEL," *Nature Methods*, vol. 11, no. 7, pp. 734–736, 2014.

[47] T. D. Murray, A. Y. Lyubimov, C. M. Ogata, H. Vo, M. Uervirojnangkoorn, A. T. Brunger, and J. M. Berger, "A high-transparency, micro-patternable chip for x-ray diffraction analysis of microcrystals under native growth conditions," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 10, pp. 1987–1997, 2015.

[48] M. G. Bowler and M. W. Bowler, "Measurement of the intrinsic variability within protein crystals: implications for sample-evaluation and data-collection strategies," *Acta Crystallographica Section F: Structural Biology Communications*, vol. 70, no. 1, pp. 127–132, 2014.

[49] M. Boudes, D. Garriga, A. Fryga, T. Caradoc-Davies, and F. Coulibaly, "A pipeline for structure determination of in vivo-grown crystals using in cellulo diffraction," *Acta Crystallographica Section D: Structural Biology*, vol. 72, no. 4, pp. 576–585, 2016.

[50] P. Roedig, I. Vartiainen, R. Duman, S. Panneerselvam, N. Stübe, O. Lorbeer, M. Warmer, G. Sutton, D. Stuart, E. Weckert *et al.*, "A micropatterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering," *Scientific Reports*, vol. 5, 2015.

[51] S. Botha, K. Nass, T. R. Barends, W. Kabsch, B. Latz, F. Dworkowski, L. Foucar, E. Panepucci, M. Wang, R. L. Shoeman *et al.*, "Room-temperature serial crystallography at synchrotron x-ray sources using slowly flowing free-standing high-viscosity microstreams," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 2, pp. 387–397, 2015.

[52] A. Y. Lyubimov, T. D. Murray, A. Koehl, I. E. Araci, M. Uervirojnangkoorn, O. B. Zeldin, A. E. Cohen, S. M. Soltis, E. L. Baxter, A. S. Brewster *et al.*, "Capture and x-ray diffraction studies of protein microcrystals in a microfluidic trap array," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 4, pp. 928–940, 2015.

[53] P. Nogly, D. James, D. Wang, T. A. White, N. Zatsepin, A. Shilova, G. Nelson, H. Liu, L. Johansson, M. Heymann *et al.*, "Lipidic cubic phase serial millisecond crystallography using synchrotron radiation," *IUCrJ*, vol. 2, no. 2, pp. 168–176, 2015.

[54] S. Tsukui, F. Kimura, E. F. Garman, S. Baba, N. Mizuno, B. Mikami, and T. Kimura, "X-ray crystal structure analysis of magnetically oriented microcrystals of lysozyme at 1.8 Å resolution," *Journal of Applied Crystallography*, vol. 49, no. 2, 2016.

[55] P. F. Tavares, S. C. Leemann, M. Sjöström, and . Andersson, "The MAX IV storage ring project," *Journal of Synchrotron Radiation*, vol. 21, no. 5, pp. 862–877, 2014.

[56] A. Photon Source. (2016) Beamline directory. [Online]. Available: http://www.aps.anl.gov/Beamlines/Directory/

[57] M. Fodje, P. Grochulski, K. Janzen, S. Labiuk, J. Gorin, and R. Berg, "08B1-1: an automated beamline for macromolecular crystallography ex-

periments at the canadian light source," *Journal of Synchrotron Radiation*, vol. 21, no. 3, pp. 633–637, 2014.

[58] D. Light Source. (2016) Welcome to i24. [Online]. Available: http://www.diamond.ac.uk/Beamlines/Mx/I24.html

[59] U. Zander, G. Bourenkov, A. N. Popov, D. de Sanctis, O. Svensson, A. A. McCarthy, E. Round, V. Gordeliy, C. Mueller-Dieckmann, and G. A. Leonard, "MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 11, pp. 2328–2343, 2015.

[60] A. Burkhardt, T. Pakendorf, B. Reime, J. Meyer, P. Fischer, N. Stübe, S. Panneerselvam, O. Lorbeer, K. Stachnik, M. Warmer *et al.*, "Status of the crystallography beamlines at PETRA III," *The European Physical Journal Plus*, vol. 131, no. 3, pp. 1–9, 2016.

[61] P. Factory. (2016) Sgbbeamlines. [Online]. Available: pfweis.kek.jp/protein/BeamLine/SBGbeamlinesE.pdf

[62] SPring-8. (2016) Bl32xu outline. [Online]. Available: www.spring8.or.jp/wkg/BL32XU

[63] ——. (2016) Bl41xu outline. [Online]. Available: www.spring8.or.jp/wkg/BL41XU

[64] S. Light Source. (2016) Pxi. [Online]. Available: https://www.psi.ch/sls/pxi/px-i-x06sa

[65] R. M. Lawrence, C. E. Conrad, N. A. Zatsepin, T. D. Grant, H. Liu, D. James, G. Nelson, G. Subramanian, A. Aquila, M. S. Hunter *et al.*, "Serial femtosecond x-ray diffraction of enveloped virus microcrystals," *Structural Dynamics*, vol. 2, no. 4, p. 041720, 2015.

[66] J. C. Falkner, A. M. Al-Somali, J. A. Jamison, J. Zhang, S. L. Adrianse, R. L. Simpson, M. K. Calabretta, W. Radding, G. N. Phillips, and V. L. Colvin, "Generation of size-controlled, submicrometer protein crystals," *Chemistry of Materials*, vol. 17, no. 10, pp. 2679–2686, 2005.

[67] H. Demirci, R. G. Sierra, H. Laksmono, R. L. Shoeman, S. Botha, T. R. Barends, K. Nass, I. Schlichting, R. B. Doak, C. Gati *et al.*, "Serial fem-

tosecond x-ray diffraction of 30S ribosomal subunit microcrystals in liquid suspension at ambient temperature using an x-ray free-electron laser," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 69, no. 9, pp. 1066–1069, 2013.

[68] L. C. Johansson, D. Arnlund, G. Katona, T. A. White, A. Barty, D. P. DePonte, R. L. Shoeman, C. Wickstrand, A. Sharma, G. J. Williams *et al.*, "Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography," *Nature Communications*, vol. 4, 2013.

[69] H. J. Kang, C. Lee, and D. Drew, "Breaking the barriers in membrane protein crystallography," *The International Journal of Biochemistry & Cell Biology*, vol. 45, no. 3, pp. 636–644, 2013.

[70] W. Liu, D. Wacker, C. Gati, G. W. Han, D. James, D. Wang, G. Nelson, U. Weierstall, V. Katritch, A. Barty *et al.*, "Serial femtosecond crystallography of G protein–coupled receptors," *Science*, vol. 342, no. 6165, pp. 1521–1524, 2013.

[71] L. Redecke, K. Nass, D. P. DePonte, T. A. White, D. Rehders, A. Barty, F. Stellato, M. Liang, T. R. Barends, S. Boutet *et al.*, "Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an x-ray laser," *Science*, vol. 339, no. 6116, pp. 227–230, 2013.

[72] D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman *et al.*, "Visualizing a protein quake with time-resolved x-ray scattering at a free-electron laser," *Nature Methods*, vol. 11, no. 9, pp. 923–926, 2014.

[73] Q. Liu, Z. Zhang, and W. A. Hendrickson, "Multi-crystal anomalous diffraction for low-resolution macromolecular phasing," *Acta Crystallographica Section D*, vol. 67, pp. 45–49, 2011.

[74] J. Garcıa-Ruiz, M. Novella, R. Moreno, and J. Gavira, "Agarose as crystallization media for proteins: I Transport processes," *Journal of Crystal Growth*, vol. 232, no. 1, pp. 165–172, 2001.

[75] B. Lorber, C. Sauter, A. Théobald-Dietrich, A. Moreno, P. Schellenberger, M.-C. Robert, B. Capelle, S. Sanglier, N. Potier, and R. Giegé, "Crystal growth of proteins, nucleic acids, and viruses in gels," *Progress in Biophysics and Molecular Biology*, vol. 101, no. 1, pp. 13–25, 2009.

[76] S. Faham and J. U. Bowie, "Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure," *Journal of Molecular Biology*, vol. 316, no. 1, pp. 1–6, 2002.

[77] V. Cherezov, M. A. Hanson, M. T. Griffith, M. C. Hilgart, R. Sanishvili, V. Nagarajan, S. Stepanov, R. F. Fischetti, P. Kuhn, and R. C. Stevens, "Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm size x-ray synchrotron beam," *Journal of The Royal Society Interface*, 2009.

[78] S. Sui, Y. Wang, K. W. Kolewe, V. Srajer, R. Henning, J. D. Schiffman, C. Dimitrakopoulos, and S. L. Perry, "Graphene-based microfluidics for serial crystallography," *Lab on a Chip*, pp. –, 2016.

[79] M. Heymann, A. Opthalage, J. Wierman, S. Akella, D. Szebenyi, S. Gruner, and S. Fraden, "Room-temperature serial crystallography using a kinetically optimized microfluidic device for protein crystallization and on-chip X-ray diffraction," *IUCrJ*, vol. 1, pp. 349–360, 2014.

[80] B. G. Abdallah, S. Roy-Chowdhury, R. Fromme, P. Fromme, and A. Ros, "Protein crystallization in an actuated microfluidic nanowell device," *Crystal Growth & Design*, vol. 16, no. 4, pp. 2074–2082, 2016.

[81] E. Chiu, F. Coulibaly, and P. Metcalf, "Insect virus polyhedra, infectious protein crystals that contain virus particles," *Current Opinion in Structural Biology*, vol. 22, pp. 234–240, 2012.

[82] B. G. Abdallah, T.-C. Chao, C. Kupitz, P. Fromme, and A. Ros, "Dielectrophoretic sorting of membrane protein nanocrystals," *ACS nano*, vol. 7, no. 10, pp. 9129–9137, 2013.

[83] D. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. Spence, and R. Doak, "Gas dynamic virtual nozzle for generation of microscopic droplet streams," *Journal of Physics D: Applied Physics*, vol. 41, no. 19, p. 195505, 2008.

[84] U. Weierstall, J. Spence, and R. Doak, "Injector for scattering measurements on fully solvated biospecies," *Review of Scientific Instruments*, vol. 83, no. 3, p. 035108, 2012.

[85] M. J. Bogan, H. Laksmono, and R. G. Sierra, "Apparatus and method for nanoflow liquid jet and serial femtosecond x-ray protein crystallography," patent US 9 275 844, 11, 2013.

[86] R. G. Sierra, C. Gati, H. Laksmono, E. H. Dao, S. Gul, F. Fuller, J. Kern, R. Chatterjee, M. Ibrahim, A. S. Brewster *et al.*, "Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II," *Nature Methods*, 2016.

[87] R. G. Sierra, H. Laksmono, J. Kern, R. Tran, J. Hattne, R. Alonso-Mori, B. Lassalle-Kaiser, C. Glöckner, J. Hellmich, D. W. Schafer *et al.*, "Nanoflow electrospinning serial femtosecond crystallography," *Acta Crystallographica Section D: Biological Crystallography*, vol. 68, no. 11, pp. 1584–1587, 2012.

[88] M. Sugahara, E. Mizohata, E. Nango, M. Suzuki, T. Tanaka, T. Masuda, R. Tanaka, T. Shimamura, Y. Tanaka, C. Suno *et al.*, "Grease matrix as a versatile carrier of proteins for serial crystallography," *Nature Methods*, vol. 12, no. 1, pp. 61–63, 2015.

[89] C. E. Conrad, S. Basu, D. James, D. Wang, A. Schaffer, S. Roy-Chowdhury, N. A. Zatsepin, A. Aquila, J. Coe, C. Gati *et al.*, "A novel inert crystal delivery medium for serial femtosecond crystallography," *IUCrJ*, vol. 2, no. 4, pp. 421–430, 2015.

[90] G. Nelson, "Sample injector fabrication and delivery method development for serial crystallography using synchrotrons and x-ray free electron lasers," Ph.D. dissertation, Arizona State University, 2015.

[91] G. D. Calvey, A. M. Katz, C. B. Schaffer, and L. Pollack, "Mixing injector enables time-resolved crystallography with high hit rate at x-ray free electron lasers," *Structural Dynamics*, vol. 3, no. 5, p. 054301, 2016.

[92] F. Mafuné, K. Miyajima, K. Tono, Y. Takeda, J.-y. Kohno, N. Miyauchi, J. Kobayashi, Y. Joti, E. Nango, S. Iwata *et al.*, "Microcrystal delivery by pulsed liquid droplet for serial femtosecond crystallography," *Acta Crystallographica Section D: Structural Biology*, vol. 72, no. 4, pp. 520–523, 2016.

[93] K. Tono, E. Nango, M. Sugahara, C. Song, J. Park, T. Tanaka, R. Tanaka, Y. Joti, T. Kameshima, S. Ono *et al.*, "Diverse application platform for hard x-ray diffraction in SACLA (DAPHNIS): application to serial protein crystallography using an x-ray free-electron laser," *Journal of Synchrotron Radiation*, vol. 22, no. 3, pp. 532–537, 2015.

[94] A. S. Soares, M. A. Engel, R. Stearns, S. Datwani, J. Olechno, R. Ellson, J. M. Skinner, M. Allaire, and A. M. Orville, "Acoustically mounted mi-

crocrystals yield high-resolution x-ray structures," *Biochemistry*, vol. 50, no. 21, pp. 4399–4401, 2011.

[95] C. G. Roessler, R. Agarwal, M. Allaire, R. Alonso-Mori, B. Andi, J. F. Bachega, M. Bommer, A. S. Brewster, M. C. Browne, R. Chatterjee *et al.*, "Acoustic injectors for drop-on-demand serial femtosecond crystallography," *Structure*, vol. 24, no. 4, pp. 631–640, 2016.

[96] A. S. Soares, J. D. Mullen, R. M. Parekh, G. S. McCarthy, C. G. Roessler, R. Jackimowicz, J. M. Skinner, A. M. Orville, M. Allaire, and R. M. Sweet, "Solvent minimization induces preferential orientation and crystal clustering in serial micro-crystallography on micro-meshes, in situ plates and on a movable crystal conveyor belt," *Journal of Synchrotron Radiation*, vol. 21, no. 6, pp. 1231–1239, 2014.

[97] A. Zarrine-Afsar, T. R. Barends, C. Müller, M. R. Fuchs, L. Lomb, I. Schlichting, and R. D. Miller, "Crystallography on a chip," *Acta Crystallographica Section D: Biological Crystallography*, vol. 68, no. 3, pp. 321–323, 2012.

[98] D. Axford, X. Ji, D. I. Stuart, and G. Sutton, "In cellulo structure determination of a novel cypovirus polyhedrin," *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 5, pp. 1435–1441, 2014.

[99] C. Mueller, A. Marx, S. W. Epp, Y. Zhong, A. Kuo, A. Balo, J. Soman, F. Schotte, H. Lemke, R. Owen *et al.*, "Fixed target matrix for femtosecond time-resolved and in situ serial micro-crystallography," *Structural Dynamics*, vol. 2, no. 5, p. 054302, 2015.

[100] S. Oghbaey, A. Sarracini, H. M. Ginn, O. Pare-Labrosse, A. Kuo, A. Marx, S. W. Epp, D. A. Sherrell, B. T. Eger, Y. Zhong *et al.*, "Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography," *Acta Crystallographica Section D: Structural Biology*, vol. 72, no. 8, 2016.

[101] P. Roedig, R. Duman, J. Sanchez-Weatherby, I. Vartiainen, A. Burkhardt, M. Warmer, C. David, A. Wagner, and A. Meents, "Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering," *Journal of Applied Crystallography*, vol. 49, no. 3, 2016.

[102] E. L. Baxter, L. Aguila, R. Alonso-Mori, C. O. Barnes, C. A. Bonagura, W. Brehmer, A. T. Brunger, G. Calero, T. T. Caradoc-Davies, R. Chatter-

jee *et al.*, "High-density grids for efficient data collection from multiple crystals," *Acta Crystallographica Section D: Structural Biology*, vol. 72, no. 1, 2016.

[103] A. E. Cohen, S. M. Soltis, A. González, L. Aguila, R. Alonso-Mori, C. O. Barnes, E. L. Baxter, W. Brehmer, A. S. Brewster, A. T. Brunger *et al.*, "Goniometer-based femtosecond crystallography with x-ray free electron lasers," *Proceedings of the National Academy of Sciences*, vol. 111, no. 48, pp. 17 122–17 127, 2014.

[104] D. Axford, P. Aller, J. Sanchez-Weatherby, and J. Sandy, "Applications of thin-film sandwich crystallization platforms," *Acta Crystallographica Section F: Structural Biology Communications*, vol. 72, no. 4, pp. 313–319, 2016.

[105] G. K. Feld, M. Heymann, W. H. Benner, T. Pardini, C.-J. Tsai, S. Boutet, M. A. Coleman, M. S. Hunter, X. Li, M. Messerschmidt *et al.*, "Low-Z polymer sample supports for fixed-target serial femtosecond x-ray crystallography," *Journal of Applied Crystallography*, vol. 48, no. 4, pp. 1072–1079, 2015.

[106] C. Broennimann, E. F. Eikenberry, B. Henrich, R. Horisberger, G. Huelsen, E. Pohl, B. Schmitt, C. Shulze-Briese, M. Suzuki, T. Tomizaki, H. Toyokawa, and A. Wagner, "The PILATUS 1M detector," *Journal of Synchrotron Radiation*, vol. 13, pp. 120–130, 2006.

[107] I. Johnson, A. Bergamaschi, H. Billich, S. Cartier, R. Dinapoli, D. Greiffenberg, M. Guizar-Sicairos, B. Henrich, J. Jungmann, D. Mezza *et al.*, "EIGER: a single-photon counting x-ray detector," *Journal of Instrumentation*, vol. 9, no. 05, p. C05032, 2014.

[108] E. N. Gimenez, R. Ballabriga, M. Campbell, I. Horswell, X. Llopart, J. Marchal, K. J. Sawhney, N. Tartoni, and D. Turecek, "Characterization of MEDIPIX3 with synchrotron radiation," *IEEE Transactions on Nuclear Science*, vol. 58, no. 1, pp. 323–332, 2011.

[109] E. Gimenez, R. Ballabriga, M. Campbell, I. Horswell, X. Llopart, J. Marchal, K. Sawhney, N. Tartoni, and D. Turecek, "Study of charge-sharing in MEDIPIX3 using a micro-focused synchrotron beam," *Journal of Instrumentation*, vol. 6, no. 01, p. C01031, 2011.

[110] R. Ballabriga, J. Alozy, G. Blaj, M. Campbell, M. Fiederle, E. Frojdh, E. H. M. Heijne, X. Llopart, M. Pichotka, S. Procz, L. Tlustos, and

W. Wong, "The Medipix3RX: a high resolution, zero dead-time pixel detector readout chip allowing spectroscopic imaging," *Journal of Instrumentation*, vol. 8, pp. C02 016–1 – C02 016–15, 2013.

[111] H. T. Philipp, L. J. Koerner, M. S. Hromalik, M. W. Tate, and S. M. Gruner, "Femtosecond radiation experiment detector for x-ray free-electron laser (XFEL) coherent x-ray imaging," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3795–3799, 2010.

[112] H. T. Philipp, K. Ayyer, M. W. Tate, V. Elser, and S. M. Gruner, "Solving structure with sparse randomly-oriented x-ray data," *Optics Express*, vol. 20, pp. 13 129–13 137, 2012.

[113] A. Mozzanica, A. Bergamaschi, M. Brueckner, S. Cartier, R. Dinapoli, D. Greiffenberg, J. Jungmann-Smith, D. Maliakal, D. Mezza, M. Ramilli *et al.*, "Characterization results of the JUNGFRAU full scale readout ASIC," *Journal of Instrumentation*, vol. 11, no. 02, p. C02047, 2016.

[114] Q. Zhang, E. M. Dufresne, P. Grybos, P. Kmon, P. Maj, S. Narayanan, G. W. Deptuch, R. Szczygiel, and A. Sandy, "Submillisecond x-ray photon correlation spectroscopy from a pixel array detector with fast dual gating and no readout dead-time," *Journal of Synchrotron Radiation*, vol. 23, no. 3, 2016.

[115] H. T. Philipp, M. W. Tate, P. Purohit, K. S. Shanks, J. T. Weiss, and S. M. Gruner, "High-speed x-ray imaging pixel array detector for synchrotron bunch isolation," *Journal of Synchrotron Radiation*, vol. 23, no. 2, 2016.

[116] Z. Zhang, N. K. Sauter, H. van den Bedem, G. Snell, and A. M. Deacon, "Automated diffraction image analysis and spot searching for high-throughput crystal screening," *Journal of Applied Crystallography*, vol. 39, no. 1, pp. 112–119, 2006.

[117] J. Aishima, R. L. Owen, D. Axford, E. Shepherd, G. Winter, K. Levik, P. Gibbons, A. Ashton, and G. Evans, "High-speed crystal detection and characterization using a fast-readout detector," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 9, pp. 1032–1035, 2010.

[118] J. A. Wojdyla, E. Panepucci, I. Martiel, S. Ebner, C.-Y. Huang, M. Caffrey, O. Bunk, and M. Wang, "Fast two-dimensional grid and transmission x-ray microscopy scanning methods for visualizing and characterizing protein crystals," *Journal of Applied Crystallography*, vol. 49, no. 3, 2016.

[119] A. J. Warren, W. Armour, D. Axford, M. Basham, T. Connolley, D. R. Hall, S. Horrell, K. E. McAuley, V. Mykhaylyk, A. Wagner *et al.*, "Visualization of membrane protein crystals in lipid cubic phase using x-ray imaging," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 7, pp. 1252–1259, 2013.

[120] D. J. Kissick, "Second-order nonlinear optical imaging of chiral crystals: Insturmentation and application to protein crystalization," Ph.D. dissertation, PURDUE UNIVERSITY, 2012.

[121] J. A. Newman, S. Zhang, S. Z. Sullivan, X. Y. Dow, M. Becker, M. J. Sheedlo, S. Stepanov, M. S. Carlsen, R. M. Everly, C. Das *et al.*, "Guiding synchrotron x-ray diffraction by multimodal video-rate protein crystal imaging," *Journal of Synchrotron Radiation*, vol. 23, no. 4, 2016.

[122] A. Barty, R. A. Kirian, F. R. Maia, M. Hantke, C. H. Yoon, T. A. White, and H. Chapman, "Cheetah: software for high-throughput reduction and analysis of serial femtosecond x-ray diffraction data," *Journal of Applied Crystallography*, vol. 47, no. 3, pp. 1118–1131, 2014.

[123] W. Minor, M. Cymborowski, Z. Otwinowski, and M. Chruszcz, "HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes," *Acta Crystallographica Section D: Biological Crystallography*, vol. 62, no. 8, pp. 859–866, 2006.

[124] W. Kabsch, "XDS," *Acta Cryst*, vol. D66, pp. 125–132, 2010.

[125] T. Battye, L. Kontogiannis, O. Johnson, H. Powell, and A. Leslie, "IMosflm: a new graphical interface for diffraction-image processing with MOSFLM," *Acta Cryst*, vol. 67, pp. 271–281, 2011.

[126] H. R. Powell, "The rossmann fourier autoindexing algorithm in mosflm," *Acta Crystallographica Section D: Biological Crystallography*, vol. 55, no. 10, pp. 1690–1695, 1999.

[127] T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty, and H. N. Chapman, "CrystFEL: a software suite for snapshot serial crystallography," *Journal of Applied Crystallography*, vol. 45, no. 2, pp. 335–341, 2012.

[128] T. A. White, A. Barty, F. Stellato, J. M. Holton, R. A. Kirian, N. A. Zatsepin, and H. N. Chapman, "Crystallographic data processing for free-electron

laser sources," *Acta Crystallographica Section D: Biological Crystallography*, vol. 69, no. 7, pp. 1231–1240, 2013.

[129] T. A. White, V. Mariani, W. Brehm, O. Yefanov, A. Barty, K. R. Beyerlein, F. Chervinskii, L. Galli, C. Gati, T. Nakane *et al.*, "Recent developments in crystFEL," *Journal of Applied Crystallography*, vol. 49, no. 2, pp. 680–689, 2016.

[130] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve *et al.*, "PHENIX: a comprehensive python-based system for macromolecular structure solution," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 2, pp. 213–221, 2010.

[131] D. Szebenyi, A. Arvai, S. Ealick, J. LaIuppa, and C. Nielsen, "A system for integrated collection and analysis of crystallographic diffraction data," *Journal of synchrotron radiation*, vol. 4, no. 3, pp. 128–135, 1997.

[132] A. Vagin and A. Teplyakov, "Molecular replacement with MOLREP," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 1, pp. 22–25, 2010.

[133] E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. Lea, and G. Bricogne, "Refinement of severely incomplete structures with maximum likelihood in BUSTER–TNT," *Acta Crystallographica Section D: Biological Crystallography*, vol. 60, no. 12, pp. 2210–2221, 2004.

[134] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy *et al.*, "Overview of the CCP4 suite and current developments," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 235–242, 2011.

[135] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot," *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 4, pp. 486–501, 2010.

[136] L. J. Koerner and S. M. Gruner, "X-ray analog pixel array detector for single synchrotron bunch time-resolved imaging," *Journal of Synchrotron Radiation*, vol. 18, pp. 157–164, 2011.

[137] M. Krieger, J. Chambers, G. Christoph, R. Stroud, and B. Trus, "Data collection in protein crystallography: capillary effects and background cor-

rections," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 30, no. 6, pp. 740–748, 1974.

[138] M. Krieger and R. Stroud, "Data collection in protein crystallography: experimental methods for reducing background radiation," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 32, no. 4, pp. 653–656, 1976.

[139] J. L. Wierman, J. S. Alden, C. U. Kim, P. L. McEuen, and S. M. Gruner, "Graphene as a protein crystal mounting material to reduce background scatter," *Journal of Applied Crystallography*, vol. 46, pp. 1501–1507, 2013.

[140] R. Glaeser, M. Facciotti, P. Walian, S. Rouhani, J. Holton, A. MacDowell, R. Celestre, D. Cambie, and H. Padmore, "Characterization of conditions required for x-ray diffraction experiments with protein microcrystals," *Biophysical Journal*, vol. 78, no. 6, pp. 3178–3185, 2000.

[141] C. U. Kim, J. L. Wierman, R. Gillilan, E. Lima, and S. M. Gruner, "A high-pressure cryocooling method for protein crystals and biological samples with reduced background x-ray scatter," *Journal of Applied Crystallography*, vol. 46, no. 1, pp. 234–241, 2013.

[142] P. D. Kwong and Y. Liu, "Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals," *Journal of Applied Crystallography*, vol. 32, no. 1, pp. 102–105, 1999.

[143] A. Riboldi-Tunnicliffe and R. Hilgenfeld, "Cryocrystallography with oil–an old idea revived," *Journal of Applied Crystallography*, vol. 32, no. 5, pp. 1003–1005, 1999.

[144] E. F. Garman and S. Doublié, "Cryocooling of macromolecular crystals: optimization methods," *Methods in Enzymology*, vol. 368, pp. 188–216, 2003.

[145] R. L. Owen, E. Rudiño-Piñera, and E. F. Garman, "Experimental determination of the radiation dose limit for cryocooled protein crystals," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 13, pp. 4912–4917, 2006.

[146] M. Haas, D.; Rossmann, "Crystallographic studies on 95 lactate dehydrogenase at -75°C," *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, vol. 26, pp. 998–1004, 1970.

[147] H. Hope, "Cryocrystallography of biological macromolecules: a generally applicable method," *Acta Crystallographica Section B: Structural Science*, vol. 44, no. 1, pp. 22–26, 1988.

[148] R. Henderson, "Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction," *Proceedings: Biological Sciences*, pp. 6–8, 1990.

[149] D. H. Juers and B. W. Matthews, "Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization," *Quarterly Reviews of Biophysics*, vol. 37, no. 02, pp. 105–119, 2004.

[150] J. M. Yuk, J. Park, P. Ercius, K. Kim, D. J. Hellebusch, M. F. Crommie, J. Y. Lee, A. Zettl, and A. P. Alivisatos, "High-resolution EM of colloidal nanocrystal growth using graphene liquid cells," *Science*, vol. 336, no. 6077, pp. 61–64, 2012.

[151] J. S. Bunch, S. S. Verbridge, J. S. Alden, A. M. Van Der Zande, J. M. Parpia, H. G. Craighead, and P. L. McEuen, "Impermeable atomic membranes from graphene sheets," *Nano Letters*, vol. 8, no. 8, pp. 2458–2462, 2008.

[152] G. Evans, D. Axford, and R. L. Owen, "The design of macromolecular crystallography diffraction experiments," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 261–270, 2011.

[153] A. J. Warren, A. D. Crawshaw, J. Trincao, P. Aller, S. Alcock, I. Nistea, P. S. Salgado, and G. Evans, "*In vacuo* x-ray data collection from graphene-wrapped protein crystals," *Acta Crystallographica Section D: Biological Crystallography*, vol. 71, no. 10, pp. 2079–2088, 2015.

[154] K. Hirata, J. Foadi, G. Evans, K. Hasegawa, and O. B. Zeldin, "Structural biology with microfocus beamlines," *Advanced Methods in Structural Biology*, pp. 241–273, 2016.

[155] A. C. Neto and K. Novoselov, "New directions in science and technology: two-dimensional crystals," *Reports on Progress in Physics*, vol. 74, no. 8, p. 082501, 2011.

[156] C. Lee, X. Wei, J. W. Kysar, and J. Hone, "Measurement of the elastic properties and intrinsic strength of monolayer graphene," *Science*, vol. 321, no. 5887, pp. 385–388, 2008.

[157] X. Li, W. Cai, J. An, S. Kim, J. Nah, D. Yang, R. Piner, A. Velamakanni, I. Jung, E. Tutuc *et al.*, "Large-area synthesis of high-quality and uniform graphene films on copper foils," *Science*, vol. 324, no. 5932, pp. 1312–1314, 2009.

[158] J. M. Yuk, K. Kim, B. Alemán, W. Regan, J. H. Ryu, J. Park, P. Ercius, H. M. Lee, A. P. Alivisatos, M. F. Crommie *et al.*, "Graphene veils and sandwiches," *Nano Letters*, vol. 11, no. 8, pp. 3290–3294, 2011.

[159] N. Mohanty, M. Fahrenholtz, A. Nagaraja, D. Boyle, and V. Berry, "Impermeable graphenic encasement of bacteria," *Nano Letters*, vol. 11, no. 3, pp. 1270–1275, 2011.

[160] T.-P. Ko, J. Day, A. Greenwood, and A. McPherson, "Structures of three crystal forms of the sweet protein thaumatin," *Acta Crystallographica Section D*, vol. 50, pp. 831–825, 1994.

[161] X. Li, C. W. Magnuson, A. Venugopal, R. M. Tromp, J. B. Hannon, E. M. Vogel, L. Colombo, and R. S. Ruoff, "Large-area graphene single crystals grown by low-pressure chemical vapor deposition of methane on copper," *Journal of the American Chemical Society*, vol. 133, no. 9, pp. 2816–2819, 2011.

[162] Z. Otwinowski and W. Minor, "Processing of x-ray diffraction data collected in oscillation mode," *Methods in Enzymology*, vol. 276, pp. 307–326, 1997.

[163] M. A. Berger, J. H. Decker, and I. I. Mathews, "Diffraction study of protein crystals grown in cryoloops and micromounts," *Journal of Applied Crystallography*, vol. 43, no. 6, pp. 1513–1518, 2010.

[164] C. U. Kim, R. Kapfer, and S. M. Gruner, "High-pressure cooling of protein crystals without cryoprotectants," *Acta Crystallographical Section D*, vol. 61, pp. 881 – 890, 2005.

[165] R. Nair, H. Wu, P. Jayaram, I. Grigorieva, and A. Geim, "Unimpeded permeation of water through helium-leak–tight graphene-based membranes," *Science*, vol. 335, no. 6067, pp. 442–444, 2012.

[166] R. D. Wampler, D. J. Kissick, C. J. Dehen, E. J. Gualtieri, J. L. Grey, H.-F. Wang, D. H. Thompson, J.-X. Cheng, and G. J. Simpson, "Selective detection of protein crystals by second harmonic microscopy," *Journal of the American Chemical Society*, vol. 130, no. 43, pp. 14 076–14 077, 2008.

[167] D. J. Kissick, E. J. Gualtieri, G. J. Simpson, and V. Cherezov, "Nonlinear optical imaging of integral membrane protein crystals in lipidic mesophases," *Analytical Chemistry*, vol. 82, no. 2, pp. 491–497, 2009.

[168] B. Zheng, L. S. Roach, and R. F. Ismagilov, "Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets," *Journal of the American Chemical Society*, vol. 125, no. 37, pp. 11 170–11 171, 2003.

[169] M. Ildefonso, N. Candoni, and S. Veesler, "Heterogeneous nucleation in droplet-based nucleation measurements," *Crystal Growth & Design*, vol. 13, no. 5, pp. 2107–2110, 2013.

[170] Š. Selimović, F. Gobeaux, and S. Fraden, "Mapping and manipulating temperature–concentration phase diagrams using microfluidics," *Lab on a Chip*, vol. 10, no. 13, pp. 1696–1699, 2010.

[171] R. D. Dombrowski, J. D. Litster, N. J. Wagner, and Y. He, "Modeling the crystallization of proteins and small organic molecules in nanoliter drops," *AIChE Journal*, vol. 56, no. 1, pp. 79–91, 2010.

[172] M. B. Romanowsky, A. R. Abate, A. Rotem, C. Holtze, and D. A. Weitz, "High throughput production of single core double emulsions in a parallelized microfluidic device," *Lab on a Chip*, vol. 12, no. 4, pp. 802–807, 2012.

[173] J.-u. Shim, G. Cristobal, D. R. Link, T. Thorsen, and S. Fraden, "Using microfluidics to decouple nucleation and growth of protein crystals," *Crystal Growth and Design*, vol. 7, no. 11, pp. 2192–2194, 2007.

[174] C. H. Schmitz, A. C. Rowat, S. Köster, and D. A. Weitz, "Dropspots: a picoliter array in a microfluidic device," *Lab on a Chip*, vol. 9, no. 1, pp. 44–49, 2009.

[175] C. L. Hansen, S. Classen, J. M. Berger, and S. R. Quake, "A microfluidic device for kinetic optimization of protein crystallization and in situ structure determination," *Journal of the American Chemical Society*, vol. 128, no. 10, pp. 3142–3143, 2006.

[176] L. Li, D. Mustafi, Q. Fu, V. Tereshko, D. L. Chen, J. D. Tice, and R. F. Ismagilov, "Nanoliter microfluidic hybrid method for simultaneous screening and optimization validated with crystallization of membrane proteins,"

Proceedings of the National Academy of Sciences, vol. 103, no. 51, pp. 19 243–19 248, 2006.

[177] K. Dhouib, C. K. Malek, W. Pfleging, B. Gauthier-Manuel, R. Duffait, G. Thuillier, R. Ferrigno, L. Jacquamet, J. Ohana, J.-L. Ferrer *et al.*, "Microfluidic chips for the crystallization of biomacromolecules by counter-diffusion and on-chip crystal x-ray analysis," *Lab on a Chip*, vol. 9, no. 10, pp. 1412–1421, 2009.

[178] S. Guha, S. L. Perry, A. S. Pawate, and P. J. Kenis, "Fabrication of x-ray compatible microfluidic platforms for protein crystallization," *Sensors and Actuators B: Chemical*, vol. 174, pp. 1–9, 2012.

[179] S. V. Akella, A. Mowitz, M. Heymann, and S. Fraden, "Emulsion-based technique to measure protein crystal nucleation rates of lysozyme," *Crystal Growth & Design*, vol. 14, no. 9, pp. 4487–4509, 2014.

[180] C. Holtze, A. Rowat, J. Agresti, J. Hutchison, F. Angile, C. Schmitz, S. Köster, H. Duan, K. Humphry, R. Scanga *et al.*, "Biocompatible surfactants for water-in-fluorocarbon emulsions," *Lab on a Chip*, vol. 8, no. 10, pp. 1632–1639, 2008.

[181] T. Sanchez, D. T. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, "Spontaneous motion in hierarchically assembled active matter," *Nature*, vol. 491, no. 7424, pp. 431–434, 2012.

[182] J. D. Tice, H. Song, A. D. Lyon, and R. F. Ismagilov, "Formation of droplets and mixing in multiphase microfluidics at low values of the Reynolds and the capillary numbers," *Langmuir*, vol. 19, no. 22, pp. 9127–9133, 2003.

[183] L. Tang and N. Y. Lee, "A facile route for irreversible bonding of plastic-PDMS hybrid microdevices at room temperature," *Lab on a Chip*, vol. 10, no. 10, pp. 1274–1280, 2010.

[184] P. R. Evans, "An introduction to data reduction: space-group determination, scaling and intensity statistics," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 282–292, 2011.

[185] H. Carrell, J. P. Glusker, V. Burger, F. Manfre, D. Tritsch, and J.-F. Biellmann, "X-ray analysis of d-xylose isomerase at 1.9 a: native enzyme in complex with substrate and with a mechanism-designed inactivator," *Proceedings of the National Academy of Sciences*, vol. 86, no. 12, pp. 4440–4444, 1989.

[186] A. T. Briinger, "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures," *Nature*, vol. 355, pp. 472–475, 1992.

[187] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin, "REFMAC5 for the refinement of macromolecular crystal structures," *Acta Crystallographica Section D: Biological Crystallography*, vol. 67, no. 4, pp. 355–367, 2011.

[188] B. Weinhausen and S. Köster, "Microfluidic devices for x-ray studies on hydrated cells," *Lab on a Chip*, vol. 13, no. 2, pp. 212–215, 2013.

[189] D. James, "Injection Methods and Instrumentation for Serial X-ray Free Electron Laser Experiments," Ph.D. dissertation, Arizona State University, 2015.

[190] A. Cheng, B. Hummel, H. Qiu, and M. Caffrey, "A simple mechanical mixer for small viscous lipid-containing samples," *Chemistry and Physics of Lipids*, vol. 95, no. 1, pp. 11–21, 1998.

[191] N. Loh, M. Bogan, V. Elser, A. Barty, S. Boutet, S. Bajt, J. Hajdu, T. Ekeberg, F. R. Maia, J. Schulz *et al.*, "Cryptotomography: reconstructing 3D fourier intensities from randomly oriented single-shot diffraction patterns," *Physical Review Letters*, vol. 104, no. 22, p. 225501, 2010.

[192] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[193] V. Elser, "Noise limits on reconstructing diffraction signals from random tomographs," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4715–4722, 2009.

[194] C. Nave and M. A. Hill, "Will reduced radiation damage occur with very small crystals?" *Journal of Synchrotron Radiation*, vol. 12, no. 3, pp. 299–303, 2005.

[195] T. Hahn, Ed., *International Tables for Crystallography*, 1st ed. Chester: International Union of Crystallography, 2006, vol. A.

[196] C. Riekel, M. Burghammer, and G. Schertler, "Protein crystallography mi-

crodiffraction," *Current Opinion in Structural Biology*, vol. 15, no. 5, pp. 556–562, 2005.

[197] D. Flot, T. Mairs, T. Giraud, M. Guijarro, M. Lesourd, V. Rey, D. van Brussel, C. Morawe, C. Borel, O. Hignette *et al.*, "The ID23-2 structural biology microfocus beamline at the ESRF," *Journal of Synchrotron Radiation*, vol. 17, no. 1, pp. 107–118, 2010.